On the Validation of Computerised Lung Auscultation

Guilherme Campos[1,2] and João Quintas[3]

[1]*Departmento de Electrónica, Telecomunicações e Informática (DETI) – Universidade de Aveiro (UA),*
Campus de Santiago, 3810-193 Aveiro, Portugal
[2]*Instituto de Engenharia Electrónica e Informática de Aveiro (IEETA) – Universidade de Aveiro (UA),*
Campus de Santiago, 3810-193 Aveiro, Portugal
[3]*Instituto de Sistemas e Robótica (ISR) – Instituto Superior Técnico (IST),*
Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal

Keywords: Adventitious Lung Sounds, Automatic Detection Algorithms, Annotation, Agreement, Performance Metrics, Validation.

Abstract: The development of computerised diagnosis tools based on lung auscultation necessitates appropriate validation. So far, this work front has received insufficient attention from researchers; validation studies found in the literature are largely flawed. We believe that building open-access crowd-sourced information systems based on large-scale repositories of respiratory sound files is an essential task and should be urgently addressed. Most diagnosis tools are based on automatic adventitious lung sound (ALS) detection algorithms. The gold standards required to assess their performance can only be obtained by human expert annotation of a statistically significant set of respiratory sound files; given the inevitable subjectivity of the process, statistical agreement criteria must be applied to multiple independent annotations obtained for each file. For these reasons, the information systems we propose should provide simple, efficient annotation tools; facilitate the formation of credible annotation panels; apply appropriate agreement criteria and metrics to generate gold-standard ALS annotation files and, based on them, allow easy quantitative assessment of detection algorithm performance.

1 INTRODUCTION

Easy, inexpensive and non-invasive, auscultation is an age-old medical diagnosis method. The stethoscope is a tribute to its paramount importance: invented by Laënnec in 1816, it has become the most universal symbol of the medical profession.

Diagnosing respiratory conditions through lung auscultation is a skill healthcare practitioners acquire by training. As shown in the diagram of Figure 1, the process can be decomposed into two steps.

The first is a sound analysis stage, based on the notion of *normal* respiratory sounds and the ability to identify *abnormal* features superimposed on them, also called *adventitious lung sounds* (ALS). ALSs are classified into various types according to their acoustic characteristics. Classification criteria and nomenclatures adopted in the literature may differ slightly, as there is no universal standardisation; for instance, Bohadana et al. (2014) list *stridors*, *wheezes*, *rhonchi*, *fine crackles*, *coarse crackles*,

pleural friction rubs and *squawks*. Different sets of clinical correlations have been established for each ALS type.
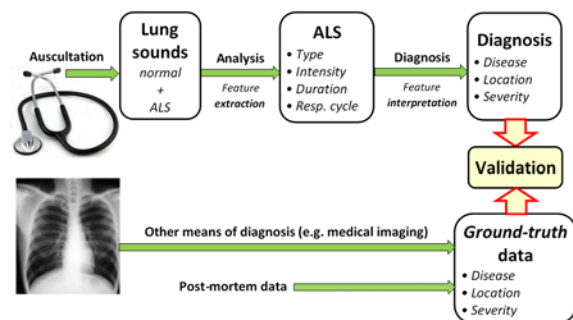

Figure 1: Lung disease diagnosis based on auscultation.

Based on this accumulated knowledge, the second step – diagnosis proper – consists in interpreting the characteristics (type, intensity, duration, instant of occurrence within the respiratory cycle…) of the ALS observed in different auscultation points in order to

establish the disease, its severity and area affected. As Figure 1 suggests, the results can only be validated against *ground-truth* data obtained through more reliable diagnosis means (e.g. medical imaging) or post-mortem examination.

2 AUTOMATIC *ALS* DETECTION

Carried out in the traditional guise (i.e. by humans), and despite constant progress towards standardisation and sophistication of auscultation training methods and technology (see, for instance, Ward and Wattier's 2011 review), the signal analysis process depicted in Figure 1 is rather subjective; obviously, it is also restricted to the human audible frequency range.

Computer-aided auscultation is potentially much more objective, reliable and efficient. With the advent of digital stethoscopy, its development became a real prospect (reflected, for example, in the 1997 review by Pasterkamp et al.). The EU-funded project *Computerised Respiratory Sound Analysis* (CORSA), involving a multinational task force of the European Respiratory Society (Sovijärvi et al. 2000), marked a research boom in this area. Naturally inspired by the human auscultation process, depicted in Figure 1, research efforts were primarily directed at automating its first step – ALS detection.

