

STUDY PROTOCOL

Open Access



Deep learning diagnostic and severity-stratification for interstitial lung diseases and chronic obstructive pulmonary disease in digital lung auscultations and ultrasonography: clinical protocol for an observational case–control study

Johan N. Siebert^{1,6*†} , Mary-Anne Hartley^{2†}, Delphine S. Courvoisier^{3,6}, Marlène Salamin⁴, Laura Robotham⁴, Jonathan Doenz², Constance Barazzone-Argiroffo^{5,6}, Alain Gervais^{1,6} and Pierre-Olivier Bridevaux⁴

Abstract

Background Interstitial lung diseases (ILD), such as idiopathic pulmonary fibrosis (IPF) and non-specific interstitial pneumonia (NSIP), and chronic obstructive pulmonary disease (COPD) are severe, progressive pulmonary disorders with a poor prognosis. Prompt and accurate diagnosis is important to enable patients to receive appropriate care at the earliest possible stage to delay disease progression and prolong survival. Artificial intelligence-assisted lung auscultation and ultrasound (LUS) could constitute an alternative to conventional, subjective, operator-related methods for the accurate and earlier diagnosis of these diseases. This protocol describes the standardised collection of digitally-acquired lung sounds and LUS images of adult outpatients with IPF, NSIP or COPD and a deep learning diagnostic and severity-stratification approach.

Methods A total of 120 consecutive patients (≥ 18 years) meeting international criteria for IPF, NSIP or COPD and 40 age-matched controls will be recruited in a Swiss pulmonology outpatient clinic, starting from August 2022. At inclusion, demographic and clinical data will be collected. Lung auscultation will be recorded with a digital stethoscope at 10 thoracic sites in each patient and LUS images using a standard point-of-care device will be acquired at the same sites. A deep learning algorithm (*DeepBreath*) using convolutional neural networks, long short-term memory models, and transformer architectures will be trained on these audio recordings and LUS images to derive an automated diagnostic tool. The primary outcome is the diagnosis of ILD versus control subjects or COPD. Secondary outcomes are the clinical, functional and radiological characteristics of IPF, NSIP and COPD diagnosis. Quality of life will be measured with dedicated questionnaires. Based on previous work to distinguish normal and pathological lung sounds, we

[†]Johan N. Siebert and Mary-Anne Hartley contributed equally to this work

*Correspondence:

Johan N. Siebert

Johan.Siebert@hcuge.ch

Full list of author information is available at the end of the article



estimate to achieve convergence with an area under the receiver operating characteristic curve of > 80% using 40 patients in each category, yielding a sample size calculation of 80 ILD (40 IPF, 40 NSIP), 40 COPD, and 40 controls.

Discussion This approach has a broad potential to better guide care management by exploring the synergistic value of several point-of-care-tests for the automated detection and differential diagnosis of ILD and COPD and to estimate severity.

Trial registration Registration: August 8, 2022. ClinicalTrials.gov Identifier: NCT05318599.

Keywords Lung diseases, Interstitial, Idiopathic pulmonary fibrosis, Idiopathic interstitial pneumonias, Pulmonary disease, Chronic obstructive, Deep learning, Artificial intelligence, Respiratory sounds, Auscultation, Ultrasonography

Background

Interstitial lung disease (ILD) encompasses several pulmonary conditions defined by an alteration of the pulmonary interstitium, a restrictive pattern of lung function, and fibrotic scarring on chest computed tomography (CT). Approximately one-third of these disorders have known endogenous or exogenous causes, including environmental or occupational factors, infections, drugs and radiation. Two-thirds are of idiopathic aetiology [1] and comprise a range of subcategories, the most common of which is idiopathic interstitial pneumonia (IIP). In turn, IIP comprises a range of sub-pathologies, such as idiopathic pulmonary fibrosis (IPF) and non-specific interstitial pneumonia (NSIP) [2]. Identifying patients with IIP at the earliest possible stage is essential for care management as treatment is aimed at slowing the irreversibly debilitating and ultimately fatal progression. Delay in specialist referral is associated with a higher mortality, irrespective of disease severity [3]. With a mean delay of 2.2 years between the onset of symptoms and specialist referral, the investigation of competing diagnoses by non-specialist providers can be costly for both patients and healthcare providers [3]. However, given the initial non-specific symptomatic presentation, the need for advanced diagnostic tools, such as high-resolution chest CT (HRCT), and an expertise in the early-stage diagnosis of IPF and NSIP remain desirable and achievable objectives [4]. Distinguishing between IPF and NSIP raises considerable diagnostic challenges as their clinical presentations share many overlapping features. However, the distinction is useful as their response to treatment differs markedly [5]. Until now, with limited treatment options benefiting mostly patients in the early stages of the disease [6, 7], many patients will progress towards disability or death [8]. Despite research and advances in therapy, ILDs remain a worldwide health challenge affecting millions of people each year [9], emphasizing the need to make progress in diagnosis and prevention.

Different measures have been proposed to improve the early diagnosis of ILDs [10, 11]. In particular, the identification of the so-called “velcro”-like crackles on lung

auscultation by primary care doctors has been suggested as an early and strongly predictive sign of IPF or fibrotic NSIP [4, 12]. For instance, while IPF and NSIP typically have fine velcro-like crackles audible on the mid-to-late inspiratory cycle, chronic obstructive pulmonary disease (COPD) tends to have coarse crackles occurring during the early inspiratory cycle [13]. As stethoscopes are readily available, inexpensive and non-invasive, they constitute an adequate tool to detect velcro-like crackles in the early stages of IPF or fibrotic NSIP to shorten the diagnostic delay and allow the prompt referral to specialised care. However, conventional auscultation is a highly subjective skill limited by inter-listener variability and human perceptual ability to distinguish between lung sounds and their temporal occurrence in the respiratory cycle. Inherent heterogeneity in stethoscope quality, background noise and patient-related factors, such as obesity or chest deformities, are other limiting factors. To overcome these drawbacks, research efforts have been devoted to improve computerised respiratory sound recording with electronic stethoscopes and an objective analysis based on advanced digital acoustic signal processing [14–17]. The advent of deep learning in recent years took the analysis of auscultation signals one step further by allowing an enhanced detection of abnormal lung sounds in patients with respiratory diseases [18–20].

Several studies have assessed the broad adoption and impact of deep learning to help diagnose COPD [21–23], the third leading cause of death worldwide [24]. The vast majority developed predictive models to cover a wide range of objectives, the main ones being the diagnosis and severity classification of the disease [25, 26]. A 2022 review of artificial intelligence (AI) techniques in COPD yielded 156 articles relevant to the application of AI in COPD research, including 56 concerning diagnosis, 65 on its prognosis, 54 on COPD severity classification, and 17 on the management of the disease [27]. Most studies have used a variety of features, including patient physiological characteristics, comorbidities, symptoms, vital signs, biomarkers, genomic information, pulmonary function tests, CT images, hospitalization

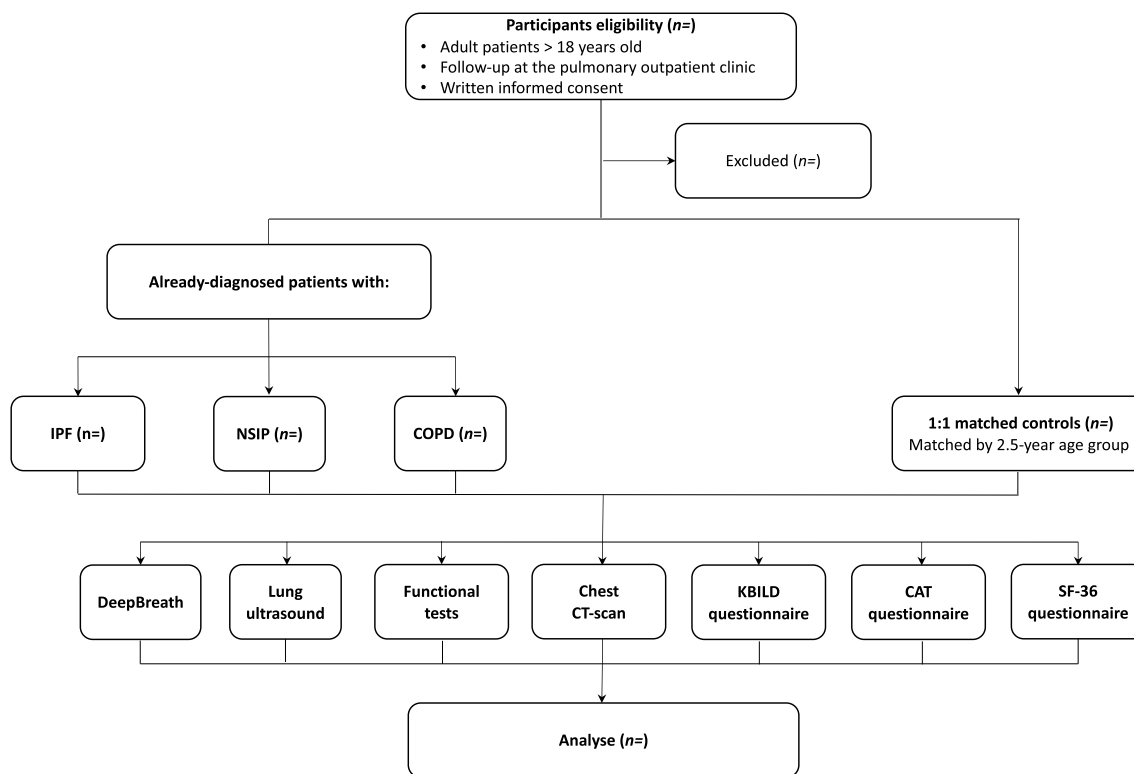


Fig. 1 Study flowchart

information, and/or breath sounds [28, 29]. Regardless of the method(s) chosen, COPD remains an incurable and progressive disease and diagnosis at the early risk stage is important. In this sense, the work of Altan et al. is innovative. The deep learning algorithms they used on analysing multiple lung auscultation points for the early diagnosis of COPD achieved high classification performance rates [30, 31]. Achieving this with a method as conventional as lung auscultation can reduce the need for additional, more extensive, time-consuming, expensive or invasive diagnostic tests.

