Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers

Luay Fraiwan [a,*], *Omnia Hassanin* [b], *Mohammad Fraiwan* [c], *Basheer Khassawneh* [d], *Ali M. Ibnian* [d], *Mohanad Alkhodari* [b]

[a] *Department of Biomedical Engineering, Jordan University of Science and Technology, Irbid, Jordan*
[b] *Department of Electrical and Computer Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates*
[c] *Department of Computer Engineering, Jordan University of Science and Technology, Irbid, Jordan*
[d] *Department of Internal Medicine, Jordan University of Science and Technology, Irbid, Jordan*

ARTICLE INFO

Article history:
Received 14 October 2020
Received in revised form
16 November 2020
Accepted 18 November 2020
Available online 01 December 2020

Keywords:
Pulmonary diseases
Lung sound analysis
Shannon entropy
Spectral entropy
Bootstrap aggregation
Adaptive boosting

ABSTRACT

This paper investigates the application of different homogeneous ensemble learning methods to perform multi-class classification of respiratory diseases. The case sample involved a total of 215 subjects and consisted of 308 clinically acquired lung sound recordings and 1176 recordings obtained from the ICBHI Challenge database. These recordings corresponded to a wide range of conditions including healthy, asthma, pneumonia, heart failure, bronchiectasis or bronchitis, and chronic obstructive pulmonary disease. Feature representation of the lung sound signals was based on Shannon entropy, logarithmic energy entropy, and spectrogram-based spectral entropy. Decision trees and discriminant classifiers were employed as base learners to build bootstrap aggregation and adaptive boosting ensembles. The optimal structure of the investigated ensemble models was identified through Bayesian hyperparameter optimization and was then compared to typical classifiers in literature. Experimental results showed that boosted decision trees provided the best overall accuracy, sensitivity, specificity, F1-score, and Cohen's kappa coefficient of 98.27%, 95.28%, 98.9%, 93.61%, and 92.28%, respectively. Among the baseline methods, SVM provided the best yet a slightly poorer performance, as demonstrated by its average accuracy (98.20%), sensitivity (91.5%), and specificity (98.55%). Despite their simplicity, the investigated ensemble classification methods exhibited a promising performance for detecting a wide range of respiratory disease conditions. The data fusion approach provides a promising insight into an alternative and more suitable solution to reduce the effect of imbalanced data for clinical applications in general and respiratory sound analysis studies in specific.

* Corresponding author at: Department of Biomedical Engineering, Jordan University of Science and Technology, Irbid, Jordan.
E-mail address: fraiwan@just.edu.jo (L. Fraiwan).

Abbreviations: RD, respiratory diseases; ALS, adventitious lung sounds; MFCC, mel frequency spectral coefficients.

https://doi.org/10.1016/j.bbe.2020.11.003

1. Introduction

According to a recent report published by the Forum of International Respiratory Societies (FIRS), respiratory diseases (RDs) are among the leading causes of severe illness worldwide, with a death toll exceeding 4 million lives annually [1]. In 2017, the World Health Organization (WHO) declared that chronic RDs accounted for more than 10% of the global disease burden, second only to cardiovascular diseases [2].

The diagnostic process of RD involves auscultation, which is a clinical examination that involves listening to the internal sounds of moving air inside and outside the lungs. Pulmonary sounds are commonly auscultated through the anterior and posterior chest walls or the trachea using a stethoscope [3]. During auscultation, a medical practitioner examines a patient to identify adventitious (atypical) lung sounds superimposing regular breathing patterns. Examples of commonly heard adventitious lung sounds (ALS) include coarse or fine crackles, pleural rubs, wheezes, and stridors [3,4]. Many RDs causing obstructed or restricted respiratory pathways are characterized by the existence of ALS while breathing. These sound types can be distinctly identified based on their characteristic frequency, pitch, intensity, and energy. For example, both wheezes and stridors are continuous high-pitched sounds occurring at a frequency range of 400 and 500 Hz, respectively. Lower pitched wheeze sounds are also known as rhonchi. High-pitched wheeze sounds may be present as a result of inflamed or narrowed bronchial tubules and are thus an indication of asthma or chronic obstructive pulmonary disease [5]. Stridors usually occur due to tracheal or laryngeal edema [6]. On the other hand, crackles are discontinuous high-pitched (fine) or low-pitched (coarse) waves associated with pneumonia, bronchitis, or heart failure conditions [7]. Similarly, pleural rubs are low-pitched rhythmic sounds associated with inflamed lung lining due to pleural effusion.