The literature evidences intense work on the development of algorithms applying pattern recognition techniques to detect and classify the various ALS types. Taking the example of crackle detection (arguably the most important and certainly one of the most challenging, given the discontinuous, non-stationary nature of crackles), a wide variety of signal processing techniques have been proposed, including digital filters (Ono et al. 1989), spectrogram analysis (Kaisla et al. 1991), auto-regressive models (Hadjileontiadis 1996), time-domain analysis (Vannuccini et al. 1998), fuzzy filters (Mastorocostas et al. 2000), wavelet and wavelet-packet transform methods (Kahya et al. 2001; Hadjileontiadis 2005; Lu and Bahoura 2006; Lu and Bahoura 2008), fractal dimension (FD) filtering (Hadjileontiadis and Rekanos 2003), Hilbert transform analysis (Li and Du 2005) and empirical mode decomposition (EMD) (Charleston-Villalobos et al. 2007; Hadjileontiadis 2007). This list is by no means exhaustive and similar efforts have gone into the development of detection algorithms for other ALS types, especially wheezes.

However, by and large, research publications in this area reveal serious imbalance between development and validation work, with insufficient attention paid to the latter. To better characterise this

problem and support the practical solution proposed for it in section 4, the next section discusses ALS detection algorithm validation and its specific requirements.

3 VALIDATION ISSUES

ALS waveforms can be characterised qualitatively, but establishing completely objective definitions is not possible (if it were, developing an algorithm with 100% detection accuracy would be a simple task). The performance of automatic ALS detection algorithms can thus only be assessed by comparing the annotations they generate with human expert annotations of the same sound files, as illustrated in Figure 2. In this context, the term *annotation* refers to a complete record of the ALS of a given type occurring in the sound file under analysis.
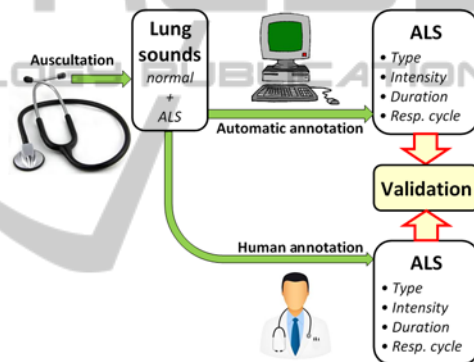

Figure 2: Validation of automatic ALS detection algorithm.

Given the subjectivity of human annotation, pointed out in the previous section, it is essential to take measures to minimise bias. For this reason, validation references should be obtained by combining multiple annotations of the same sound file, each carried out independently by a different human expert, into a single *gold-standard* annotation. The criteria governing this combination or *agreement* process must be explicit. For instance, the pilot study by Quintas et al. (2013) used agreement by majority, but other approaches can and should be explored.

Performance tests reported in the literature are very often based on annotations by a single expert, thus lacking credibility. In the rare instances of multi-annotation, the criteria used for generating gold standards are normally not clarified.

For statistical significance, both the panels of expert annotators and the sets of annotated sound files should be as large and diverse as possible. The development of pattern recognition algorithms often

relies on training; obviously, training and test sets must be separate i.e. performance tests cannot be based on the same files used for training. This constitutes an additional argument in favour of building large, diverse repositories of sound files and corresponding gold-standard annotations, but the repositories actually used in practice tend to be very small and relatively homogeneous.

It is clear from the previous discussion that the availability of complete, reliable and user-friendly computational tools for respiratory sound annotation is essential. The use of open annotation file formats is desirable. The crackle, wheeze and respiratory cycle annotation application RSAS (Dinis et al., 2012) was an effort in this direction. Regrettably, making this kind of tools publicly available is not yet the rule.

In general, replicating the detection algorithm tests described in the literature is virtually impossible, as there is no easy access to the relevant data (sound files and reference annotations). Any performance claims under these circumstances would lack credibility. Since absolute agreement between the annotations used to build a gold standard is extremely unlikely (the small pilot study on multi-annotation presented in Dinis et al. (2012) strongly supports this idea), some extreme performance claims found in the literature may be signs of methodological flaws related to the use of single-annotator data, artificially homogeneous sound repositories (Quintas et al. 2013) or even performance indices measured on training set files.

The creation of a Web-based open information system to stimulate the development and sharing of respiratory sound data and annotation repositories, annotation tools, gold standards, agreement metrics and criteria, as well as detection algorithms, is essential to solve the difficulties discussed and advance research in this area.

4 *ALS* INFORMATION SYSTEM

The information system we propose is outlined in Figure 3. The idea is to base it on an Internet platform and feed it through crowdsourcing i.e. by attracting contributions from the respiratory healthcare community worldwide. This point is emphasised in the figure by the association of the various functional modules with user classes, loosely labelled *managers*, *practitioners*, *annotators*, *developers* and *trainees*.

At the core of this information system lies a repository of lung auscultation sound files obtained through digital stethoscopy. The aim is to make it as expanded and diversified as possible. The online sound file **submission** module must therefore be versatile and user-friendly. It must accommodate multi-channel stethoscopy data.