Conversely, research regarding IPF and NSIP are scarce and have focused mostly on datasets collected through radiological [32–34], genomic [35, 36] or functional tests [37]. Pancaldi et al. described the use of an AI algorithm to detect the presence of velcro-like crackles in patients with rheumatoid arthritis and a suspicion of ILD [17, 38]. However, to our knowledge, no study has investigated the benefit of deep learning-aided diagnostic tools for early IPF and NSIP diagnosis using respiratory sound analysis in adults. This might allow doctors to assess acoustic signatures more objectively and thus allow a more standardised and potentially earlier diagnosis in patients presenting at primary care clinics with non-specific, chronic respiratory symptoms. On the other hand,

lung ultrasound (LUS) is already the standard of care for detecting consolidations, diagnosing pneumonia and guiding pleural taps. The distinction between A (normal aeration) and B (alveolar-interstitial syndrome) lines on LUS is clinically important and forms the backbone of multiple clinical decision trees for real-time respiratory diagnoses and treatment choices [39]. As such, not only is LUS a relevant gold standard for lung pathology, but it could also benefit from automation by deep learning.

We developed a series of deep learning algorithms on digital lung auscultation (*DeepBreath*) and LUS to detect a range of physiological and pathological lung diseases, including (COVID-19) [40]. This study will seek to explore the synergistic value of several point-of-care-tests for the AI-aided detection and differential diagnosis of ILD and COPD, as well as estimate of severity, with the aim to better guide and improve care management in adults.

Methods

Study design

This is a single-centre, prospective, population-based, case–control study that will be carried out in subjects with IPF, NSIP and COPD within a pulmonology outpatient clinic in Switzerland, with a total of approximately

Table 1 Study schedule

	Study period		
	Pre-enrolment	Enrolment	Close-out
Timepoint	$-t_1$	t_0	
Study procedures			
Recruitment		✓	
Eligibility screening		✓	
Informed consent form		✓	
Assessments			
Case report form (demographic + clinical data)		✓	✓
EKO CORE digital auscultation collection		✓	✓
Pulmonary function tests		✓	
Chest CT-scan/X-ray	✓		
Lung ultrasound		✓	
K-BILD questionnaire		✓	
CAT questionnaire		✓	
SF-36 questionnaire		✓	

✓: Performed

7000 specialised consultations per year. Figure 1 shows the study flowchart and Table 1 details the study schedule. The present study protocol adheres to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) Statement [41].

Population

Inclusion criteria are consecutive, consenting adult outpatients (> 18 years) with IPF (group 1), NSIP (group 2) and COPD (group 3) already diagnosed prior to the consultation (index) date. Probable and definitive IPF diagnosis will be made according to the Fleischner Society Consensus [42], NSIP diagnosis with the American Thoracic Society classification [2, 43], and COPD with the Global Initiative for Chronic Obstructive Lung Disease criteria [44]. Consenting, age-matched (± 2.5 years) individuals with normal lung function (spirometry, lung volume and transfer factor for carbon monoxide [TL_{CO}]) followed in the outpatient clinic with a similar quality of electronic medical records, but for diseases other than the outcome of interest, will serve as the 1:1 control group (group 4). This latter group will comprise patients with obstructive sleep apnoea, follow-up of occupational lung diseases (miners, chemical workers, etc.), and follow-up of pulmonary nodules (considered benign after 2 years). Reasons for pulmonary follow-up among the controls will be reviewed and reported in a supplementary file. Identifying all cases with the outcome of interest and selecting controls for comparison is a more efficient and resource-sparing study design than a full cohort study. Exclusion criteria are: (1) patients who cannot be

mobilised for posterior auscultation; (2) those known for severe cardiovascular disease with pulmonary repercussions; (3) patients known for a concurrent, acute, infectious pulmonary disease (e.g., pneumonia, bronchitis); (4) patients known for asthma exclusive from COPD; (5) patients with alpha-1-antitrypsin deficit; (6) a physical inability to follow procedures; and 7) inability to give informed consent.

Recruitment and informed consent procedure

Patients will be recruited from an outpatient pulmonology clinic in Switzerland in daily clinical practice. Participants will provide written informed consent, provided that they have had sufficient time for consideration and the opportunity to ask questions. Important concepts will be highlighted via bulleted text. A checkbox will assess whether participants understand key consent information in the presence of study investigators. These consent forms will be collected and countersigned by the study investigators and stored securely in an access-controlled room. No financial compensation will be offered to participants.

We anticipate that withdrawal and discontinuation will be limited as the study offers the advantage of taking place in a single centre and during a single, short (i.e., 60 min) intervention period on the day of a routine clinical visit. In the case of withdrawal after informed consent, the individual’s data collected so far and related to the intervention will be destroyed/deleted. Any withdrawal and/or discontinuation will be justified and reported in final publications in anonymised form.

Hypothesis and objectives

Primary hypothesis

We hypothesise that point-of-care digital lung auscultation and LUS have a clinically exploitable predictive performance for the detection of pathological acoustic and sonographic signatures in patients with ILD. Furthermore, we propose that these signatures are sufficiently unique to not only discriminate ILD patients from control subjects, but also from COPD and other respiratory diseases, and perhaps even to categorise the various severity grades and subtypes of ILD, when determined. We further hypothesise that the automated interpretation of lung auscultation and LUS by deep learning could match or outperform expert evaluation and standardise lung auscultation and LUS interpretation.

Primary objective

To collect a systematic sound bank of digital lung auscultation and images for the development of deep learning algorithms that predict pathological signatures of ILD in an adult population to: (1) discriminate ILD from non-ILD lung sounds and images; (2) predict ILD clinical severity; (3) differentiate ILD from COPD; and (4) possibly determine the subcategories of ILD (i.e., IPF versus NSIP).

Secondary hypothesis

International clinical practice guidelines recommend to suspect IPF and NSIP in the presence of velcro-like crackles [45] (and similarly for coarse crackles in COPD). However, there are few data indicating whether these sounds are associated with clinical, functional, and radiological characteristics upon ILD diagnosis [4, 46].

Secondary objective

To investigate whether velcro-like crackles labelled by human experts are associated with the aforementioned characteristics in patients with IPF and NSIP (and similarly for coarse crackles and COPD). The impact of the diseases on patients' health-related quality of life will be measured with standardised questionnaires.

Our overall hypothesis is that the use of *DeepBreath* might substantially improve the early and accurate diagnosis of patients with chronic lung disease.

Primary and secondary outcomes

The primary outcome is the diagnostic of ILD, both IPF and NSIP, versus control subjects or COPD. We will assess the predictive performance of the *DeepBreath* algorithm-evaluated lung auscultation and LUS in the following identification and risk stratification tasks as follows: (1) to discriminate ILD from control subjects (according to expert clinical diagnosis [42]); (2) to

differentiate ILD from COPD; (3) to predict ILD clinical severity (according to a HRCT grading scale¹ and lung function tests²); and (4) to differentiate the subcategories of ILD (such as IPF, NSIP) according to the gold standard diagnosis [2, 43, 44]).

Secondary outcomes are the clinical, functional and radiological characteristics of IPF, NSIP and COPD diagnosis. We will: (1) compare the predictive performance of human, expert-identified acoustic and LUS signatures in the above predictive tasks (Kappa coefficient); (2) assess diagnostic performance of a model trained to detect crackles; (3) explore the utility of adding clinical data (signs, symptoms, demographics, medical history and basic paraclinical tests) to the breath sound algorithms; and (4) determine the impact of the diseases on subjects' health-related quality of life measured with the standardised King's Brief Interstitial Lung Disease (K-BILD) [47], the COPD assessment test (CAT) [48], and the 36-item Medical Outcomes Study Short-Form Health Survey (SF-36) [49] severity assessment questionnaires.