Regardless of the type of stethoscopic system used, auscultation is considered one of the safest, easiest, and cheapest examination procedures. Moreover, it provides a patient-friendly and non-invasive solution to monitoring the function of the lungs and other respiratory organs [4]. These qualities are of great value in resource-constrained primary care settings where advanced diagnostic tools are technologies, such as spirometry and radiography, are inaccessible. However, despite being routinely used in healthcare settings, it is a consensus among clinicians that standard pulmonary auscultation has some notable limitations. Firstly, acquiring good auscultatory skills requires extensive training and expertise. Moreover, efficient detection of adventitious sounds is sensitive to the level of experience and auditory acuity of the healthcare professional. Even if auscultation is performed by an expert practitioner, abnormal patterns are sometimes overlooked or misinterpreted during the examination [8]. Thus, subjectivity and inter-variability in observations and interpretations may limit the diagnostic effectiveness of such an approach. These challenges have given rise to the significance of computer-aided auscultation systems that can perform automated identification of ALS and RDs.

Recently, researchers have proposed various artificial intelligence solutions for the identification of adventitious lung sounds. Generally, proposed approaches were based on feature extraction paired with different classification models. At the feature extraction stage, breathing sounds were commonly characterized using several signal processing techniques, including higher-order statistics [9], spectrograms or scalograms [10,11], wavelet transform coefficients [12,13], Hilbert–Huang transform [14], and mel-frequency cepstral coefficients (MFCC) [15]. These feature extraction methods have been utilized in conjunction with the standard machine or deep learning methods such as naive Bayes classifiers [9], k-nearest neighbors [14], support vector machines [16], artificial neural networks (ANN) [13], convolutional neural networks (CNN) [17], and recurrent neural networks (RNN) [18]. Generally, obtained accuracy results varied between 97% and 70.2% for wheeze [19–22], 97.5% and 86% for crackle [23,24], and 99% for normal sound types. In [9], discriminating between normal, fine crackles, coarse crackles, mono-wheezes, and ploy-wheezes sound using a tree-based classifier provided an overall accuracy of 94%.

Despite the various efforts to develop automated adventitious lung sound detection algorithms, their usefulness in identifying RDs is still limited. Recent studies have shown that the presence of abnormal respiratory sounds is not a distinctive characteristic of impaired respiratory functions [3,25]. For example, atypical respiratory sounds might not reflect impaired breathing patterns, and abnormalities do not always translate into audible sounds. These findings necessitate the need for first-hand computer-aided tools that are capable of identifying RDs directly from lung sound signals regardless of the existence of adventitious sounds. In this regard, a recent subclass of studies in the literature focused on exploring different machines and deep learning techniques to perform binary (normal vs. pathological) [26,27], ternary (normal vs. chronic vs. non-chronic) [28], or multi-class [28,29] classification of RDs. Investigated disease conditions included respiratory tract infections, pneumonia, bronchiectasis, bronchiolitis, asthma, and COPD. These studies reported accuracies up to 93.3%, 99%, and 98% for binary, ternary, and multi-class classification, respectively.

It is worth noting that most of the achieved satisfactory results in the context of multi-class classification were based on hybrid deep learning approaches. Despite exhibiting a highly promising performance without the need to incorporate sophisticated feature engineering techniques, training a reliable deep network architecture can be time-consuming and requires significant computational resources. Besides, the training process is iterative, and it involves multiple model parameters and enormous datasets. In clinical contexts, the lack of sufficient high-quality, diverse, and annotated training data is considered among the main limitations. To address the issue of data available, previous studies have employed several data augmentation techniques to over-sample minority classes such as variational autoencoder, adaptive synthetic sampling, and synthetic minority oversampling. However, minority oversampling techniques can introduce data leakage during the validation process, and thus, obtained results might be biased towards high fake accuracies. In fact, the majority of data augmentation approaches are mainly employed to improve the classification performance, without addressing the imperative requirement of using fine-grained datasets that represent the studied population.