The records associated with the submitted sound files should be as complete as possible (without compromising patient anonymity), since successful data-mining using the system will depend crucially on access to data on the patient (age, gender, ethnicity, weight, clinical antecedents,…), auscultation conditions (location, equipment, procedures,…) and results from other means of diagnosis (e.g. medical imaging).

Academic research projects may be particularly valuable in building a repository of this kind, inasmuch as they can contribute large-scale data-sets
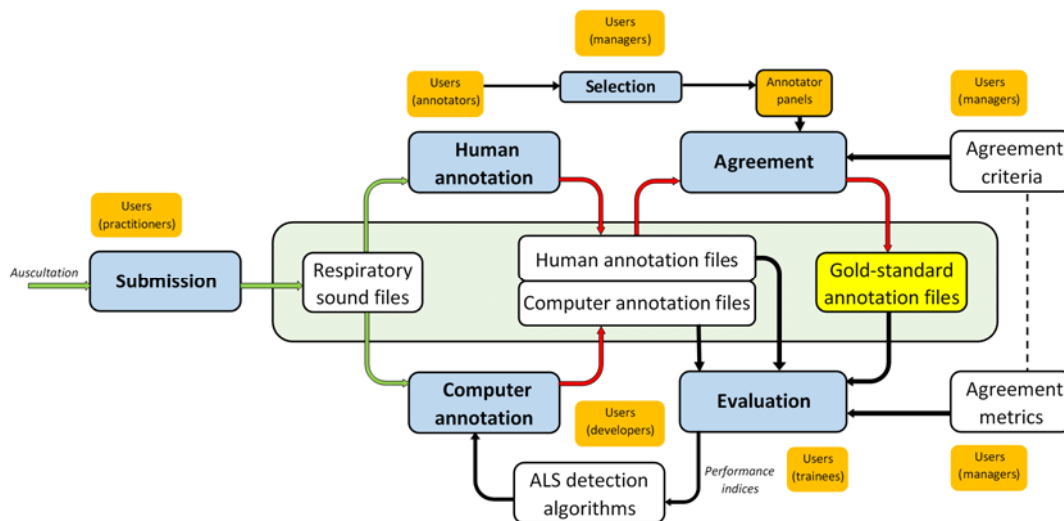

Figure 3: Web-based respiratory sound information system.

obtained under controlled conditions.

It must be possible to define and label sets of sound files within the repository, for the purposes of generating gold standards, training detection algorithms and testing their performance.

An essential tool of this system is the **human annotation** module: a graphical user interface (GUI) along the lines of RSAS (Dinis et al. 2012). It should allow simple, intuitive annotation of any respiratory sound file stored in the repository, the result being a new file (annotation file) tagged to the corresponding sound file/annotator pair and stored in a repository of annotation files. Dinis et al. (2012) propose formats for crackle, wheeze and respiratory cycle annotation files.

Annotating files may be of interest to users of very different levels. For example, the system can assist non-experts (trainees) practice and assess their performance. For the purpose of generating gold-standards, however, it is important to select expert annotator panels from the pool of annotators. As seen in the previous section, the gold standards, generated by the **agreement** module, combine multiple annotations of the same sound file (one per panel member) according to explicit agreement criteria.

The system must, of course, support **computer annotation** through an interface to automatic ALS detection algorithms; these must be able to collect sound files (from test sets or training sets) and submit their corresponding annotations, which must be tagged accordingly and stored in the repository as any other annotation.

The **evaluation** module applies appropriate agreement metrics, consistent with the criteria used for generating the gold standard annotations, to compute detection performance indices. This can be used both on computer annotations (to assist the process of ALS detection algorithm training and validation) and human annotations (to assist the training and assessment of healthcare practitioners).

5 MACHINE LEARNING

ALS detection algorithms are intended to automate the first step of the process outlined in Figure 1, assuming that diagnosis proper will remain a human task. However, with the unceasing progress of computing, signal processing and communication technologies, it is possible to envisage fully automated respiratory disease diagnosis and monitoring systems. This involves automating both the feature extraction and the interpretation steps.

In this scenario, adventitious lung sounds lose importance. Pattern recognition can be applied with no *a priori* restrictions on which features to be considered. This may prove a significant advantage with machine learning techniques such as genetic algorithms, support vector machines or neural networks, as different features (for example in the ultrasound frequency range, completely disregarded by ALS) may contribute to more accurate diagnosis results. In this regard, an analogy may be drawn with music genre classification algorithms, whose performance has improved significantly with the increasing use of machine-selected low-level features with no obvious musical meaning and seemingly unrelated to the human process of musical style identification.