Study procedure

The study will be performed over a period of 6 months. Recruitment can be stopped before the anticipated end if the inclusion of 160 patients is reached before. A trained research nurse/doctor (MS/LR) will recruit the subjects during a single routine consultation in the outpatient clinic. This will include checking the selection criteria for each patient prior to study participation, obtaining written informed consent, administering questionnaires on demographic characteristics (age, sex, occupation, long-term exposure to occupational or environmental agents, etc.), relevant medical history, and symptomatic presentation (Additional file 1), which will be captured by an electronic case report form to be completed by the study coordinator on a tablet. The nurse/doctor will also administer the standardised K-BILD [47], CAT [48], and SF-36 [49] severity assessment questionnaires, and record lung sounds during 5–7 min with an electronic stethoscope in the same zones as LUS acquisition, as previously proposed by our group [40]. For the LUS examination using a standard point-of-care ultrasound device, an adapted version of our previous 10-point acquisition

¹ HRCT severity markers: traction bronchiectasis (3 levels); presence of honeycombing (3 levels); ground glass opacities (3 levels); reticulation (3 levels); and emphysema (3 levels). This analysis will be performed in all subjects with an available CT, and separately in IID and COPD patients and control subjects. Chest CT-scans will be reviewed independently by two radiologists blinded to each other.

² Functional lung test severity markers: forced expiratory volume in 1 s (FEV₁); forced vital capacity (FVC); FEV₁/FVC; total lung capacity (TLC); functional respiratory capacity (FRC); transfer capacity for CO (TL_{CO}); K_{CO}; alveolar volume (AV).

protocol [50] will be used, which involves scanning the anterior superior, anterior inferior, posterior superior, posterior inferior and lateral thorax regions. In addition, pulmonary functional tests (conducted with patients in a stable condition) will be collected, as well as chest X-rays with two incidences (posterior-anterior and lateral) and HRCT scans for IIP patients (within 12 months). Controls will not be exposed to a chest X-ray and/or CT scan unless required as part of their routine follow-up; cases will have undergone such imaging given that it is part of their diagnostic evaluation.

All LUS images and lung sounds captured will be digitally recorded and transferred via a secured internet connection together with relevant metadata to a secure server. For study quality control purposes, the quality of the image and the interpretation of a random sample of images will be evaluated retrospectively by an experienced radiologist. The images will be further used for secondary studies developing machine learning algorithms and AI for LUS diagnosis. As this study will take place during outpatient visits under usual conditions and with conventional diagnostic measurement tools, we do not expect any problems that would put participants at a greater risk than normal exposure in daily clinical practice.

Lung sound recording

The frequency range of normal lung sounds extends from below 100 Hz to 1000 Hz, with a sharp drop at approximately 100 to 200 Hz [51], whereas tracheal sound extends between 100 to 5000 Hz. In the lower band range (under 100 Hz), heart and thoracic muscle sounds overlap. Abnormal lung sounds (wheezing, rhonchi etc.) have characteristic frequencies and duration, differentiating them from each other [51]. In particular, fine velcro-like crackles are caused by explosive openings of the small airways, have a distinguishable high-pitched frequency of about 650 Hz, and a typical short duration of about 5 ms.

In this study, the lung sounds will be gathered digitally in all subjects with the same Eko CORE digital stethoscope (Eko Devices, Inc., CA, USA). Four anterior thoracic sites (superior and inferior bilaterally), 4 posterior sites (superior and inferior bilaterally), and 2 lateral sites (right, left) will be auscultated per patient using the stethoscope. For each auscultation site, a 30-s digital recording will be acquired. Patients will be informed of the necessity to breathe deeply. All signals will be saved as 16-bit resolution, 4 kHz-sampled WAV files. The built-in filter will range from 20 to 2000 Hz. Heart and thoracic muscle sounds, as well as other background low-frequency noises, will be filtered out through EKO software's high-pass filters. Coded recorded sounds will be synced in real-time to a General Data Protection Regulation

(GDPR)-compliant secured cloud-storage location. Random auscultatory recordings will be reviewed by the study investigators for quality control.

LUS

LUS is a well-established, consumable-free and non-invasive point-of-care respiratory examination. While it is less ubiquitous than the stethoscope, its new portable and affordable ultrasound-on-a-chip design, pluggable into a mobile device, has the potential to be integrated into the standard clinical examination without incurring extra costs, time, radiation or specialist consultation. It has been shown to be highly effective in detecting lung consolidation in pneumonia [52]. For COVID-19, its diagnostic accuracy matches that of chest CT [53] and it was previously demonstrated that it has an excellent performance for risk-stratification [50]. LUS has been found to be very sensitive to detect subtle changes in the subpleural space. Fibrosis presents as diffuse, multiple B-lines where thickening or irregularity of the pleural line is associated with scarring and disease advancement. Disease severity is also seen in the total number of B-lines, while the average distance between two adjacent B-lines is an indicator of a particular pattern of fibrosis (e.g., pure reticular fibrosis as in IPF compared with the predominant ground glass pattern seen in fibrotic, nonspecific, interstitial fibrosis). The anatomic distribution of these anomalies may also have some relevance to fibrosis type.

In this study, a trained doctor (LR) will perform all LUS at inclusion. Acquisition will be standardised according to protocol [50]. Two images (sagittal and transverse) and 5-s video clips will be systematically recorded for each of the 10 thoracic sites with a Butterfly IQ (Butterfly Network, Guilford, CT, USA), using the lung preset. Reporting of pathological LUS features will be standardised. For every zone, the following patterns will be reported: (1) normal appearance (A lines, <3 B lines); (2) pathologic B lines (≥ 3 B lines); (3) confluent B lines; (4) thickening of the pleura with pleural line irregularities (subpleural consolidation <1 cm); (5) consolidation (≥ 1 cm); (6) presence of subpleural nodules; (7) presence of pleural effusion; (8) diaphragmatic excursion (in mm); and (9) diaphragmatic thickening (in mm). The LUS score, used as a correlate of loss of lung tissue aeration, as well as a normalised LUS score (nLUS score) corrected for the number of examined zones, will be calculated in every patient [54].

AI algorithms

Diagnostic and risk stratification algorithms

We will develop *DeepBreath*, a deep learning algorithm to detect the acoustic signatures of IPF, NSIP and COPD from lung sounds. While several state-of-the-art