In this study, we carried out a multi-class RD classification task considering six different conditions, namely normal, asthma, pneumonia, heart failure, bronchiectasis and bronchitis (BRON disorders), and chronic obstructive pulmonary disease (COPD). To this end, a novel stethoscopic lung sound dataset was collected locally at King Abdullah University Hospital, Jordan University of Science and Technology, Irbid, Jordan. This dataset was complemented by the publicly available ICBHI Challenge database to obtain a more balanced distribution among the respiratory disease classes. In terms of validity in the context of clinical applications, we believe that this data fusion approach provides a better alternative to data augmentation techniques employed in the literature. We propose to tackle the classification problem through a simple yet effective framework utilizing entropy features along with homogeneous ensemble classification methods. Subsequently, a comparative investigation is carried out to compare the proposed ensemble models to several baseline machine learning classifiers that were repeatedly employed in previous works.

The rest of this paper is organized as follows. Section 2 describes the methods used to acquire the lung sound signals along with the mathematical formulation of the entropy features. A brief overview touching on the mathematical groundwork and the implementation of the classification models is also provided. Section 3 presents the experimental results, and Section 4 provides a discussion of these results. Finally, Section 5 concludes this paper.

2. Materials and methods

As shown in Fig. 1, the adopted methodology consists of the following main phases: data acquisition and preparation, feature extraction, construction and training of the ensemble and baseline classifiers, and finally performance evaluation. These steps are detailed below.

2.1. *Lung sound signal acquisition*

The data used in this study incorporated two datasets, both consisting of stethoscopic lung sounds corresponding to different pulmonary diseases. The primary dataset was acquired as part of an ongoing project at King Abdullah University Hospital, Jordan University of Science and Technology, Irbid, Jordan. The second dataset was extracted from the publicly available ICBHI Challenge database to complement the primary dataset [30]. Table 1 provides a demographic overview of the subjects included in each disease category and dataset.

2.1.1. *Primary dataset*

The primary dataset involved a total of 70 patients with different respiratory conditions such as asthma, pneumonia, heart failure, bronchiectasis or bronchitis (BRON disorders), and chronic obstructive pulmonary disease (COPD). Data were also recorded from 35 healthy controls. To ensure an impartial investigation, the age was not a variable of interest in this study. The participants spanned all age groups of children, adults, and elderly. All participants provided written consent after fully understanding the terms of the study and the procedure involved. The study protocol was designed following the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) at King Abdullah University Hospital.

Two thoracic professionals performed the diagnostic and recording procedures. Lung sounds were collected from one of the following standard anterior or posterior chest locations,

