Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Visualization Analysis of Research Status and Hotspots of Computerized Respiratory Sound Analysis Based on CiteSpace

Yan Wang, Yu Xu, Yali Fan, Yang Zhang, Zixuan Qi, Yang Cheng*

Department of Respiratory and Critical Care Medicine, Beijing Jishuitan Hospital, Capital Medical University, Beijing 100035, China

**Corresponding author: Yang Cheng, chengyang0323@outlook.com*

Abstract: Objective: To explore the research landscape and hotspots of Computerized Respiratory Sound Analysis (CORSA) and provide a reference for future in-depth studies. **Methods:** Literature related to CORSA published up to August 27, 2020, was retrieved from the Web of Science Core Collection. CiteSpace 5.6.R3 was used to perform co-authorship analysis, institutional collaboration analysis, keyword co-occurrence analysis, and co-citation analysis. **Results**: A total of 1,897 publications were included. Co-authorship analysis identified several influential contributors, including Zahra Moussavi, Kenneth Sundaraj, and H. Pasterkamp. Major research institutions included the University of Manitoba, the University of Queensland, and Aristotle University of Thessaloniki. Keyword co-occurrence analysis indicated that "respiratory sound," "lung sound," "asthma," "children," and "classification" were major research themes. The most frequently co-cited articles were published by Arati Gurung (2011), Mohammed Bahoura (2009), and H. Pasterkamp (1997). Highly cited journals included Chest, the American Journal of Respiratory and Critical Care Medicine, and IEEE Transactions on Biomedical Engineering. **Conclusion:** CORSA research is primarily driven by European and North American scholars and institutions, with China still in an early stage of development. Current hotspots include respiratory sound acquisition and processing, feature extraction methods such as Mel-frequency cepstral coefficients (MFCCs), and classification techniques based on machine learning and deep learning. CORSA is suitable for diverse populations and is widely applied in respiratory diseases, especially bronchial asthma. Its non-invasive nature offers particular advantages for infants and pregnant women. Although CORSA demonstrates strong clinical potential, its clinical translation remains limited. Advancing clinical applications and bridging the gap between research and practice will be key directions for future development. The prominence of top-tier respiratory and engineering journals among citations suggests that CORSA is an emerging and influential research frontier.

Keywords: Respiratory sounds; Sound analysis; Computer; CiteSpace

Online publication: December 16, 2025

1. Introduction

Respiratory diseases such as pulmonary infections, chronic obstructive pulmonary disease (COPD), bronchial asthma, and lung cancer have become major public health concerns in China, ranking as the fourth leading cause of death among urban residents and the leading cause in rural areas [1]. Although modern medicine has introduced numerous objective diagnostic tools, including computed tomography (CT), magnetic resonance imaging (MRI), pulmonary function tests, pulmonary angiography, and radioactive isotope scanning, none can fully replace pulmonary auscultation. Subtle respiratory sounds, such as dry and moist rales, remain difficult to detect using these advanced imaging or physiological techniques. Since its invention by Laënnec in 1821, the stethoscope has remained the most widely used clinical instrument for bedside assessment [2]. However, traditional auscultation suffers from several limitations, including a narrow frequency response, high dependence on clinician experience, lack of objective recording and storage mechanisms, and difficulty in achieving quantitative comparisons. These challenges hinder its effectiveness in respiratory disease diagnosis and long-term symptom monitoring.

With advancements in digital signal processing, respiratory sounds can now be recorded, analyzed, and automatically classified through Computerized Respiratory Sound Analysis (CORSA) systems [3]. CORSA effectively addresses many drawbacks of conventional stethoscopes, providing objective, repeatable, and quantifiable assessments of respiratory sounds [4]. Despite increasing research interest in CORSA in recent years, standardized analytical methods and systematic comparisons have not yet been established.

Bibliometrics provides a quantitative research approach that examines literature and related scholarly outputs using mathematical and statistical methods to reveal the research status, development trends, and hotspots in a given field [5]. By focusing on publication characteristics such as contributing countries, institutions, journals, authors, and keywords, bibliometric analysis enables researchers to summarize existing research landscapes, identify developmental trajectories, and highlight potential future directions [6–8]. Although bibliometrics has been widely applied across a range of scientific disciplines, there is currently no dedicated scientometric investigation of CORSA. As an accessible and efficient tool for both qualitative and quantitative evaluation, bibliometrics can help uncover thematic evolution, major research areas, and emerging topics within CORSA-related studies [9–12].

To address this gap, the present study conducts a comprehensive bibliometric analysis of CORSA research published over the past three decades. Using the Web of Science Core Collection as the data source and CiteSpace software as the analytical platform, we map research trends, hotspots, collaborative networks, and influential contributions, with the aim of providing a systematic overview and a foundation for further advancement in the CORSA field [13].

2. Materials and methods

2.1. Data sources

We conducted a comprehensive search of the Web of Science Core Collection to identify publications related to CORSA. The search covered all records from the database's inception through December 31, 2022. Inclusion criteria were restricted to English-language publications indexed in the Science Citation Index Expanded (SCIE), with document types limited to "articles" and "reviews." All retrieved records, including full bibliographic information and cited references, were exported in plain text format and saved under the unified filename structure download_.txt to ensure compatibility and accurate recognition by CiteSpace.

2.2. Search strategy

A search strategy was developed by combining subject terms and free-text terms related to respiratory sounds and computer-assisted analysis. The search field was restricted to "Subject" (TS). The final search strategy was as follows:

(1) Respiratory sound–related terms: TS = ("respiratory sound*" OR "lung sound*" OR "breath* sound*" OR wheez* OR crackle* OR squawk* OR stridor* OR rhonch*
OR rale* OR "pleural rub*" OR "abnormal lung sound*" OR "abnormal respiratory sound*"
OR "adventitious lung sound*" OR "adventitious respiratory sound*" OR "lung disorder")

(2) Computer-analysis–related terms: TS = ("Numerical Analysis, Computer-Assisted" OR "Image Processing, Computer-Assisted" OR "Fourier Analysis"
OR "Machine Learning" OR "digital auscultation" OR "electronic auscultation" OR "sound analysis"
OR "computerized analysis" OR "computerized lung sound analysis" OR "signal process*"
OR "automated classification of lung sound")

2.3. Data analysis methods and metric interpretation

This study employed CiteSpace 5.6.R3 to conduct bibliometric visualization analysis. The time-slicing range was set from 1987 to 2020, with 3-year intervals, enabling the identification of developmental trajectories and emerging research trends in CORSA while improving computational efficiency. The threshold was set to Top 50, and the Pathfinder algorithm was applied. The selected node types included Author, Institution, Country, Keyword, Reference (document co-citation), and Cited Journal (journal co-citation) [14].