The difficulty of this approach, in this case, is the long validation loop. The intermediate validation of the feature extraction step (see Figure 2) is no longer applicable; the performance of automatic diagnosis algorithms must be directly compared with ground-truth results from other means of diagnosis, as shown in Figure 4.
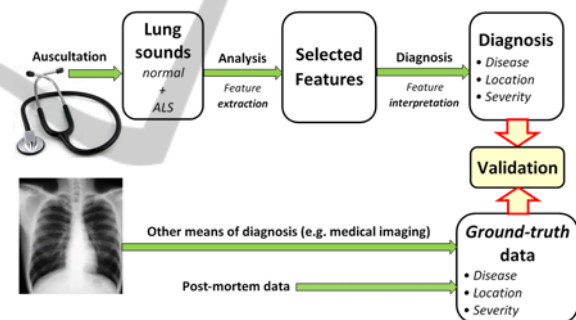

Figure 4: Automatic respiratory disease diagnosis.

This makes it even more indispensable to create an information system with an extensive lung sound repository fed by crowd-sourcing, as described in the previous section; naturally, the modules related to ALS annotation would not be necessary in this approach.

REFERENCES

Bohadana A, Izbicki G, Kraman SS (2014) "Fundamentals of lung auscultation." The New England Journal of Medicine, 370(8): 744-751.

Charleston-Villalobos S, González-Camarena R, Chi-Lem G, Aljama-Corrales T (2007) "Crackle sounds analysis by empirical mode decomposition. Nonlinear and nonstationary signal analysis for distinction of crackles

in lung sounds." IEEE Engineering in Med. and Biology Magazine 26(1): 40-47.

Dinis J, Campos G, Rodrigues J, Marques, A (2012) "Respiratory sound annotation software." Int. Conf. on Health Informatics (*HEALTHINF'12*), 183-188. Vilamoura, Portugal, February 1-4.

Hadjileontiadis LJ (1996) "Nonlinear separation of crackles and squawks from vesicular sounds using third-order statistics." 18th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society 5: 2217-2219.

Hadjileontiadis LJ (2005) "Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding-part I: methodology." IEEE Trans. on Biomedical Engineering 52(6): 1143-1148.

Hadjileontiadis LJ (2007) "Empirical mode decomposition and fractal dimension filter. A novel technique for denoising explosive lung sounds." IEEE Engineering in Med. and Biology Magazine 26(1): 30-39.

Hadjileontiadis LJ, Rekanos T (2003) "Detection of explosive lung and bowel sounds by means of fractal dimension." IEEE Signal Processing Letters 10(10): 311-14.

Kahya YP, Yerer S, Cerid O (2001) "A wavelet-based instrument for detection of crackles in pulmonary sounds." 23rd Annual Int. Conf. of the IEEE Engineering in Med. and Biology Society, 2001. 4: 3175-3178.

Kaisla T, Sovijärvi ARA, Piirilä P, Rajala HM, Haltsonen S, Rosqvist T (1991) "Validated method for automatic detection of lung sound crackles." Medical & biological engineering & computing 29(5): 517-521.

Li Z, Du M (2005) "HHT based lung sound crackle detection and classification." 2005 Int. Symposium on Intelligent Signal Processing and Communication Systems 385-388.

Lu X, Bahoura M (2006) "Separation of crackles form vesicular sounds using wavelet packet transform." Int. Conf. on Acoustics, Speech and Signal Processing (*ICASSP 2006*).

Lu X, Bahoura M (2008) "An integrated automated system for crackles extraction and classification." Biomedical Signal Processing and Control 3(3): 244-254.

Mastorocostas PA, Tolias YA, Theocaris JB, Hadjileontiadis LJ, Panas SM (2000). "An orthogonal least squares-based fuzzy filter for real-time analysis of lung sounds." IEEE Transactions on Bio-medical Engineering 47(9): 1165-76.

Ono M, Arakawa K, Mori M, Sugimoto T, Harashima H (1989). "Separation of fine crackles from vesicular sounds by a nonlinear digital filter." IEEE transactions on bio-medical engineering 36(2): 286-291.

Pasterkamp H, Kraman SS, Wodicka G (1997) "Respiratory Sounds. Advances Beyond the Stethoscope." American Journal of Respiratory and Critical Care Medicine 156(3): 974-87.

Quintas J, Campos G, Marques A (2013) "Multi-algorithm respiratory crackle detection." Int. Conf. on Health Informatics (*HEALTHINF'13*), 239-244. Barcelona, Spain, February 11-14.

Sovijärvi ARA, Vanderschoot J, Earis J E (2000) "Standardization of computerized respiratory sound analysis". European Respiratory Review 10:77, 585.

Vannuccini L, Rossi M, Pasquali G (1998) "A new method to detect crackles in respiratory sounds." Technology and Health Care 6(1): 75-79.

Ward JJ, Wattier BA (2011) "Technology for enhancing chest auscultation in clinical simulation". Respiratory Care 56(6): 834-845.