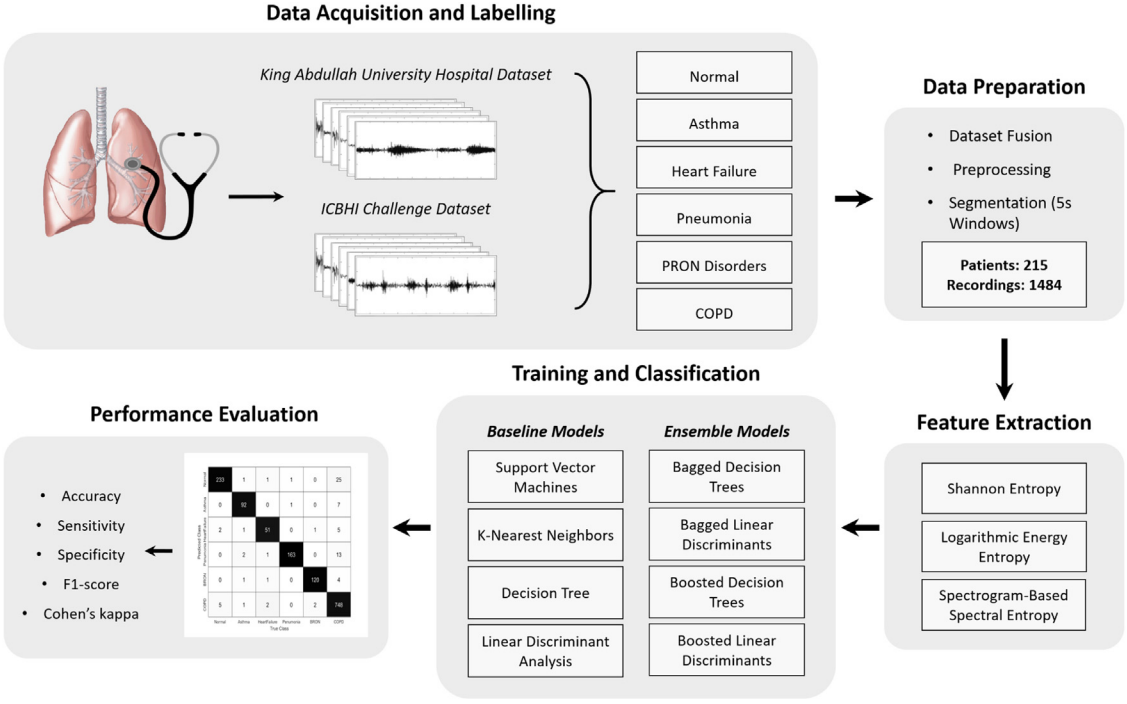

Fig. 1 – The complete analysis framework employed in this study.

Table 1 – Demographic summary of the data within different diagnostic categories.

Category	Primary dataset			ICBHI database		
	# subjects	Age (mean ± SD)	# recordings	# subjects	Age (mean ± SD)	# recordings
Normal	35 (11F, 24M)	43.11 ± 20.12	105	26 (13F, 13M)	12 ± 20.01	135
Asthma	32 (17F, 15M)	45.94 ± 15.55	94	1 (F)	70	4
Heart Failure	21 (9F, 12M)	58.78 ± 18.66	56	–	–	–
Penumonia	5 (2F, 3M)	55.6 ± 10.09	17	6 (2F, 4M)	62 ± 29.11	148
BRON	3 (1F, 2M)	37.33 ± 26.63	7	13 (7F, 6M)	25.04 ± 20.62	116
COPD	9 (1F, 8M)	57.22 ± 9.46	29	64 (16F, 48M)	69.22 ± 8.36	773

Note: SD: Standard Deviation, M: Male, F: Female.

based on the diagnostic requirements of each case: upper left or right, middle left or right, or middle left or right. Recordings were performed on patients in a setting, supine, or prone position and with the stethoscope placed tightly at the location of interest to minimize man-made artifacts. The audio signals were recorded using a single channel stethoscope-based acquisition system (electronic stethoscope 3200, 3M Littmann). The system provides a built-in ambient and frictional reduction technology. All signals were acquired at a sampling rate of 4 kHz using a 16-bit quantizer and were band-limited to an extended frequency range of 20-2 kHz. Moreover, the frequency response between 50 and 500 Hz was attenuated at the acquisition stage to minimize the interference of heart beats sounds [31].

The overall dataset consisted of a total of 308 lung sound recordings, each is of 5 s duration. Such a duration is sufficient to cover at least one respiratory cycle based on the average resting respiration rates for adults (12–20 breaths per minute) and was repeatedly adopted in previous studies. [32–35]. In general, using small-length data windows relaxes the challenge of medical data availability and improves the computational time efficiency of the model. Moreover, training the model on lung sound signals independent of the respiratory cycles simplifies the data curation and labeling process. These advantages are usually favorable in clinical settings and real-time applications. Fig. 2 shows samples of the acquired lung sound signals for normal and RDs conditions with adventitious wheezes or crackles.