In CiteSpace, node size is proportional to the frequency of the node type; growth-ring colors represent citation time, and ring thickness reflects the quantity of items within that time slice. The thickness and density of lines indicate the strength of relationships between nodes. Core authors were identified using Price's Law, calculated as $N = 0.749 \times (\eta max)^{1/2}$, where ηmax is the publication count of the most productive author [15]. Authors with publication counts greater than N were classified as core authors. Betweenness centrality was used to evaluate the importance of nodes within the network, where nodes with centrality ≥ 0.1 were emphasized with purple rings. Burst terms, detected using burst-detection algorithms, were used to identify emerging research trends and potential future hotspots [16].

In addition, this study used CiteSpace (6.1.R4) and the Bibliometrix 4.1.0 package in R (https://www.bibliometrix.org) for comprehensive bibliometric analysis. CiteSpace supports visualization of knowledge structures and emerging trends, including cluster analysis, dual-map citation overlays, timelines, and burst detection of keywords and references [9,11]. Cluster modularity (Q value) and silhouette score (S value) were used to evaluate clustering quality, where $Q > 0.3$ indicates significant cluster structure and $S > 0.5$ indicates reliable clustering.

The Bibliometrix package, a widely used R-based tool for bibliometrics, was applied to analyze thematic evolution and to identify annual publication output, productive countries/regions, institutions, journals, authors, and highly co-cited journals, authors, and references [17]. Nodes in Bibliometrix maps represent countries, institutions, or journals, with node size indicating research volume or co-occurrence frequency. Connections between nodes represent co-occurrence strength.

The CiteSpace parameter settings were as follows: LRF = 3, LBY = −1, Top N = e (e = 2), Time Span = 2010–2019, Years Per Slice = 1, Link Strength = Cosine, Scope = Within Slices, Selection Criteria = g-index (K =

25), and Minimum Duration (MD) = 5.

3. Results

3.1. Overall distribution

Based on the above search strategy, a total of 2,563 CORSA-related publications were retrieved from 1977 to 2022, including 2,395 research articles and 168 reviews. These publications were primarily written in English (n = 2,487; 97.03%), followed by German (n = 35; 1.37%), Turkish (n = 17; 0.66%), French (n = 15; 0.59%), Spanish (n = 15; 0.59%), Portuguese (n = 13; 0.51%), Russian (n = 12; 0.47%), Chinese (n = 2; 0.08%), Croatian (n = 2; 0.08%), and Italian (n = 2; 0.08%).

The annual output of CORSA-related publications shows a clear upward trajectory from 1977 to 2022, reflecting growing scholarly attention and continuous expansion of the field. According to publication volume and growth rate, this period can be divided into two main phases:

(1) Initial stage (1977–1990): Fewer than 5 publications were produced per year;

(2) Development stage (1991–2022):

  (i) 1991–2012: Fewer than 100 publications per year, with an average annual increase of 3 publications and an average growth rate of 5.2%;

  (ii) 2013–2022: A rapid growth phase in which annual publications exceeded 100, with an average annual increase of 16 publications and an average growth rate of 10.2%.

Curve fitting analysis demonstrated a strong correlation between publication year and research output, with a coefficient of determination $R^2 = 0.9397$ ($p < 0.001$), indicating a significant growth trend in CORSA-related studies.

3.2. Country/region and institutional analysis

A total of 2,563 publications were produced by 202 institutions across 113 countries/regions. The top 10 countries/regions and institutions ranked by publication volume are presented in **Table 1**. The United States contributed the highest number of publications (n = 575), followed by China (n = 224), the United Kingdom (n = 195), and Canada (n = 191). The collaboration network highlights major contributing countries, with the United States, China, the United Kingdom, Canada, India, and Germany represented by larger nodes, reflecting higher productivity. Extensive international collaborations were observed: for example, the United States collaborated closely with China, Canada, the United Kingdom, and Germany; France collaborated with the United Kingdom, Canada, Iran, and Malaysia; and China collaborated with South Korea, Australia, and the United States.

The top ten institutions were distributed across seven countries/regions. Four institutions produced more than 30 publications: the University of Manitoba (n = 60), the University of Queensland (n = 35), the University of Toronto (n = 32), and Johns Hopkins University (n = 30). As shown in Figure 3, the co-authorship network reveals substantial institutional collaboration. The University of Manitoba, the University of Queensland, and the University of Toronto appear as larger nodes due to their high productivity. Notable collaboration patterns include strong ties between the University of Manitoba and institutions such as the University of Toronto, the University of Porto, and Monash University. Similarly, Johns Hopkins University has collaborated extensively with the Indian Institute of Technology, Stanford University, and the University of Colorado.

Table 1. Main core authors in the field of CORSA (frequency ≥ 12)

No.	Author	Frequency	Affiliation	Representative paper
1	Zahra Moussavi	18	Dept. of Electrical & Computer Engineering, University of Manitoba, Canada	Automatic breath and snore sounds classification from tracheal and ambient sounds recordings (doi: 10.1016/j.medengphy.2010.06.013)
2	Kenneth Sundaraj	17	AI-Rehab Research Group, University Malaysia Perlis, Malaysia	Machine learning in lung sound analysis: A systematic review (doi: 10.1016/j.bbe.2013.07.001)
3	H. Pasterkamp	17	Dept. of Pediatrics & Child Health, University of Manitoba, Canada	Respiratory sounds. Advances beyond the stethoscope (doi: 10.1164/ajrccm.156.3.9701115)
4	Yukio Nagasaka	15	Kyoto Respiratory Center, Otowa Hospital, Japan	Lung sound analysis and airway inflammation in bronchial asthma (doi: 10.1016/j.jaip.2016.02.001)
5	Raimon Jané	14	Dept. of Automatic Control, Universitat Politècnica de Catalunya, Spain	Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency (doi: 10.1109/JBHI.2015.2396636)
6	L. J. Hadjileontiadis	13	Dept. of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Greece	Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter (doi: 10.1109/10.649999)
7	Rajkumar Palaniappan	13	AI-Rehab Research Group, University Malaysia Perlis, Malaysia	A comparative study of the SVM and K-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals (doi: 10.1186/1471-2105-15-223)
8	Z. Moussavi	13	Dept. of Electrical & Computer Engineering, University of Manitoba, Canada	Heart sound cancellation from lung sound recordings using time-frequency filtering (doi: 10.1007/s11517-006-0030-8)
9	Udantha R. Abeyratne	12	School of IT & Electrical Engineering, The University of Queensland, Australia	Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis (doi: 10.1109/TBME.2014.2381214)
10	A. R. A. Sovijärvi	12	Dept. of Medicine, Helsinki University Central Hospital, Finland	Crackles: recording, analysis and clinical significance (doi: 10.1183/09031936.95.08122139)