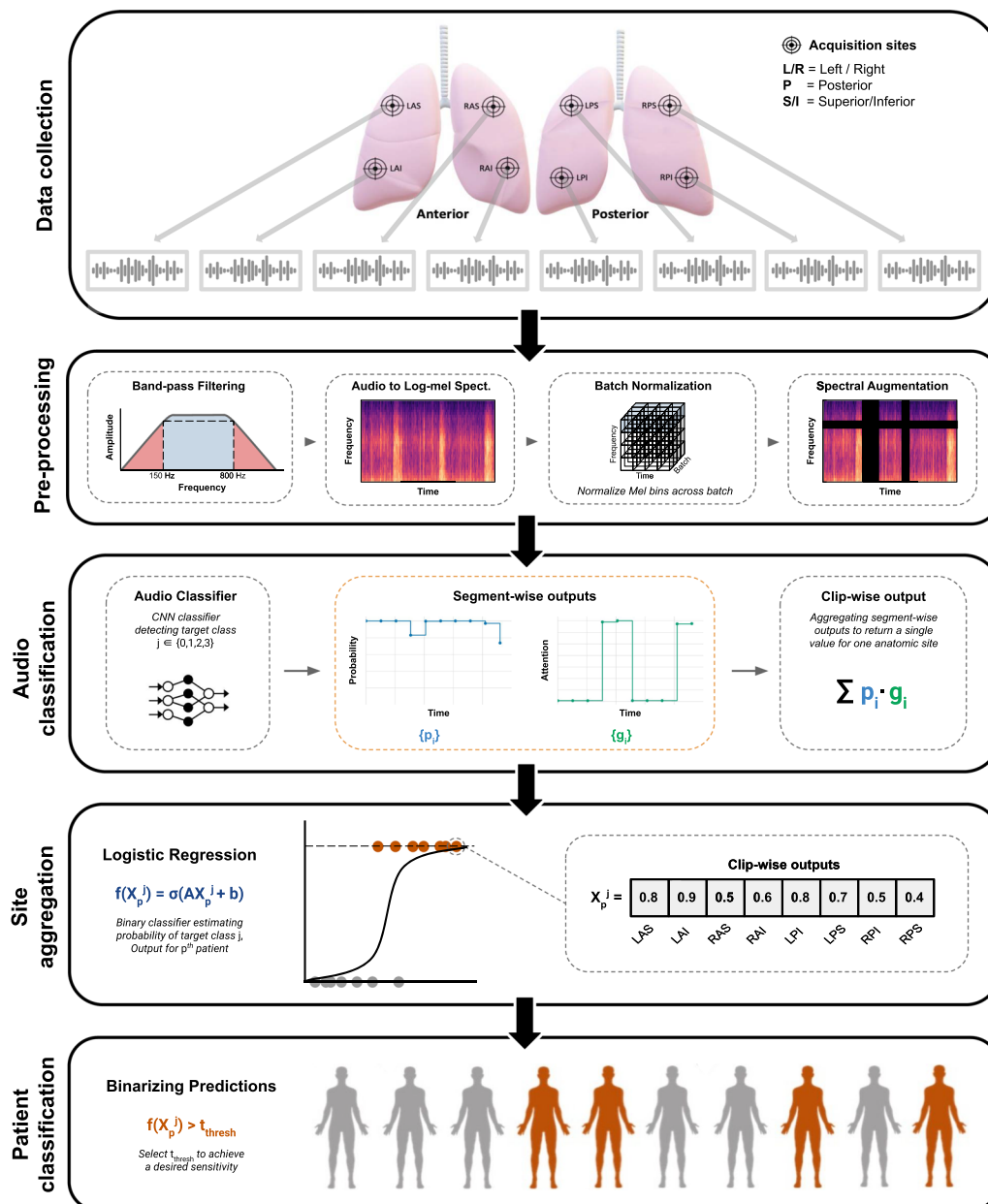


Fig. 2 Overview of the DeepBreath binary classification model. Top to bottom: Data collection. Every patient will have 10 lung recordings corresponding to 1 per 10 anatomical sites (LAS, RAS: Left and Right Anterior Superior; LAI, RAI: Left and Right Anterior Inferior; LPS, RPS: Left and Right Posterior Superior; LPI, RPI: Left and Right Posterior Inferior; Left and Right Lateral [not shown on the figure]). Pre-processing. A band-pass filter is applied to clips before transformation to log-mel spectrograms which are batch-normalised and augmented and then fed into an audio classifier. Here, a CNN outputs both segment-level prediction and attention values which are aggregated into a single clip-wise output for each site. These are then aggregated by concatenation to obtain a feature vector of size for every patient, which is evaluated by a logistic regression. Finally, patient-level classification is performed by thresholding to get a binary output. The segment-wise outputs of the audio classifier are extracted for further analysis. Used with permission from Heitmann et al. (<https://doi.org/10.1038/s41746-023-00838-3>, Nature Digital Medicine) (Swiss Federal Institute of Technology EPFL, Lausanne, Switzerland)

approaches will be tested, the general framework is summarised in Fig. 2. Digital lung auscultations will first be cleaned to crop non-biological frequencies and amplitudes generated by ambient noise not filtered by

the stethoscope’s active noise cancelling. The sounds will then be divided into overlapping time windows of between 1 and 10 s and transformed to Mel Frequency Cepstral Coefficients (MFCCs). Several data

augmentation techniques will be explored, such as amplitude scaling, pitch shift and random time shift. The effect of each pre-processing method will be tested and the best performing approach according to sensitivity and specificity will be reported. This dataset will then be fed into various deep learning networks (such as convolutional neural networks, Long Short-Term Memory models [LSTM], Transformer architectures, etc.). A prediction on each segment will then be aggregated to represent a patient (including all anatomic sites) and binary classification into positive *vs* negative for diagnostic results will be performed for ILD or control subjects, ILD or COPD, and (if ILD-positive) IPF or NSIP. The same prediction will also be made using LUS images. Risk stratification will use multiclass or regression according to scales obtained from clinical interpretation of LUS, lung function tests, HRCT imagery, K-BILD or CAT, and SF-36 quality of life questionnaires.

Exploring the synergy of clinical data with breath sounds

Clinical data will be explored for its predictive capacity in the above tasks and added to the breath sound analysis either as a support vector machine (SVM) or in conditional feature extraction upstream of the neural network.

Clinical assessment of lung auscultation and LUS

The following data will be reviewed by external experts and interpreted using standardised report forms noting the binary presence/absence of several anomalies as well as a text field for other notable observations: routine chest X-ray films (usually 2 incidences, posterior-anterior and lateral); lung auscultation audio clips (10 anatomic regions represented, 30 s recordings of each region); and LUS images (10 anatomic regions represented, 5 s video clips of each region). The analysis will be blinded and the assessor will not have any knowledge of the linked clinical data or association between the various imaging modalities (i.e., IDs are scrambled between media sources and chest X-ray images will be reviewed blinded to the patient's LUS images, auscultation audio, clinical data, etc.). These blinding and standardisation procedures are expected to minimise performer and study management bias respectively.

Pulmonary function tests and chest CT scan

For all subjects, spirometry, body-plethysmographic parameters (see details above) and lung diffusion capacity for carbon monoxide (TLco/Kco) will be measured. Participants' lung images recorded from previous HRCT (or X-rays) during past routine visits will be used. No chest CT scans will be performed as part of this study; only lung images of participants previously recorded as part of their regular follow-up or to be performed in this

context will be used. The presence on the chest CT scan of honeycombing, traction bronchiectasis, reticulations, ground-glass opacities, and emphysema will be measured for patients with IPF, NSIP or emphysema. The main chest CT scan features of IPF are reported to be basal and peripheral reticulations, traction bronchiectasis, minimal ground glass opacities, and moderate or extensive honeycombing. For NSIP, the CT scan typically demonstrates bilateral lung involvement and invariably some extent of ground-glass opacities, mainly in the lower zones with fibrotic changes, while honeycombing is not a common feature [55]. The presence on the chest CT scan of structural abnormality, such as $\geq 5\%$ emphysema and/or $\geq 15\%$ gas-trapping and/or airway wall thickness ≥ 2.5 mm [56], will be measured for patients with COPD. The protocol assumes normal lung parenchyma in the control group, which will not be exposed to radiation unless controls have already undergone a recent (i.e., < 5 years) chest CT scan for other reasons. This will be taken into consideration as we will investigate the association of lung sounds with radiological characteristics.

Questionnaires

Demographics including age, sex, ethnicity, environment (smoking status, long-term exposure to occupational or environmental agents, etc.), treatments, presence of chronic respiratory symptoms or repeated lower respiratory tract infectious diseases, as well as a diagnosis of other comorbidities (obesity, immunodeficiency, alpha-1 antitrypsin deficiency, etc.) will be reported in a questionnaire (Additional file 1). Severity of functional limitations according to the New York Heart Association (NYHA) functional classification [57] will be also reported if available.

The impact of IPF and NSIP on subjects' health-related quality of life will be measured with the standardised K-BILD questionnaire [47] (use with license agreement), which covers 15 questions exploring 3 health dimension scores (psychological, breathlessness and activities, and chest symptoms) using a 7-point Likert response scale (scores range from 0 to 100, a higher score indicating better health status). The impact of COPD will be assessed with the CAT [48] (use with license agreement) that measures eight items: cough; phlegm; chest tightness; breathlessness; limited activities; confidence leaving home; sleeplessness; and energy. Scores range from 0 to 40, with a higher score indicating a more severe impact of COPD on a patient's life. The SF-36 will also be used to assess the impact of IPF, NSIP or COPD on patients' quality of life. The investigators will double-check on-site that the questionnaires are fully and accurately completed. Data collection will be carried out using the online REDCap database (REDCap, Vanderbilt

University, Nashville, TN, USA; <https://www.project-red-cap.org/resources/citations/>). Conditions or complaints occurring after enrolment will not be considered in the statistical analyses. Current symptoms at enrolment will be registered.

Sample size calculation

Each patient will provide 10 audio recordings of 30 s. Samples will be considered at the patient level with all 10 recordings. In deep learning, sample size calculation is an intractable problem that is usually discovered through empirical investigation. The number of samples required to reach a certain performance criterion is dependent on the characteristics of the dataset, the diversity and number of the classes, the degree of data augmentation possible, as well as the complexity of the learning algorithm. Thus, sample size calculations cannot rely on the traditional statistical heuristics that are often used in biostatistics. Rather, sample size estimations in deep learning are mostly made by analogy. Evaluating existing knowledge on similar datasets, we find that the expected proportion of velcro-like crackles in IIP patients is nearly 100% [4], whereas the prevalence of coarse crackles is 71% in COPD patients [58]. The exclusivity of these sounds among groups is not known, but overlap is assumed to be minimal and pathological sounds are by definition absent in non-ILD and non-COPD control subjects. Assuming a similar discriminative power compared to a previous work done by our group (personal communication) to distinguish between normal and pathological lung sounds in pneumonia from 80 patients in balanced classes (40 pathological; 40 controls) with 8 auscultation sites of 30 secs each, we estimate using the same number of patients in each class to achieve convergence at above 80% of the area under the receiver operating characteristic curve (AUROC). Thus, we will aim to enrol at least as many patients in each group: 80 ILD (40 IPE, 40 NSIP); 40 COPD; and 40 controls (i.e., known not to have ILD or COPD, and with normal lung function). As the recruitment site would expect 120 ILD patients (40% with IPF; 60% with NSIP) and 100 with COPD over the space of one year, this number is achievable in the time frame of the study (6 months), even with a 70% consent rate.

This sample size is also predicted to be sufficient for deep learning on LUS. Our preliminary results (personal communication) on COVID diagnosis using deep learning achieved 90% AUROC with 150 patients (balanced classes of 75 COVID+ and 75 COVID-). As human experts cannot perceive a COVID-specific signature in LUS with high specificity, this is likely a more technically difficult task than distinguishing ILD and COPD from healthy patients. Indeed, there is ample evidence on the visible signs of ILD on LUS [59].