2.1.2. *ICBHI challenge dataset*

The ICBHI challenge database was made publicly available for research as part of the scientific challenges announced at the International Conference on Biomedical and Health Informatics 2017. The overall database involved a total of 126 participants and 920 lung sound recordings, spanning different aged groups and respiratory conditions. Recordings were collected by two independent research groups at the (1) Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences, University of Aveiro, Aveiro, Portugal and (2) Papanikolaou General Hospital and the General Hospital of Imathia, Aristotle University of Thessaloniki and the University of Coimbra, Thessaloniki, Greece. The audio signals were recorded using one of the following stethoscope systems: (1) Electronic Stethoscope 3200, 3M Littmann, (2) Classic II SE Stethoscope, 3M Littmann (3) C417 L Professional Lavalier Microphone, AKG HARMAN, and (4) Meditron Master Elite Electronic Stethoscope, Welch Allyn. Recording sites included the trachea and anterior, posterior, and lateral left and right chest locations. In this study, recordings corresponding to one of the respiratory diseases of interest were only selected to complement the primary dataset described earlier. To be consistent with the primary dataset, the respiratory recordings were divided into 5 s nonoverlapping windows in a consecutive manner and without particularly extracting respiratory cycles. Moreover, all signals were re-sampled at a sampling frequency of 4000 Hz. Table 1 lists the number of patients in each disease category and the corresponding number of recordings used from this dataset. Altogether, a total of 110 subjects and 1176 lung sound recordings were included from this dataset. Each recording was considered as an independent sample in the training-validation data.

2.2. *Feature extraction*

In this work, three well-known entropy-based algorithms in information theory were employed for feature extraction, namely the Shannon entropy [36], the logarithmic energy entropy [37], and the spectral entropy [38,39]. Generally, the entropy statistic quantifies the degree of information content or surprise inherited within a signal. Greater entropy values are usually associated with increased randomness and less disorder. The standard definition of Shannon entropy (ShaEn) for a random variable X, with N possible outcomes $\{x_1, x_2, \ldots, x_i, \ldots, x_N\}$, is mathematically formulated as:

$$ShaEn = -\sum_{i=1}^{N} P(x_i)\log(P(x_i)), \tag{1}$$

where $P(x_i)$ is the probability of $x_i \in X$ satisfying:

$$\sum_{i=1}^{N} P(x_i) = 1. \tag{2}$$

Herein, each audio time series signal was realized as a random variable. The underlying probability distribution was non-parametrically estimated by binning the signal to a uniform-bin-width histogram. The optimal bin-width for each time series signal was adaptively determined based on the Scott's normal reference rule [40]:

$$Pin\ \ Width = \frac{3.49\hat{\sigma}}{\sqrt[3]{n}}, \tag{3}$$

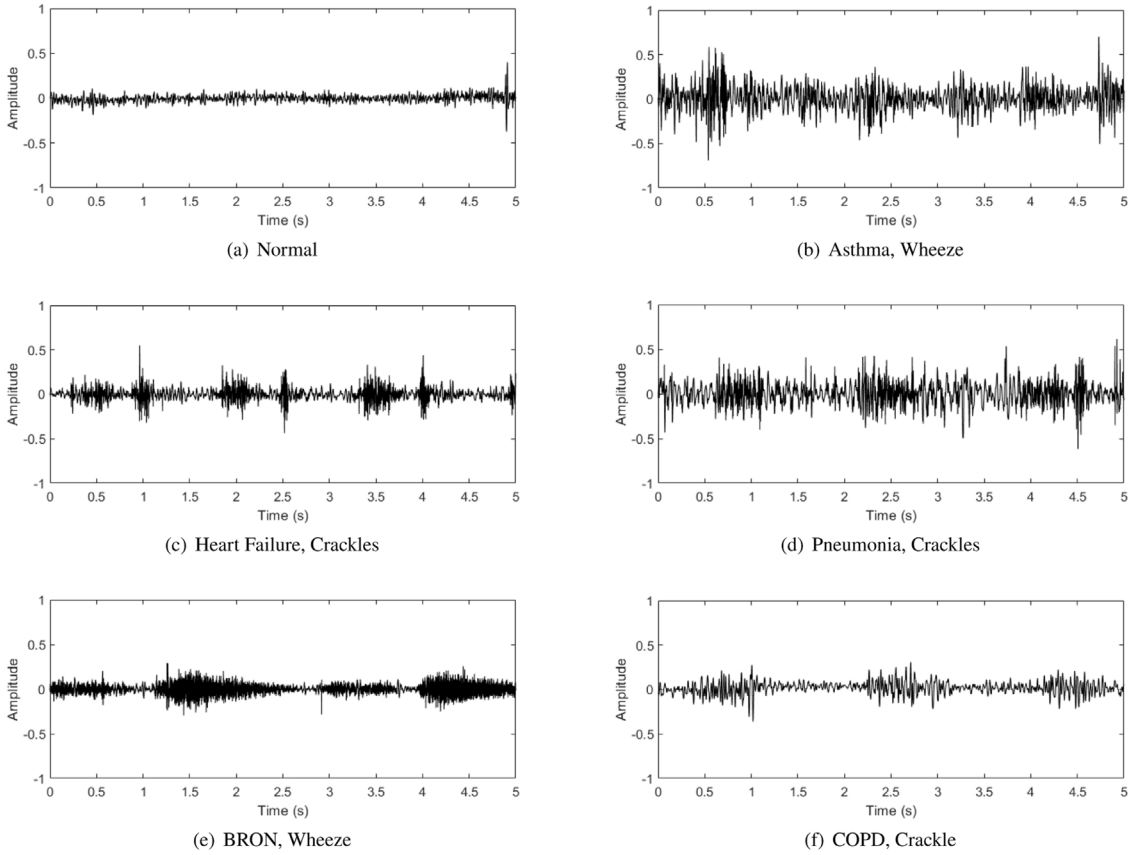

Fig. 2 – Examples of lung sound signals with adventitious wheezes or crackles for different respiratory disease conditions: (a) normal; (b) asthma; (c) heart failure; (d) pneumonia; (e) bronchitis (BRON disorders); (f) chronic obstructive pulmonary disease (COPD).

where σ is the standard deviation of the random variable X or the time series signal, and n is the total number of observations within the time series. In effect, the Shannon entropy can be seen as a measure of the uniformity of the For classification, each 5s signal sample was represented by a 1×3 feature vector formed by concatenating the 3 entropy values. The same feature vectors were used for all classification models, probability distribution, with greater values reflecting a more uniform distribution. Likewise, the logarithmic energy entropy (LogEn) is mathematically expressed as:

$$LogEn = -\sum_{i=1}^{N}(\log P(x_i))^2, \tag{4}$$

The spectral entropy quantifies the degree of randomness within the spectral distribution of the signal. Adopting a similar concept of the Shannon entropy, it thus can be calculated based on a probability density function corresponding to the normalized power spectral density of the signal [41]. In this work, the spectral entropy feature was obtained based on the time-frequency power spectrogram of the lung sound signals. Given a time-frequency spectrogram $S(t, f)$, the probability distribution can be found as:

$$P(m) = \frac{\sum_t S(t,m)}{\sum_f \sum_t S(t,f)} \tag{5}$$

Correspondingly, the spectral entropy (SpeEn) is estimated based on the Shannon definition as:

$$SpeEn = -\sum_{m=1}^{N_F} P(m)\log(P(m)), \tag{6}$$

where N_F is the total number of frequency points. Finally, each feature was independently normalized to the [0, 1] range. This step reduces the standard deviation of the obtained features and thus suppresses the effects of outliers. Give a feature component Y with a minimum bound Y_{min} and a maximum bound Y_{max}, constituent values can be normalized using:

$$Y_{norm} = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \tag{7}$$

The entropy features were extracted independently from each 5 s audio signal. For classification, each 5s signal was represented by a 1×3 feature vector formed by concatenating the three entropy values.

Table 2 – *p*-Values of the Kolmogorov–Smirnov normality test for different disease categories and entropy features.

Category	Feature		
	ShaEn	LogEn	SpeEn
Normal	1.57×10^{-50}	9.35×10^{-51}	1.44×10^{-25}
Asthma	1.68×10^{-50}	9.35×10^{-51}	2.48×10^{-25}
Heart Failure	3.24×10^{-50}	9.35×10^{-51}	3.43×10^{-25}
Penumonia	1.82×10^{-50}	9.35×10^{-51}	8.92×10^{-26}
BRON	1.13×10^{-50}	9.35×10^{-51}	2.05×10^{-25}
COPD	2.64×10^{-50}	9.35×10^{-51}	1.85×10^{-26}

2.3. Statistical significance analysis

A one-way analysis of variance was conducted for each entropy feature independently to test whether different respiratory disease conditions exhibit distinct distributions. For each feature type, each disease category was represented by 56 randomly selected samples. This sample size was selected to match the heart failure category having the smallest number of recordings. Before the statistical analysis, a Kolmogorov–Smirnov normality test was conducted. As shown in Table 2, all feature samples do not follow a normal distribution, and therefore a non-parametric statistical test should be used. Consequently, the Kruskal–Wallis one-way analysis of variance was employed. Moreover, the post-hoc Dunn–Sidák approach was used to perform pairwise comparisons between disease conditions. The significance level for all tests was set to 5%.