3.3. Journals and co-cited academic journals

A total of 1,995 DILI-related publications were disseminated across 592 academic journals. Six journals published more than 35 papers, five of which were based in the United States; the only exception was Archives of Toxicology, published in Germany (Table 2). Toxicology had the highest publication count (n = 79; IF2018 = 3.564; Q1), followed by Hepatology (n = 48; IF2018 = 14.971; Q1), Journal of Hepatology (n = 38; IF2018 = 5.542; Q1), PLOS ONE (n = 38; IF2018 = 2.776; Q2), and Archives of Toxicology (n = 37; IF2018 = 5.741; Q1).

A citation network was constructed using journals with ≥ 21 publications (T = 21), accounting for 3.55% of all journals (21/592). Toxicology, Hepatology, Journal of Hepatology, and PLOS ONE display larger node sizes, reflecting higher publication outputs. Journals such as Toxicology, Chemical Research in Toxicology, Journal of Hepatology, and PLOS ONE show dense citation linkages. Notably, Toxicology exhibits positive citation relationships with Archives of Toxicology and Chemical Research in Toxicology.

Co-citation analysis was used to identify journals frequently cited together within the same publications (Gao 2019). Hepatology had the highest co-citation frequency (n = 7,383; IF2018 = 14.971; Q1), followed by Gastroenterology (n = 3,326; IF2018 = 19.233; Q1), Journal of Hepatology (n = 2,892; IF2018 = 18.946; Q1), Toxicology (n = 2,317; IF2018 = 3.564; Q1), and Drug Metabolism and Disposition (n = 2,062; IF2018 = 3.354; Q2) (**Table 2**). Among the top 15 co-cited journals, The New England Journal of Medicine had the highest impact factor (IF2018 = 70.67). A total of 21 journals with ≥ 700 co-citations (T = 700) formed the journal co-citation network, representing 7,645 articles (0.27%).

Hepatology and Gastroenterology are represented by larger nodes due to their high co-citation frequencies. Hepatology demonstrates strong co-citation linkages with Gastroenterology, Journal of Hepatology, Toxicological Sciences, and Drug Metabolism and Disposition.

The dual-map overlay of journals (**Figure 1**) illustrates the thematic distribution of citing and cited journals (Chen, 2017; Miao *et al.*, 2018). Cited journals appear on the left and citing journals on the right, with colored paths indicating citation flows. Four primary citation paths were observed, two orange and two green. The orange paths indicate that research published in molecular/biology/genetics and health/nursing/medical journals is commonly cited by studies appearing in molecular/biology/immunology journals. The green paths show that research in molecular/biology/genetics and health/nursing/medical journals is frequently cited by medical/clinical journals.

Table 2. High-frequency keywords in the field of CORSA (top 30)

Rank	Frequency	Keyword
1	307	Respiratory sound
2	304	Lung sound
3	192	Asthma
4	164	Children
5	159	Classification
6	148	Normal breath sound
7	147	Pulmonary disease
8	140	Spectral analysis
9	135	Algorithm
10	129	Sleep apnea
11	116	Time-frequency analysis
12	109	Wavelet transform
13	108	Feature extraction
14	107	Wheeze
15	104	Copd
16	101	Heart sound
17	100	Machine learning
18	100	Air flow
19	99	Pneumonia
20	77	Auscultation
21	76	Crackle
22	75	Acoustic analysis
23	67	Signal processing
24	65	Sensor
25	58	Lung sound analysis
26	57	Spectra
27	56	Fractal dimension
28	56	Snoring
29	54	Parameter
30	53	Artificial neural network

3.4. Analysis of collaboration networks

3.4.1. Analysis of author collaboration networks

The analysis of author collaboration networks reveals that Zahra Moussavi is the most prolific author, with a total of 18 publications. According to Price's Law, with N = 6.7, core authors in CORSA-related research must have published at least 7 papers. Based on CiteSpace statistics, there are 39 core authors in total. Due to space constraints, this report lists only the main core authors who have published 12 or more papers (**Table 1**). As illustrated in **Figure 1**, the CORSA research field currently comprises three main teams, with Yukio Nagasaka, L Hadjileontiadis, and A Sovijärvi serving as the cores of these teams, respectively. Within each team, there is a relatively high level of collaboration, but there is limited collaboration between the teams. The institutions to which the main core authors belong indicate that scholars engaged in CORSA-related research are primarily from disciplines related to electronics and computer engineering. Medical scholars focusing on this field are mainly concentrated in the fields of asthma and pediatrics.