Statistical analysis plan

For descriptive statistics related to the clinical data collected, all continuous variables will be reported as medians with their interquartile ranges. Binary and categorical variables distribution will be reported as proportions and percentages. To evaluate baseline demographic differences and outcomes differences between the case and control patients, conditional logistic regressions will be used to account for the matched design. Pearson's and Spearman's correlation coefficients will be used to assess the relationship between continuous variables normally and non-normally distributed, respectively. For the primary outcome, each task will be quantified using descriptive statistics (i.e., proportion and type of abnormalities), as well as the AUROC, sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratios (with 95% CIs over a fivefold cross-validation).

The diagnostic accuracy of each echographic sign will be assessed and sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio with their 95% CIs will be calculated. To find the combination of echographic signs with the best diagnostic accuracy, we will compare the performance of several multivariable models, such as logistic regression, random forest and neural networks. Performance will be reported on a test set comprising 20% of the data in a tenfold cross-validation with 95% CIs. As a secondary objective, we will aim to compare the predictive performance of human expert-identified acoustic signatures in the above predictive tasks. First, we will describe the expert labels by the percentage of sound labels attributed to each diagnosis. A multivariable logistic regression will be derived using the clinical data and sound labels to estimate the diagnoses, as for the primary objective. A kappa score will be used to assess the concordance between *DeepBreath* and expert diagnosis consolidated into a basic predictive model.

The K-BILD, CAT and SF-36 questionnaires will be analysed with descriptive statistics. Associations between the questionnaires' sum scores and lung function parameters will be quantified by Spearman's rank correlation coefficient. We will consider correlations < 0.3 as negligible, ≥ 0.3 to < 0.5 as low, ≥ 0.5 to 0.7 as moderate, and ≥ 0.7 as strong.

Missing data will be reported and padded with zero in the deep learning network and also assessed according to other labels. Features with more than 50% missing values or with a significant bias in missing data fields will be removed and reported. All statistical tests will be two-sided with a type-I error risk of 5%. Data analysis will be carried out using the latest version of R (R Foundation, Vienna, Austria) for descriptive statistics and statistical tests.

Discussion

Untreated IPF has the worst prognosis among the different forms of ILD, with median survival ranging from 3 to 5 years from diagnosis [60]. Recent studies suggest that if novel anti-fibrotic medications (pirfenidone and nintedanib) are started early, they can slow the rate of lung function decline and prevent IPF exacerbation, thus reducing mortality [6, 7]. Unfortunately, because of the unspecific nature of the symptoms, the early stage of IPF remains underdiagnosed and many patients will progress to advanced disease and may require lung transplantation [8]. By contrast, the prognosis of NSIP is generally better than that of IPF, with a median survival time of more than 9 years. Systemic steroids and immunosuppressive therapy may be attempted to slow or reverse the course of the disease, but non-responsive individuals may also be considered for lung transplantation [8]. When left untreated, NSIP tends to progress toward fibrotic changes and persistent debilitating symptoms. COPD is also a leading cause of disability worldwide. Patients are generally unaware of their condition for years, leading to a significant delay in diagnosis, the application of preventive measures such as a smoking cessation intervention, and potential treatment [61]. Being able to recognise and diagnose these lung diseases earlier is of the utmost clinical importance.

This study will aim to collect a standardised dataset of digital lung auscultations and derive a deep learning model able to detect the acoustic and sonographic signatures of the presence and severity of IPF, NSIP and COPD. Recent advances in deep learning are promising to support doctors in standardising the detection and interpretation of complex patterns in pulmonary diseases and AI has proven to outperform doctors in discriminating respiratory pathologies via respiratory functional explorations [62], symptoms [63, 64] and/or radiological examinations [34]. To overcome the subjectivity of human auscultation and the discrepancy in auscultation ability between doctors [16], the development of AI algorithms for the analysis of respiratory acoustic signals has been proposed [19, 20]. In order to meet this requirement, many attempts have been made to develop and apply neural networks to automate the detection and classification of various disease-related breath sounds using machine learning and deep learning-based analysis [14, 64–66]. In particular, recent literature reviews have summarized advances in the implementation of respiratory sound-based AI algorithms in the screening, diagnosis, and classification of COPD [26, 65]. Conversely, the current state of knowledge on the computerized analysis of breath sounds in patients with ILD using AI techniques has not been assessed. Table 2 summarizes the published studies most similar to our research.

However, there are some notable differences in these studies, which justify the present work. The main ones are the frequent absence of healthy subjects as a control group and the almost unanimous lack of severity classification or joint use of LUS images. As stated by Charleston-Villalobos et al. [68], a comparison with other attempts to diagnose and classify lung sounds is difficult due to the difference in data acquisition, type of classification scheme, lack of gold standards allowing standardization between studies, and their distinct exploratory nature. In particular, a major flaw of most anterior studies aimed at building deep learning models for diagnostic classification from digital lung sounds is the use of publicly available databases, such as the R.A.L.E repository [82] or the International Conference on Biomedical Health Informatics [83]. These databases have inherent acquisition flaws due to heterogeneity in data collection and methods that create systematic biases between the predicted labels on which new algorithms are built. This is then reflected in the results of studies with an exaggerated excellent predictive performance that prevents their evaluation and comparison with each other. On the contrary, in our study, sounds will come from a cohort of patients under standardised and homogeneous recording conditions. It remains to be determined whether an AI algorithm using respiratory sounds and/or LUS analysis can be used as an initial and accurate diagnostic tool for patients with ILDs or COPD. The diagnosis of IPF, NSIP and COPD in early stages may allow practitioners to appropriately recognise exacerbations of a chronic lung disease, whereas patients may initially be diagnosed as having multiple bouts of acute disease (e.g., bronchitis) without this defined diagnosis [61]. Early diagnosis with AI may therefore allow patients to benefit from prevention measures and the allocation of appropriate treatments aiming to reduce the progression to permanent lung damage and improve the overall prognosis in patients presenting at primary care clinics for non-specific chronic respiratory symptoms. As research in this area is scarce, it is anticipated that the results generated from this study will be of great importance and may be sufficient to change and improve pulmonary primary care practice in a vulnerable population by proposing a faster diagnosis.

This study has several limitations. First, the interpretation and data generated by the algorithm at this stage of our research will not be used for diagnostic purposes or treatment decisions. Both of these points will require further dedicated validation studies in clinical contexts. Second, selection bias can occur in case-control studies when control subjects are not truly representative of the population that produces the cases. In this study, both populations will stem from the same source population

Table 2 Overview on computerized respiratory sound analysis in ILDs using AI techniques

Related works	Sample size	Purpose	Features	Performance	Issues
Aykanat et al. [67]	40 IPF, 211 COPD, and 574 other pulmonary single or mixed conditions, 805 healthy subjects	Binary classification (healthy vs pathological) and 12-class lung disease classification	SVM, k-NN, GB	Performance for binary classification: Acc 88% to 92% Se 85% to 92% Sp 85% to 88% Performance for lung disease classification: Acc 43% to 68% Se 53% to 96%	Severity classification not done. LUS not used
Charleston-Villalobos et al. [68]	19 ILD (12 IPF, 7 extrinsic allergic alveolitis), 8 healthy subjects	Binary classification (healthy subjects vs patients with ILD)	AR model, SNN	Performance evaluation of the neural network: Acc 76.9% to 98.8%, Se 80.4% to 100%, Sp 73.3% to 100%	Severity classification not done. LUS not used
Fletstra et al. [69]	39 IPF, 95 CHF, 123 PN	Binary classification (IPF vs CHF and IPF vs PN)	BPNN, SVM	Acoustic properties of fine crackles of IPF help distinguish them from crackles of CHF and PN	No healthy subjects. Severity classification not done. LUS not used
Fukumitsu et al. [70]	71 ILD (24 honeycombing +, 47 honeycombing-)	Predicting honeycombing on HRCT by the acoustic properties of fine crackles	FFT	Acoustic properties of fine crackles distinguish the honeycombing from the non-honeycombing group	No healthy subjects. Severity classification not done. LUS not used
Horimasu et al. [71]	34 ILD, 8 COPD or asthma, 7 lung tumor, 5 lung nodule, 6 other	Comparisons of machine-learning-based quantification of four types of lung sounds between lung fields with and without ILD in HRCT and chest X-ray	Described in [72]	AUROC 0.855, Acc 75%, Se 76.1%, Sp 73.6%	Severity classification not done. Results suffered from the presence of background noise. LUS not used
Kahya et al. [73]	23 ILD, 28 COPD, 18 healthy	Binary classification (healthy vs pathological)	AR model, k-NN	Acc 71.1%	Severity classification not done. LUS not used
Kim et al. [74]	112 ILD, 211 COPD, 497 other pulmonary conditions, 51 healthy	Binary classification (healthy vs pathological) and three-class (crackles vs wheezes vs rhonchi) classification	CNN	Performance for binary classification: AUROC 0.93 Acc 86.5% Performance for abnormal sounds classification: AUROC 0.92 Acc 85.7%	Severity classification not done. LUS not used
Malmberg et al. [75]	8 fibrosing alveolitis, 8 emphysema, 8 asthma, 8 healthy subjects	Diagnosis agreement between clinical and machine-learning-based classification of lung sounds	FFT, SOM	Kappa for fibrosing alveolitis 0.54	Small number of patients. Severity classification not done. LUS not used
Manfredi et al. [76]	98 CTD patients (42 ILD +, 56 ILD-)	Identifying CTD patients with possible ILD using lung sound-based binary classification (presence vs absence of ILD) compared with chest HRCT	FFT	Acc 82.6%, Se 88.1%, Sp 78.6%	No healthy subjects. Severity classification not done. LUS not used