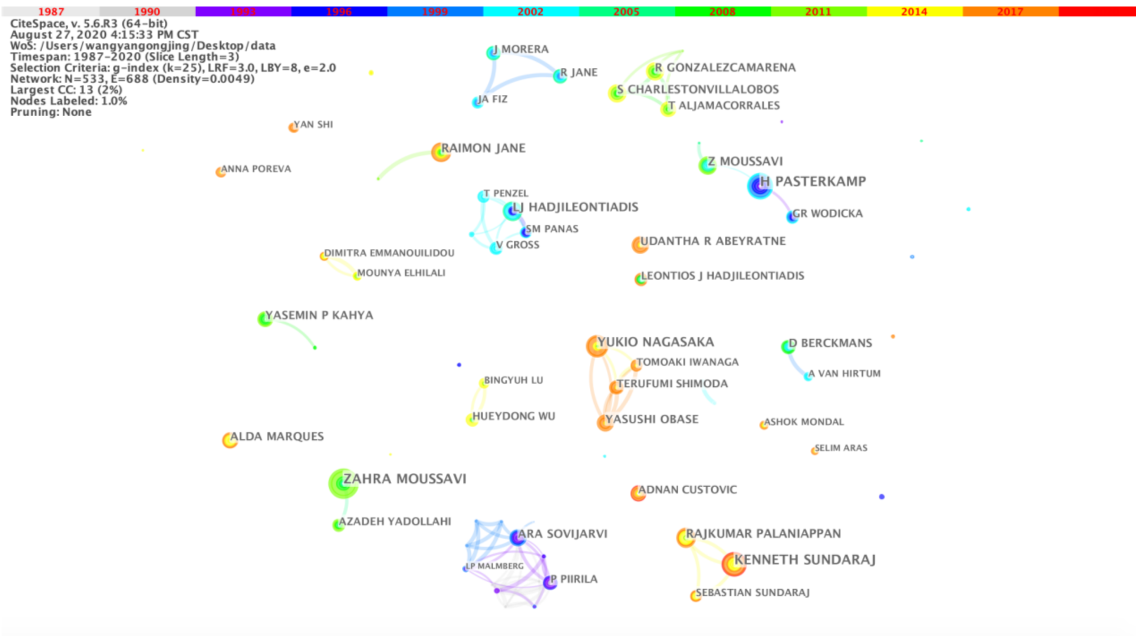

Figure 1. Visualization map of co-author analysis in the field of CORSA.

3.4.2. Analysis of institutional collaboration networks

As shown in **Figure 2**, the top 10 institutions in terms of publication volume are, in order, the University of Manitoba in Canada (55 papers), The University of Queensland in Australia (27 papers), Aristotle University of Thessaloniki in Greece (26 papers), Bogazici University in Turkey (23 papers), the University of Toronto in Canada (21 papers), Katholieke Universiteit Leuven in Belgium (19 papers), Johns Hopkins University in the United States (18 papers), Polytechnic University of Catalonia in Spain (17 papers), Universiti Malaysia Perlis in Malaysia (16 papers), and the University of Washington in the United States (15 papers). It can be seen that there is relatively close cooperation among universities in Europe and the United States, with a large number of publications. The institutions in China with a relatively high number of publications are Zhejiang University, Beihang University (10 papers), and Guangzhou Medical University (9 papers), ranked 24th and 33th respectively. Domestic universities have not yet established close cooperation in the research field related to CORSA.

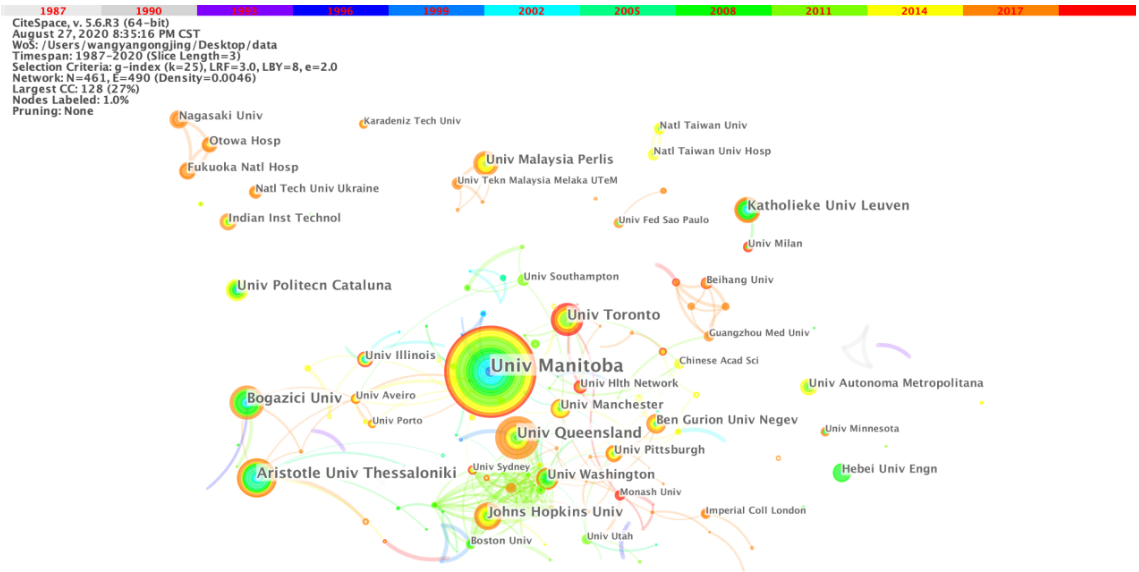

Figure 2. Visualization map of co-institutions analysis in the field of CORSA.

3.4.3. Analysis of national cooperation networks

As shown in **Figure 3**, the top 10 countries in terms of publication volume are the United States (442 papers), Canada (155 papers), China (138 papers), the United Kingdom (138 papers), Japan (106 papers), Germany (100 papers), India (83 papers), Italy (74 papers), Australia (73 papers), and France (69 papers). Among them, the United States (0.55), the United Kingdom (0.20), Germany (0.17), Australia (0.12), and France (0.12) have a betweenness centrality of ≥ 0.1, serving as key hubs connecting different countries and are highlighted with purple circles. The size of the network nodes in the figure represents the contribution of each country, while the number and thickness of the connecting lines indicate the strength of cooperative relationships between countries.

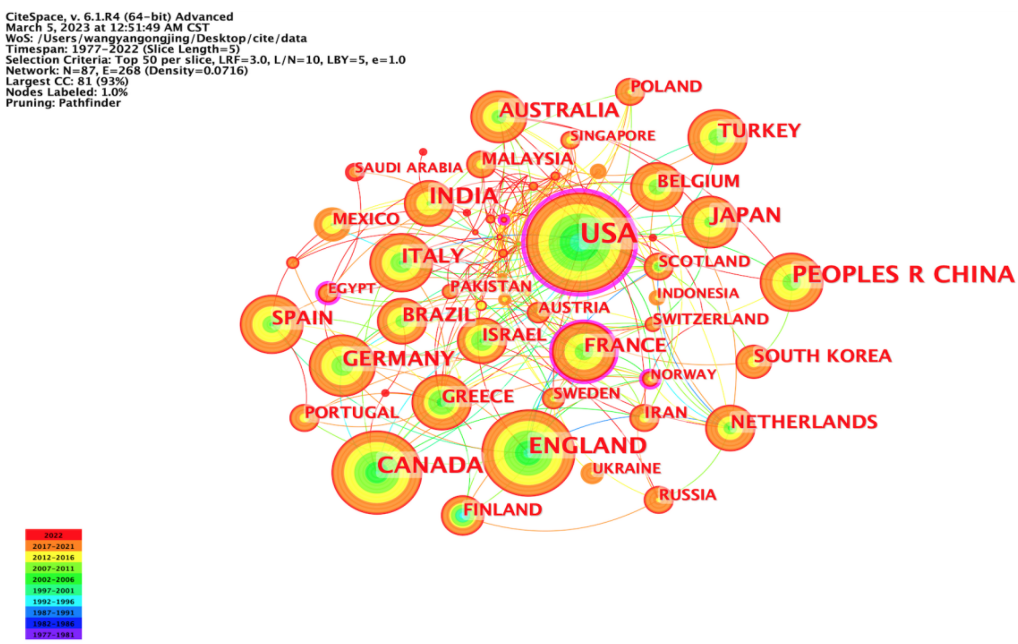

Figure 3. Visualization map of co-countries analysis in the field of CORSA.

3.5. Keyword co-occurrence analysis

Keywords in literature are one of the important labels reflecting the content of the literature. Keywords with high frequency of occurrence within a certain period may represent research hotspots or focal points in the field at that time [14]. The visualization map of keyword co-occurrence is shown in **Figure 4**, and the top 30 high-frequency keywords with larger nodes, indicating higher co-occurrence frequencies, are listed in **Table 2**. After manually clustering the keywords based on reading the literature, the specific clusters are shown in **Table 3**.

It can be seen that the keywords related to CORSA research mainly involve five aspects: respiratory sounds, signal acquisition and processing, signal feature extraction, signal classification, and clinical applications. Burst words are keywords with high burst rates that are widely concerned in a short period and can represent the research frontiers of CORSA-related research. By intercepting these words through CiteSpace, a table of high-burst-value keywords related to CORSA research is obtained (**Figure 5**). The red sections represent the start and end time periods during which keywords appear frequently in the short term. As shown in **Figure 5**, regarding research related to CORSA, in terms of signal feature extraction, the specific methods have gradually evolved from early waveform analysis to fractal dimension and wavelet transform. In recent years, MFCC have gradually become a research hotspot. In terms of the selection of signal classifiers, there has been a gradual shift from traditional neural networks in the early days to support vector machines. Currently, machine learning and deep learning are the research hotspots in this field.

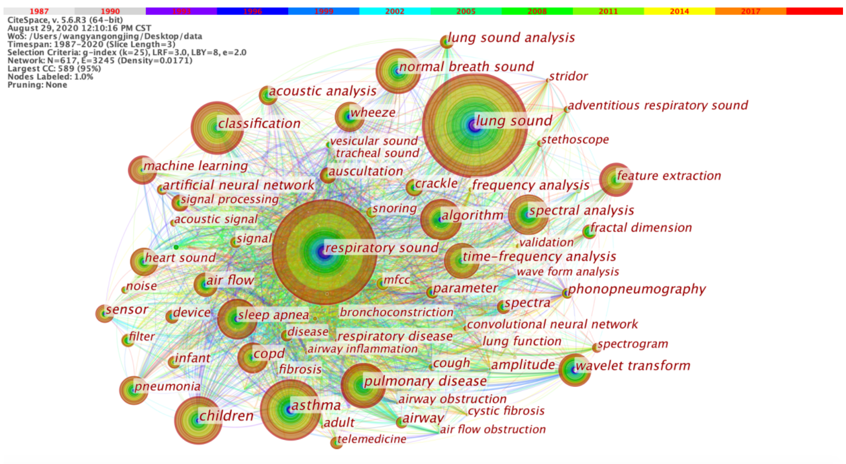

Figure 4. Visualization map of co-occurrence network analysis of keywords in the field of CORSA.

Table 3. Keywords co-occurrence clustering table of CORSA

Cluster ID	Cluster name	Main included keywords
1	Respiratory sounds	Respiratory sound, normal breath sound, tracheal sound, lung sound, vesicular sound, adventitious respiratory sound, crackle, wheeze, stridor, snoring, auscultation, stethoscope
2	Signal acquisition & processing	Sensor, electronic stethoscope, signal, acoustic signal, signal processing, acoustic transmission, noise, heart sound, bowel sound, filter
3	Signal feature extraction	Feature extraction, parameter, spectral analysis, wavelet transform, phonopneumography, spectrogram, amplitude, time-frequency analysis, MFCC, frequency analysis, waveform analysis, entropy, fractal dimension
4	Signal classification	Classification, algorithm, machine learning, artificial neural network, convolutional neural network, support vector machine, deep learning, validation
5	Clinical application	Adult, adolescent, children, infant, respiratory disease, pulmonary disease, asthma, sleep apnea, COPD, pneumonia, community acquired pneumonia, interstitial lung disease, bronchiectasis, cystic fibrosis, cough, device, telemedicine