Table 2 (continued)

Related works	Sample size	Purpose	Features	Performance	Issues
Manfredi et al. [38]	137 RA patients (59 ILD +, 78 ILD-)	Identifying RA patients with possible ILD using lung sound-based binary classification (presence vs absence of ILD) compared with chest HRCT	FFT	Overall performance: Acc 83.9%, Se 93.2%, Sp 76.9% Performance for velcro-like crackles detection: Acc 67.2%, Se 69.1%, Sp 65.7% F-score 92.4% Se 85.9%	No healthy subjects. Severity classification not done. LUS not used
Messner et al. [77]	7 IPF, 16 healthy subjects	Binary classification (healthy subjects vs patients with IPF)	CRNN	F-score 72.1% Se 71.5%	Small number of patients. Manual crackles labelling required. Severity classification not done. LUS not used
Messner et al. [78]	5 IPF, 10 healthy subjects	Binary classification (healthy subjects vs patients with IPF)	GRNN	Spectral analysis of lung sounds is useful in the diagnosis and evaluation of the severity of IPN	Small number of patients. Manual crackles labelling required. Severity classification not done. LUS not used
Ono et al. [79]	21 IPN, 10 healthy subjects	IPN detectability and severity classification	FFT	Acc 90%, Se 92.6%, Sp 88.4%	LUS not used
Pancaldi et al. [17]	70 RA patients (27 ILD +, 43 ILD-)	Identifying RA patients with possible ILD using lung sound-based binary classification (presence vs absence of ILD) compared with chest HRCT	FFT	Performance evaluation of the neural network: Acc 90.7% to 97.3%, Se 91.8% to 98.3%, Sp 87.5% to 96.3%	No healthy subjects. Severity classification not done. LUS not used
Santiago-Fuentes et al. [80]	19 ILD (10 IPF, 9 CPFES)	Binary classification (IPF vs CPFES)	AR model, SNN	Se 85%, Sp 85% Se 100% (three-class classifier)	No healthy subjects. Severity classification not done. LUS not used
Sen et al. [81]	10 obstructive bronchiectasis, 10 ILD, 20 healthy subjects	Binary (healthy vs pathological) and three class (healthy, bronchiectasis and ILD) classifiers	AR model, SVM		Small number of patients. Severity classification not done. LUS not used

ACC, accuracy; AR, autoregressive; AUROC, area under the receiver-operating curve; BPNN, backpropagation neural networks; CHF, congestive heart failure; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; CPFES, combined pulmonary fibrosis with emphysema syndrome; CTD, connective tissue diseases; CRNN, convolutional recurrent neural network; FFT, fast Fourier transform; GB, Gaussian Bayes; GRNN, gated recurrent neural network; HRCT, high-resolution computed tomography; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; IPN, interstitial pneumonia; k-NN, k-nearest neighbor; LUS, lung ultrasound; MPNN, message passing neural network; PN, pneumonia; RA, rheumatoid arthritis; Se, sensitivity; SNN, supervised neural network; SOM, self-organizing map; Sp, specificity; SVM, support vector machine

in a single-centre outpatient clinic, which may suggest more acute symptoms and pathological lung sounds than those encountered in ambulatory care services. Third, since patients with already-diagnosed IPF, NSIP and COPD will be enrolled, we will not be able to confirm whether *DeepBreath* would have detected these patients at earlier stages. Finally, we acknowledge that the sample size is modest, but it appears to be sufficiently powered in the context of a pilot study.

Conclusion

The *DeepBreath* model could offer a robust, promising and realistic predictive potential of deep learning to be used as a decision support system by health specialists to better guide clinical care management by exploring the synergistic value of digital lung auscultation and ultrasonography for the automated detection and differential diagnosis of ILD and COPD and to estimate severity. This could be the next frontier in the early diagnosis of COPD and ILD to help improve patient outcomes and quality of life. Furthermore, this study may pave the way for future research based on non-invasive AI models combining point-of-care techniques already commonly used in clinical practice for application to other pulmonary pathologies or even to decentralised care in low-resource settings.

Abbreviations

AI	Artificial intelligence
AV	Alveolar volume
AUROC	Area under the receiver operating characteristic curve
CAT	COPD assessment test
CI	Confidence intervals
COPD	Chronic obstructive pulmonary disease
CT	Computerised tomography
FEV ₁	Forced expiratory volume in 1 s
FCV	Forced vital capacity
FRC	Functional respiratory capacity
GDPR	General Data Protection Regulation
HRCT	High resolution chest CT
IIP	Idiopathic interstitial pneumonia
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
K-BILD	King's Brief Interstitial Lung Disease Questionnaire
K _{CO}	Carbon monoxide transfer coefficient
LSTM	Lung short-term memory
LUS	Lung ultrasound
NSIP	Non-specific interstitial pneumonia
STROBE	Strengthening the Reporting of Observational studies in Epidemiology
SVM	Support vector machine
TLC	Total lung capacity
TL _{CO}	Transfer factor for carbon monoxide

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12890-022-02255-w>.

Additional file 1. Self-administered questionnaire on demographic characteristics (occupation, long-term exposure to occupational or environmental agents, etc.), relevant medical history and symptom presentation.

Acknowledgements

This study was supported by the Promotion Santé Valais, Ligue Pulmonaire Valaisanne. We gratefully acknowledge the support of Rosemary Sudan for editorial assistance.

Author contributions

JNS is the study coordinator, co-conceived the study, led the design, and drafted the protocol. He will contribute to the study analyses and will write the final manuscript. MAH contributed to the study design, drafted the protocol, is responsible for the development of the deep learning algorithms, and will contribute to statistical analyses. JD contributed to the development of the deep learning algorithms. DSC contributed to the design and will perform statistical analyses. MS will be the research nurse, enrol the patients and collect the data. LR assisted in drafting the protocol, contributed to the design, will be the research doctor, include clinical data at enrolment, enrol patients, collect the data and implement the research. CBA, JD, MS, LR and AG contributed to the critical review of the protocol. POB will be the principal investigator. He co-conceived the study, contributed to the study design, and oversaw the drafting of the protocol. All authors significantly contributed to the development of the protocol, had full access to the manuscript and approved submission of the manuscript to *BMC Pulmonary Medicine*. All authors read and approved the final manuscript.

Funding

The present study has financial support from the Promotion Santé Valais, Ligue Pulmonaire Valaisanne, a branch of the Swiss Pulmonary League. The funders of the study had no role in study design and will have no role in data collection, data analysis, data interpretation, or writing of the report. This study is not commercially funded.

Availability of data and materials

All pertinent data generated or analysed during this study will be included in the published article (and its supplementary information files). The audio used in the study will not be publicly available to protect participant privacy. An anonymous copy of the final (anonymised) datasets used and/or analysed during the current study will be available from the corresponding author on reasonable request upon approval of a proposal and with a signed data access agreement. Data will be made available for a specified research purpose to qualified external researchers whose proposed use of the data has been approved by their institutional review board. The request proposal must include a statistician. Data will be available beginning 6 months and ending 5 years following article publication. The full code and test sets will be made available on publication at the GitHub repository (<https://github.com>).

Declarations

Ethics approval and consent to participate

The study was approved by SwissEthics and the Vaud Cantonal Ethics Committee on August 4 2022 (study reference PB_2021-01688). This protocol is adapted for publication. The study will be conducted according to the principles of the Declaration of Helsinki [84], and Good Clinical Practice [85]. All patients will be required to provide written informed consent. It is our intention to present the results at scientific congresses and to publish the results in an international peer-reviewed journal, irrespective of the magnitude or direction of effect. This protocol adheres to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) Statement [41]. Information about study subjects will be kept confidential. All data will be entered into the REDCap data management system. Hard copy records will be stored in a locked cabinet in a secure location. Access to records and data will be limited to study personnel. Study data will be de-identified and a master linking log with identifiers will be kept and stored separately from the data.

Consent for publication

Not applicable.

Competing interests

AG intend to develop a smart stethoscope 'Onescope', which may be commercialised. All other authors declare that they have no competing interests.

Author details

¹Division of Paediatric Emergency Medicine, Department of Women, Child and Adolescent, Geneva University Hospitals, 47 Avenue de la Roseraie, 1211 Geneva 14, Switzerland. ²Machine Learning and Optimization (MLO) Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. ³Quality of Care Unit, Geneva University Hospitals, Geneva, Switzerland. ⁴Division of Pulmonology, Hospital of Valais, Sion, Switzerland. ⁵Division of Paediatric Pulmonology, Department of Women, Child and Adolescent, Geneva University Hospitals, Geneva, Switzerland. ⁶Faculty of Medicine, University of Geneva, Geneva, Switzerland.