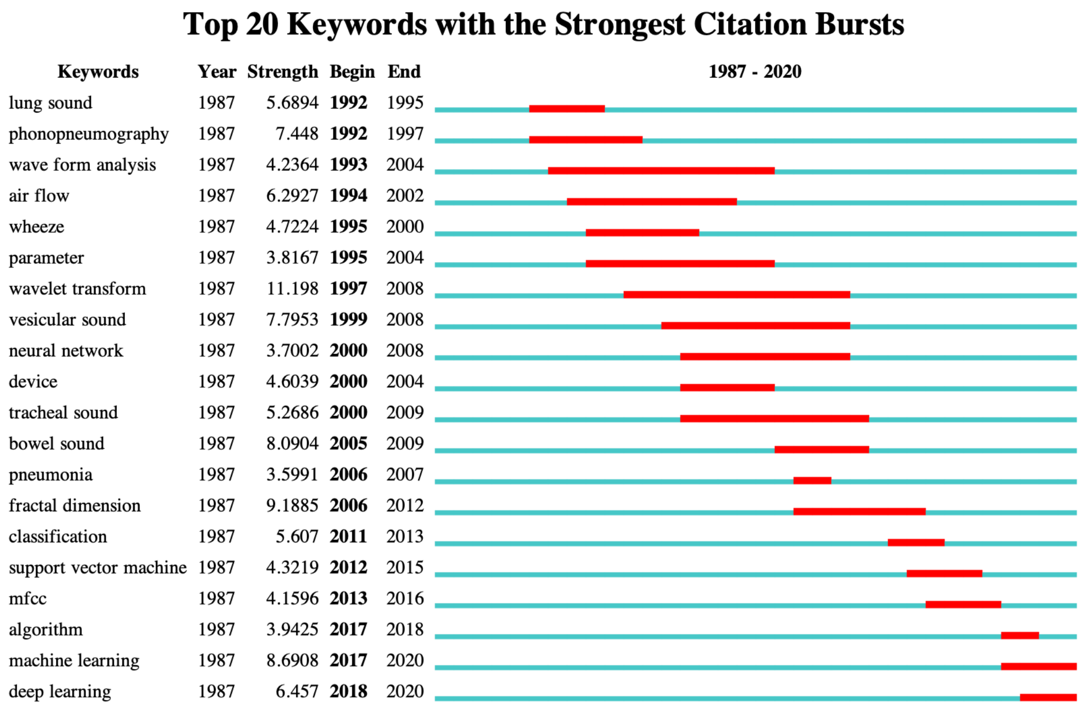

Figure 5. Keywords with the strongest citation bursts in the field of CORSA.

3.6. Co-citation analysis

3.6.1. Document co-citation analysis

Document co-citation analysis provides insight into the academic influence of publications within a field, as highly co-cited documents typically represent foundational or widely recognized research. As shown in **Figure 6**, which presents the visualization map of literature co-citation analysis in the CORSA field, the three most frequently co-cited papers are as follows.

(1) Gurung *et al.* (2011) (50 co-citations): This systematic review and meta-analysis evaluated studies on computerized lung sound analysis (CLSA). The authors reported that most studies utilized electret microphones or piezoelectric sensors for auscultation and employed Fourier transform and neural network algorithms for lung sound analysis and automatic classification. CLSA demonstrated an overall sensitivity of 80% (95% CI: 72–86%) and a specificity of 85% (95% CI: 78–91%) for detecting wheezes or crackles [18];

(2) Bahoura (2009) (44 co-citations): This study proposed a pattern-recognition approach to classify breath sounds into normal and wheezing categories. By comparing feature extraction methods, including Fourier Transform, Linear Predictive Coding, Wavelet Transform, and MFCC, and combining them with vector quantization, Gaussian Mixture Models (GMM), and artificial neural networks (ANN), the study demonstrated that MFCC combined with GMM yields the most effective performance for distinguishing wheezing from normal breath sounds [19];

(3) Pasterkamp (1997) (39 co-citations): This influential review examined the thoracic and upper respiratory tract as acoustic systems, described methods for measuring lung sounds, and summarized existing knowledge on the origins and clinical relevance of normal and adventitious breath sounds. It also provided insights into potential clinical applications of breath sound analysis [4].

Collectively, these three publications represent classic and highly influential references that have shaped the development of CORSA research.

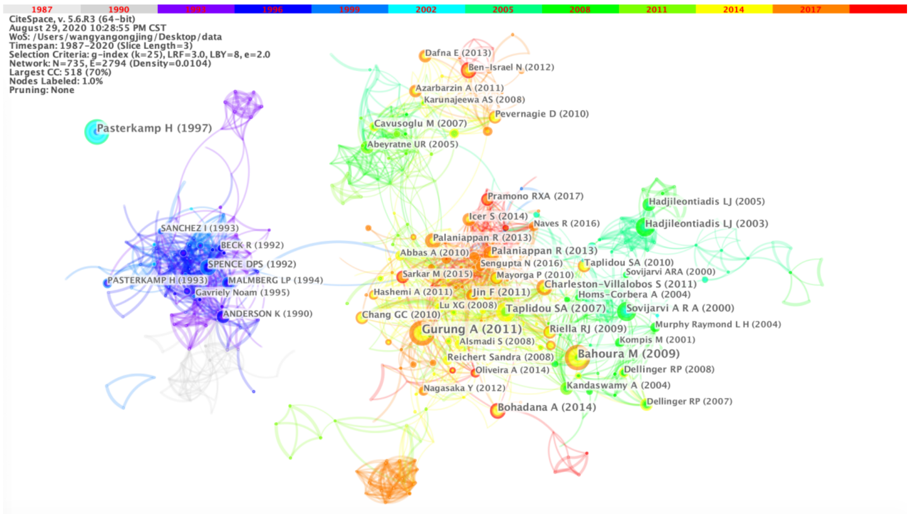

Figure 6. Visualization map of literature co-citation analysis in the field of CORSA.

3.6.2. Journal co-citation analysis

Journal co-citation analysis can reveal the disciplinary scope of a research field. As illustrated in **Figure 7**, the top 10 journals by co-citation frequency are: Chest (584), American Journal of Respiratory and Critical Care Medicine (510), IEEE Transactions on Biomedical Engineering (456), European Respiratory Journal (430), Journal of Applied Physiology (380), New England Journal of Medicine (361), Thorax (358), American Review of Respiratory Disease (338), Lancet (289), and Medical & Biological Engineering & Computing (266). These journals span multiple disciplines, including general medicine, biomedicine, physiology, respiratory systems, critical care, sleep medicine, computer science, mathematics, computational biology, and engineering technology. The distribution of co-cited journals indicates that CORSA research is interdisciplinary, with broad applications across both medical and engineering fields.

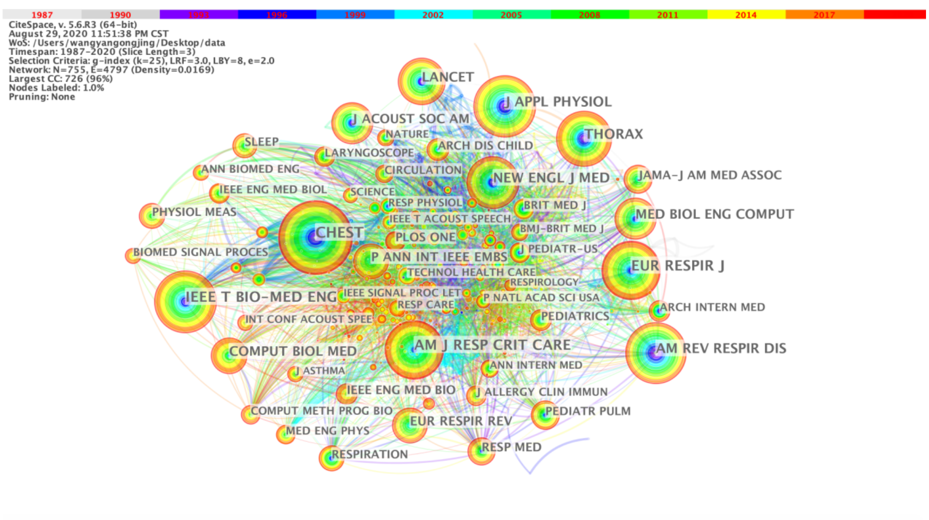

Figure 7. Visualization map of journal co-citation analysis in the field of CORSA.

4. Discussion

CORSA involves recording respiratory sounds using electronic devices and subsequently analyzing them through computational methods to classify signals based on their characteristic features. The considerable overlap among different respiratory sounds, coupled with their non-stationary and nonlinear nature, makes the extraction of universal and representative features challenging. Advances in computational techniques have facilitated research in this area, prompting increased attention from various research teams. In this study, CiteSpace 5.6.R3 was employed to visually analyze CORSA-related literature from the Web of Science Core Collection, providing an objective overview of the current research status and hotspots.

Analysis of author, institution, and country collaboration networks allowed the identification of leading contributors and their cooperation patterns. Over the past three decades, CORSA research has been conducted worldwide, with scholars and institutions from Europe and the United States dominating the field and gradually undertaking deeper investigations, while China remains in the early stages of development. Keyword co-occurrence, cluster analysis, and burst term identification highlighted several main research directions as follows:

(1) Signal acquisition and processing: The most commonly used sensors are piezoelectric and electret microphones [2]. Typical acquisition sites include the suprasternal notch, posterior thoracic wall (5 cm lateral to the midline, 7 cm below the inferior scapular angles), anterior thoracic wall (second intercostal space along the midclavicular line), lateral thoracic wall (fourth and fifth intercostal spaces in the axilla), and the trachea. Minimum recommended acquisition sites include the trachea and the left and right bases of the posterior thoracic wall [20]. Noise reduction, including environmental and physiological interference such as heart and bowel sounds, is essential. Methods include shielding sensors with soundproof materials, using high-pass filters (70–100 Hz), and applying adaptive digital signal processing techniques in computers or microprocessors [21–23];

(2) Signal feature extraction: Feature extraction identifies distinguishing attributes of signals and is crucial for effective classification [24]. Since respiratory sounds are non-stationary and nonlinear, time-frequency domain analyses are particularly useful [3,25]. Common methods include Short-Time Fourier Transform (STFT), Autoregressive Model (AR), Fractal Dimension (FD) analysis, MFCC, Wavelet Transform (WT), and Wavelet Packet Transform (WPT) [25–29]. STFT remains widely used, while MFCC has emerged as a recent research hotspot;

(3) Signal classification: Classification algorithms have evolved from early empirical rule-based approaches (e.g., threshold or peak selection) to advanced machine learning techniques such as Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs), k-Nearest Neighbors (k-NN), Gaussian Mixture Models (GMMs), genetic algorithms, fuzzy logic, and Support Vector Machines (SVMs) [19,30–32]. ANNs efficiently handle complex nonlinear data and provide accurate classification [33]. For example, Kandaswamy *et al.* reported training and testing classification accuracies of 100% and 94.02% for normal, wheeze, crackle, and rhonchi sounds [25]. The k-NN algorithm offers computational simplicity but requires large datasets for optimal accuracy [33,34]. Despite limitations, ANNs and k-NN remain widely used in CORSA due to their superior performance and precision in detecting respiratory sounds;

(4) Clinical applications: CORSA has evolved from sound identification to diagnostic applications for pulmonary diseases. Its non-invasive, safe, real-time, and recordable nature makes it suitable for adults, children, infants, and pregnant women, who may be unable to undergo radiological or invasive procedures. CORSA has been applied to asthma, sleep apnea, COPD, pneumonia, and pulmonary

fibrosis, with bronchial asthma being the most frequent target. Its use spans outpatient clinics, hospital wards, home monitoring, and telemedicine. During the COVID-19 pandemic, CORSA devices facilitated early detection of cases with high risk of progression, improving early diagnosis, treatment efficiency, and reducing infection risk. Nevertheless, challenges such as device portability, signal complexity, and variability have hindered widespread clinical adoption.

Document co-citation analysis provided insights into influential studies, including those by Gurung (2011), Bahoura (2009), and Pasterkamp (1997), facilitating a comprehensive understanding of CORSA research progress [4,18,19]. Journal co-citation analysis indicated that CORSA-related research spans multiple disciplines, with high co-citation frequencies in top-tier journals in respiratory medicine, critical care, and electronic engineering, highlighting the interdisciplinary nature and potential future directions of this field.

5. Conclusion

In summary, despite significant innovations in computer-based analysis of breath sounds over the past 30-plus years, it has not yet assumed a significant position in clinical medicine. It goes without saying that modern technology offers immense advantages in signal acquisition, processing, storage, analysis, and communication. However, standard validation and data management methods have not yet been established. Therefore, how to routinely and effectively utilize these facilities to aid in the daily diagnosis and management of patients with respiratory diseases remains to be determined. Researchers in the engineering and technology fields have primarily focused their attention on CORSA in terms of signal feature extraction methods and classifier selection, while medical researchers have concentrated on its clinical applications. The effective integration of medicine and engineering, leveraging advanced signal processing and artificial intelligence technologies to develop computer-based real-time breath sound analysis systems for disease diagnosis and monitoring, and commercializing these systems represent key areas for future research. This study has certain limitations, primarily stemming from the analysis of only English-language literature published in the Web of Science Core Collection, which may impact the overall research findings when used to represent the entire body of research. This study visually presents the main core authors, research institutions, countries, current research status and hotspots, historical evolution trends of research hotspots, classic cited literature, and publishing journals related to CORSA research, with the aim of providing references for further in-depth research and facilitating the early development of a widely applicable computerized real-time breath sound analysis system for clinical use.

Funding

Beijing Hospitals Authority Clinical Medicine Development of Special Funding Support, Code (Project No.: YGLX202514)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang C, Xu J, Yang L, et al., 2018, Prevalence and Risk Factors of Chronic Obstructive Pulmonary Disease in China (The China Pulmonary Health [CPH] Study): A National Cross-Sectional Study. The Lancer, 391(10131): 1706–1717.