Received: 3 September 2022 Accepted: 20 November 2022
Published online: 02 June 2023

References

- Raghu G, Nyberg F, Morgan G. The epidemiology of interstitial lung disease and its association with lung cancer. *Br J Cancer*. 2004;91(Suppl 2):S3-10.
- Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2013;188(6):733-48.
- Lamas DJ, Kawut SM, Bagiella E, Philip N, Arcasoy SM, Lederer DJ. Delayed access and survival in idiopathic pulmonary fibrosis: a cohort study. *Am J Respir Crit Care Med*. 2011;184(7):842-7.
- Sellares J, Hernandez-Gonzalez F, Lucena CM, Paradela M, Brito-Zeron P, Prieto-Gonzalez S, Benegas M, Cuerpo S, Espinosa G, Ramirez J, et al. Auscultation of velcro crackles is associated with usual interstitial pneumonia. *Medicine (Baltimore)*. 2016;95(5): e2573.
- Mink SN, Maycher B. Comparative manifestations and diagnostic accuracy of high-resolution computed tomography in usual interstitial pneumonia and nonspecific interstitial pneumonia. *Curr Opin Pulm Med*. 2012;18(5):530-4.
- Nathan SD, Albera C, Bradford WZ, Costabel U, Glaspole I, Glassberg MK, Kardatzke DR, Daigl M, Kirchgassler KU, Lancaster LH, et al. Effect of pirfenidone on mortality: pooled analyses and meta-analyses of clinical trials in idiopathic pulmonary fibrosis. *Lancet Respir Med*. 2017;5(1):33-41.
- Richeldi L, Cottin V, du Bois RM, Selman M, Kimura T, Bailes Z, Schlenker-Herceg R, Stowasser S, Brown KK. Nintedanib in patients with idiopathic pulmonary fibrosis: Combined evidence from the TOMORROW and INPULSIS(R) trials. *Respir Med*. 2016;113:74-9.
- Brown AW, Kaya H, Nathan SD. Lung transplantation in IPF: a review. *Respirology*. 2016;21(7):1173-84.
- Rivera-Ortega P, Molina-Molina M. Interstitial lung diseases in developing countries. *Ann Glob Health*. 2019;85(1):2414.
- Antoniou KM, Symvoulakis EK, Margaritopoulos GA, Lionis C, Wells AU. Early diagnosis of IPF: time for a primary-care case-finding initiative? *Lancet Respir Med*. 2014;2(1): e1.
- Cordier JF, Cottin V. Neglected evidence in idiopathic pulmonary fibrosis: from history to earlier diagnosis. *Eur Respir J*. 2013;42(4):916-23.
- Cottin V, Cordier JF. Velcro crackles: the key for early diagnosis of idiopathic pulmonary fibrosis? *Eur Respir J*. 2012;40(3):519-21.
- Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. *Ann Thorac Med*. 2015;10(3):158-68.
- Andres E, Gass R, Charolux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J Med Life*. 2018;11(2):89-106.
- Pinho C, Oliveira A, Jácome C, Rodrigues JM, Marques A. Integrated approach for automatic crackle detection based on fractal dimension and box filtering. In: *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications*. edn. Edited by Global I; 2020: 815-832.
- Gurung A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med*. 2011;105(9):1396-403.
- Pancaldi F, Sebastiani M, Cassone G, Luppi F, Cerri S, Della Casa G, Manfredi A. Analysis of pulmonary sounds for the diagnosis of interstitial lung diseases secondary to rheumatoid arthritis. *Comput Biol Med*. 2018;96:91-7.
- Abbas A, Fahim A. An automated computerized auscultation and diagnostic system for pulmonary diseases. *J Med Syst*. 2010;34(6):1149-55.
- Grzywalski T, Piecuch M, Szajek M, Breborowicz A, Hafke-Dys H, Kocinski J, Pastusiak A, Belluzzo R. Practical implementation of artificial intelligence algorithms in pulmonary auscultation examination. *Eur J Pediatr*. 2019;178(6):883-90.
- Palaniappan R, Sundaraj K, Sundaraj S. Artificial intelligence techniques used in respiratory sound analysis—a systematic review. *Biomed Tech (Berl)*. 2014;59(1):7-18.
- Bhatt SP, Washko GR, Hoffman EA, Newell JD Jr, Bodduluri S, Diaz AA, Galban CJ, Silverman EK, San Jose Estepar R, Lynch DA. Imaging advances in chronic obstructive pulmonary disease. Insights from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) study. *Am J Respir Crit Care Med*. 2019;199(3):286-301.
- Das N, Topalovic M, Janssens W. Artificial intelligence in diagnosis of obstructive lung disease: current status and future potential. *Curr Opin Pulm Med*. 2018;24(2):117-23.
- Mlodzinski E, Stone DJ, Celi LA. Machine learning for pulmonary and critical care medicine: a narrative review. *Pulm Ther*. 2020.
- Forum of International Respiratory Societies: The global impact of respiratory disease. In., Second Edition edn. Sheffield: European Respiratory Society; 2017.
- De Ramon FA, Ruiz Fernandez D, Gilart Iglesias V, Marcos Jorquera D. Analyzing the use of artificial intelligence for the management of chronic obstructive pulmonary disease (COPD). *Int J Med Inform*. 2021;158: 104640.
- Feng Y, Wang Y, Zeng C, Mao H. Artificial intelligence and machine Learning in chronic airway diseases: focus on asthma and chronic obstructive pulmonary disease. *Int J Med Sci*. 2021;18(13):2871-89.
- Exarchos K, Aggelopoulou A, Oikonomou A, Biniskou T, Beli V, Antoniadou E, Kostikas K. Review of artificial intelligence techniques in chronic obstructive lung disease. *IEEE J Biomed Health Inform*. 2022;26(5):2331-8.
- Nikolaou V, Massaro S, Fakhimi M, Stergioulas L, Price D. COPD phenotypes and machine learning cluster analysis: a systematic review and future research agenda. *Respir Med*. 2020;171: 106093.
- Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med*. 2020;14(6):559-64.
- Altan G, Kutlu Y, Allahverdi N. Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE J Biomed Health Inform*. 2019;24(5):1344-50.
- Altan G, Kutlu Y, Pekmezci AÖ, Nural S. Deep learning with 3D-second order difference plot on respiratory sounds. *Biomed Signal Process Control*. 2018;45:58-69.
- Abe H, Ashizawa K, Li F, Matsuyama N, Fukushima A, Shiraishi J, MacMahon H, Doi K. Artificial neural networks (ANNs) for differential diagnosis of interstitial lung disease: results of a simulation test with actual clinical cases. *Acad Radiol*. 2004;11(1):29-37.
- Fukushima A, Ashizawa K, Yamaguchi T, Matsuyama N, Hayashi H, Kida I, Imafuku Y, Egawa A, Kimura S, Nagaoki K, et al. Application of an artificial neural network to high-resolution CT: usefulness in differential diagnosis of diffuse lung disease. *Am J Roentgenol*. 2004;183(2):297-305.
- Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med*. 2018;6(11):837-45.
- Kim SY, Diggins J, Pankratz D, Huang J, Pagan M, Sindy N, Tom E, Anderson J, Choi Y, Lynch DA, et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med*. 2015;3(6):473-82.
- Pankratz DG, Choi Y, Imtiaz U, Fedorowicz GM, Anderson JD, Colby TV, Myers JL, Lynch DA, Brown KK, Flaherty KR, et al. Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Ann Am Thorac Soc*. 2017;14(11):1646-54.
- Topalovic M, Laval S, Aerts JM, Troosters T, Decramer M, Janssens W. Belgian Pulmonary Function Study i: Automated interpretation of pulmonary function tests in adults with respiratory complaints. *Respiration*. 2017;93(3):170-8.
- Manfredi A, Cassone G, Cerri S, Venerito V, Fedele AL, Trevisani M, Furini F, Addimanda O, Pancaldi F, Della Casa G, et al. Diagnostic accuracy of a