[2] Welsby P, Parry G, Smith D, 2003, The Stethoscope: Some Preliminary Investigations. Postgrad Med J, 79(938): 695–698.

[3] Sovijärvi A, Vanderschoot J, Malmberg L, et al., 2000, Definition of Terms for Applications of Respiratory Sounds. Eur Respir Rev, 2000(10): 597–610.

[4] Pasterkamp H, Kraman S, Wodicka G, 1997, Advances Beyond the Stethoscope. American Journal of Respiratory and Critical Care Medicine, 1997(156): 974–987.

[5] Daim T, Rueda G, Martin H, et al., 2006, Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis. Technological Forecasting & Social Change, 73(8): 981–1012.

[6] Ellegaard O, Wallin J, 2015, The Bibliometric Analysis of Scholarly Production: How Great is the Impact? Scientometrics, 105(3): 1809–1831.

[7] Gao Y, Shi S, Ma W, et al., 2019, Bibliometric Analysis of Global Research on PD-1 and PD-L1 in the Field of Cancer. Int Immunopharmacol, 2019(72): 374–384.

[8] Lu C, Bing Z, Bi Z, et al., 2019, Top 100 Most Cited Publications on Network Pharmacology: A Bibliometric Analysis. Evid Based Complement Alternat Med, 2019: 1704816.

[9] Chen C, 2004, Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization. Proc Natl Acad Sci USA, 101(Suppl 1): 5303–10.

[10] Chen C, 2017, Science Mapping: A Systematic Review of the Literature. Journal of Data and Information Science, 2(2): 1–40.

[11] Chen C, Song M, 2019, Visualizing a Field of Research: A Methodology of Systematic Scientometric Reviews. PLoS One, 14(10): e0223994.

[12] Wang S, Zhou H, Zheng L, et al., 2021, Global Trends in Research on Macrophages Associated with Acute Lung Injury Over the Past 10 Years: A Bibliometric Analysis. Front Immunol, 2021(12): 669539.

[13] Synnestvedt M, Chen C, Holmes J, 2005, CiteSpace II: Visualization and Knowledge Discovery in Bibliographic Databases. AMIA Annu Symp Proc, 2005: 724–8.

[14] Li J, Chen C, 2016, CiteSpace: Scientific Text Mining and Visualization, Capital University of Economics and Business Press, Beijing, 196–198.

[15] Zong S, 2016, Evaluation of Core Authors Based on Price's Law and the Composite Index Method: A Case Study of "Chinese Journal of Scientific and Technical Periodicals". Chinese Journal of Scientific and Technical Periodicals, 27(12): 1310–1314.

[16] Chen C, 2006, CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. Journal of the American Society for Information Science and Technology, 57(3): 359–377.

[17] Aria M, Cuccurullo C, 2018, Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. Journal of Informetrics, 11(4): 959–975.

[18] Gurung A, Scrafford C, Tielsch J, et al., 2011, Computerized Lung Sound Analysis as Diagnostic Aid for the Detection of Abnormal Lung Sounds: A Systematic Review and Meta-Analysis. Respir Med, 105(9): 1396–1403.

[19] Bahoura M, 2009, Pattern Recognition Methods Applied to Respiratory Sounds Classification into Normal and Wheeze Classes. Comput Biol Med, 39(9): 824–843.

[20] Sovijärvi A, Vanderschoot J, Earis J, 2000, Standardization of Computerized Respiratory Sound Analysis. Eur Respir Rev, 10(77): 585.

[21] Kaisla T, Sovijärvi A, Piirilä P, et al., 1991, Validated Method for Automatic Detection of Lung Sound Crackles. Med Biol Eng Comput, 29(5): 517–521.

[22] Ono M, Arakawa K, Mori M, et al., 1989, Separation of Fine Crackles from Vesicular Sounds by a Nonlinear Digital Filter. IEEE Trans Biomed Eng, 36(2): 286–291.

[23] Suzuki A, Sumi C, Nakayama K, et al., 1995, Real-Time Adaptive Cancellation of Ambient Noise in Lung Sound Measurement. Medical and Biological Engineering and Computing, 33(5): 704–708.

[24] Mariello A, Battiti R, 2018, Feature Selection based on Neighborhood Entropy. IEEE Transactions on Neural Networks and Learning Systems, 29(12): 6313–6322.

[25] Kandaswamy A, Kumar C, Ramanathan R, et al., 2004, Neural Classification of Lung Sounds using Wavelet Coefficients. Computers in Biology and Medicine, 34(6): 523–537.

[26] Flietstra B, Markuzon N, Vyshedskiy A, et al., 2011, Automated Analysis of Crackles in Patients with Interstitial Pulmonary Fibrosis. Pulmonary Medicine, 2011: 7.

[27] Riella R, Nohama P, Maia J, 2009, A Method for Automatic Detection of Wheezing in Lung Sounds. Brazilian Journal of Medical and Biological Research, 42(7): 674–684.

[28] Abbas A, Fahim A, 2010, An Automated Computerized Auscultation and Diagnostic System for Pulmonary Diseases. Journal of Medical Systems, 34(6): 1149–1155.

[29] Azarbarzin A, Moussavi Z, 2010, Unsupervised Classification of Respiratory Sound Signals into Snore/No-Snore Classes. Conference Proceedings of the IEEE Engineering in Medicine and Biology Society, 2010: 3666–3669.

[30] Wiśniewski M, Zieliński T, 2015, Joint Application of Audio Spectral Envelope and Tonality Index in an E-Asthma Monitoring System. IEEE Journal of Biomedical and Health Informatics, 19(3): 1009–1018.

[31] Sengupta N, Sahidullah M, Saha G, 2016, Lung Sound Classification Utilizing Cepstral-Based Statistical Features. Computers in Biology and Medicine, 2016(75): 118–29.

[32] Alsmadi S, Kahya Y, 2008, Development of a DSP-Based Device for Real-Time Classification of Pulmonary Sounds. Computers in Biology and Medicine, 38(1): 53–61.

[33] Tu J, 1996, The Pros and Cons of Employing Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. Journal of Clinical Epidemiology, 49(11): 1225–31.

[34] Yang M, Chen C, 1998, On the Modified K-Nearest Neighbor Rule. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 28(3): 461–6.