- velcro sound detector (VECTOR) for interstitial lung disease in rheumatoid arthritis patients: the InSPIRAte validation study (InTerstitial pneumonia in rheumatoid Arthritis with an electronic device). *BMC Pulm Med.* 2019;19(1):111.
39. Arntfeld R, Wu D, Tschirhart J, VanBerlo B, Ford A, Ho J, McCauley J, Wu B, Deglint J, Chaudhary R, et al. Automation of lung ultrasound interpretation via deep learning for the classification of normal versus abnormal lung parenchyma: a multicenter study. *Diagnostics (Basel).* 2021;11(11):2049.
 40. Glangetas A, Hartley MA, Cantais A, Courvoisier DS, Rivollet D, Shama DM, Perez A, Spechbach H, Trombert V, Bourquin S, et al. Deep learning diagnostic and risk-stratification pattern detection for COVID-19 in digital lung auscultations: clinical protocol for a case-control and prospective cohort study. *BMC Pulm Med.* 2021;21(1):103.
 41. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, Initiative S. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453–7.
 42. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, Goldin JG, Hansell DM, Inoue Y, Johkoh T, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society white paper. *Lancet Respir Med.* 2018;6(2):138–53.
 43. Travis WD, Hunninghake G, King TE Jr, Lynch DA, Colby TV, Galvin JR, Brown KK, Chung MP, Cordier JF, du Bois RM, et al. Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med.* 2008;177(12):1338–47.
 44. Global Initiative for Chronic Obstructive Lung Disease, Inc.: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2020 report. In.; 2020.
 45. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med.* 2018;198(5):e44–68.
 46. Sgalla G, Walsh SLF, Sverzellati N, Fletcher S, Cerri S, Dimitrov B, Nikolic D, Barney A, Pancaldi F, Larcher L, et al. "Velcro-type" crackles predict specific radiologic features of fibrotic interstitial lung disease. *BMC Pulm Med.* 2018;18(1):103.
 47. Patel AS, Siegert RJ, Brignall K, Gordon P, Steer S, Desai SR, Maher TM, Renzoni EA, Wells AU, Higgins IJ, et al. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax.* 2012;67(9):804–10.
 48. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J.* 2009;34(3):648–54.
 49. Ware JE Jr. SF-36 health survey update. *Spine.* 2000;25(24):3130–9.
 50. Brahier T, Meuwly JY, Pantet O, BrochuVez MJ, Gerhard Donnet H, Hartley MA, Hugli O, Boillat-Blanco N. Lung ultrasonography for risk stratification in patients with coronavirus disease 2019 (COVID-19): a prospective observational cohort study. *Clin Infect Dis.* 2021;73(11):e4189–96.
 51. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med.* 2014;370(8):744–51.
 52. Chavez MA, Shams N, Ellington LE, Naithani N, Gilman RH, Steinhoff MC, Santosham M, Black RE, Price C, Gross M, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. *Respir Res.* 2014;15:50.
 53. Sorlini C, Femia M, Nattino G, Bellone P, Gesu E, Francione P, Paterno M, Grillo P, Ruffino A, Bertolini G, et al. The role of lung ultrasound as a front-line diagnostic tool in the era of COVID-19 outbreak. *Intern Emerg Med.* 2021;16(3):749–56.
 54. Mayo PH, Copetti R, Feller-Kopman D, Mathis G, Maury E, Mongodi S, Mojoli F, Volpicelli G, Zanobetti M. Thoracic ultrasonography: a narrative review. *Intensive Care Med.* 2019;45(9):1200–11.
 55. Tomassetti S, Ryu JH, Piciucchi S, Chilosi M, Poletti V. Nonspecific interstitial pneumonia: what is the optimal approach to management? *Semin Respir Crit Care Med.* 2016;37(3):378–94.
 56. Lowe KE, Regan EA, Anzueto A, Austin E, Austin JHM, Beaty TH, Benos PV, Benway CJ, Bhatt SP, Bleecker ER, et al. COPDGene(R) 2019: redefining the diagnosis of chronic obstructive pulmonary disease. *Chronic Obstr Pulm Dis.* 2019;6(5):384–99.
 57. White P, Myers M. The classification of cardiac diagnosis. *JAMA.* 1921;77(18):1414–5.
 58. Murphy RL. In defense of the stethoscope. *Respir Care.* 2008;53(3):355–69.
 59. Manolescu D, Davidescu L, Traila D, Oancea C, Tudorache V. The reliability of lung ultrasound in assessment of idiopathic pulmonary fibrosis. *Clin Interv Aging.* 2018;13:437–49.
 60. Oldham JM, Collard HR. Comorbid conditions in idiopathic pulmonary fibrosis: recognition and management. *Front Med (Lausanne).* 2017;4:123.
 61. Csiksz NG, Gartman EJ. New developments in the assessment of COPD: early diagnosis is key. *Int J Chron Obstruct Pulmon Dis.* 2014;9:277–86.
 62. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, Janssen R, Kerstjens HAM, Liistro G, Louis R, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J.* 2019;53(4):1801660.
 63. Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. *Artif Intell Med.* 2018;88:58–69.
 64. Chamberlain D, Kodgule R, Ganelin D, Miglani V, Fletcher RR. Application of semi-supervised deep learning to lung sound analysis. *Conf Proc IEEE Eng Med Biol Soc.* 2016;2016:804–7.
 65. Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: a systematic review. *PLoS ONE.* 2017;12(5):e0177926.
 66. Kim Y, Hyon Y, Lee S, Woo SD, Ha T, Chung C. The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulm Med.* 2022;22(1):119.
 67. Aykanat M, KILIÇ Ö, Bahar K, SARYAL SB. Lung disease classification using machine learning algorithms. *Int J Appl Math Electron Comput.* 2020;8(4):125–32.
 68. Charleston-Villalobos S, Martinez-Hernandez G, Gonzalez-Camarena R, Chi-Lem G, Carrillo JG, Aljama-Corrales T. Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. *Comput Biol Med.* 2011;41(7):473–82.
 69. Flietstra B, Markuzon N, Vyshedskiy A, Murphy R. Automated analysis of crackles in patients with interstitial pulmonary fibrosis. *Pulm Med.* 2011;2011: 590506.
 70. Fukumitsu T, Obase Y, Ishimatsu Y, Nakashima S, Ishimoto H, Sakamoto N, Nishitsuji K, Shiwa S, Sakai T, Miyahara S, et al. The acoustic characteristics of fine crackles predict honeycombing on high-resolution computed tomography. *BMC Pulm Med.* 2019;19(1):153.
 71. Horimasu Y, Ohshimo S, Yamaguchi K, Sakamoto S, Masuda T, Nakashima T, Miyamoto S, Iwamoto H, Fujitaka K, Hamada H, et al. A machine-learning based approach to quantify fine crackles in the diagnosis of interstitial pneumonia: a proof-of-concept study. *Medicine (Baltimore).* 2021;100(7): e24738.
 72. Ohshimo S, Sadamori T, Tanigawa K. Innovation in analysis of respiratory sounds. *Ann Intern Med.* 2016;164(9):638–9.
 73. Kahya YP, Guler EC, Sahin S. Respiratory disease diagnosis using lung sounds. In: *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society Magnificent Milestones and Emerging Opportunities in Medical Engineering (Cat No 97CH36136): 1997: IEEE; 1997: 2051–2053.*
 74. Kim Y, Hyon Y, Jung SS, Lee S, Yoo G, Chung C, Ha T. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Sci Rep.* 2021;11(1):17186.
 75. Malmberg LP, Kallio K, Haltsonen S, Katila T, Sovijarvi AR. Classification of lung sounds in patients with asthma, emphysema, fibrosing alveolitis and healthy lungs by using self-organizing maps. *Clin Physiol.* 1996;16(2):115–29.
 76. Manfredi A, Cassone G, Vacchi C, Pancaldi F, Della Casa G, Cerri S, De Pasquale L, Luppi F, Salvarani C, Sebastiani M. Usefulness of digital velcro crackles detection in identification of interstitial lung disease in patients with connective tissue diseases. *Arch Rheumatol.* 2021;36(1):19–25.
 77. Messner E, Fediuk M, Swatek P, Scheidl S, Smolle-Juttner FM, Olschewski H, Pernkopf F. Multi-channel lung sound classification with convolutional recurrent neural networks. *Comput Biol Med.* 2020;122: 103831.
 78. Messner E, Fediuk M, Swatek P, Scheidl S, Smolle-Juttner FM, Olschewski H, Pernkopf F. Crackle and breathing phase detection in lung sounds with deep bidirectional gated recurrent neural networks. *Annu Int Conf IEEE Eng Med Biol Soc.* 2018;2018:356–9.
 79. Ono H, Taniguchi Y, Shinoda K, Sakamoto T, Kudoh S, Gemma A. Evaluation of the usefulness of spectral analysis of inspiratory lung sounds recorded with phonopneumography in patients with interstitial pneumonia. *J Nippon Med Sch.* 2009;76(2):67–75.

80. Santiago-Fuentes LM, Charleston-Villalobos S, Gonzalez-Camarena R, Mejia-Avila M, Mateos-Toledo H, Buendia-Roldan I, Aljama-Corrales T. A multichannel acoustic approach to define a pulmonary pathology as combined pulmonary fibrosis and emphysema syndrome. *Annu Int Conf IEEE Eng Med Biol Soc.* 2017;2017:2757–60.
81. Sen I, Saraclar M, Kahya YP. Computerized diagnosis of respiratory disorders. SVM based classification of VAR model parameters of respiratory sounds. *Methods Inf Med.* 2014;53(4):291–5.
82. Owens D. RALE lung sounds 3.0. *Comput Inform Nurs.* 2002;5(3):9–10.
83. Rocha BM, Filos D, Mendes L, Serbes G, Ulukaya S, Kahya YP, Jakovljevic N, Turukalo TL, Vogiatzis IM, Perantoni E, et al. An open access database for the evaluation of respiratory sound classification algorithms. *Physiol Meas.* 2019;40(3): 035001.
84. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191–4.
85. International Conference on Harmonisation: Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med* 1999;18(15):1905–1942.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