2.4. Ensemble classification models

Ensemble classification refers to the process by which multiple classifiers are strategically trained and combined to solve a particular classification problem. With the objective of reducing bias and variance within the classification process, this approach is expected to provide a composite model that performs better than its constituent weak learners [42,43]. In this work, four different ensemble classification models were tested to evaluate the diagnostic effectiveness of discriminating between the six RD conditions based on the entropy of lung sounds. In specific, linear discriminant analysis and decision tree classifiers were employed as weak learned for bagging and boosting ensemble approaches. For each of these classification models, the audio recordings were represented as a feature vector consisting of the three entropy features concatenated.

2.4.1. Bagging ensembles

Bagging is a method suggested by Leo Breiman in 1996 [42,44] by which a classification decision is made by aggregating the predictions of different versions of a weak learner. The weak learners are made distinct by being trained over slightly different samples drawn from the overall dataset. Generating different training samples is achieved through bootstrapping, which means that a subset is selected with replacement, maintaining the size of the original dataset [42,45]. More specifically, given a training dataset X of size N, bootstrapping produces a sequence of training subsets X_i, each consisting of N' responses randomly selected from X with replacement. Each training subset is then used to train a weak classification model, and the final classification decision is made based on simple majority voting [43].

2.4.2. Boosting ensembles

Boosting works similarly to bagging; it is a technique to create a more robust learner by combining multiple variations of a base learner. However, boosting adopts a sequential learning approach in which each weak model considers the misclassifications of previous learners during the training process. As such, this slightly sophisticated approach considers the areas where the system might not be performing well. Generally, boosted ensembles focus on reducing bias as opposed to reducing the variance via bagging.

Adaptive boosting, also known as AdaBoost, is the most applicable implementation of the concept of boosting ensembles. Following Freund and Schapire's introduction of adaptive boosting in 1997, the algorithm was modified to account for multi-class classification as per the AdaBoostM2 implementation [42,45]. Generally, the AdaBoost approach assigns a confidence weigh to each weak learner considering the number of misclassified instances during the testing phase. Moreover, this confidence weight is used to increase the weights for misclassified observations and reduce the weights for the correctly classified ones. For the next learner, the newly weighted subsample is used for training and for calculating its corresponding classification error. The AdaBoostM2 algorithm calculates the classification error of each learner based on the weighted pseudo-loss [45]. For N training observations and M classes, the misclassification error ϵ at the training step t is:

$$\epsilon_t - = \frac{1}{2}\sum_{n=1}^{N}\sum_{m \neq y_n} d_{n,k}^{(t)}(1 - h_t(x_n, y_n) + h_t(x_n, m)) \quad (8)$$

where $h_t(x_n, m)$ is the prediction confidence of the t_{th} learner, $d_{n,k}^{(t)}$ the weight of the n_{th} observation at the t_{th} step, and y_n is a true class label. Accordingly, the confidence weight of the learner t is:

$$\alpha_t = \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}. \quad (9)$$

After each step, the learner's confidence weight α_t can be used to update the weights of the training observations according to:

$$W^{t+1} = W^t e^{\pm\alpha_t} \quad (10)$$

After training T weak learners, class predictions for new features can be found using:

$$f(x) = \sum_{t=1}^{T}\alpha_t h_t(x). \quad (11)$$

For both bagging and boosting models, the number of training subsets (i.e. the number of learners) and the hyperparameters of each weak learner were iteratively optimized during the training process, as described in Section 2.6. The size of each training subset was set to $\sqrt{N}$.

2.5. Baseline classification models

The performance of the ensemble models was compared to that of benchmark classifiers that are popular in lung sound analysis literature: decision trees (DT), linear discriminant analysis (LDA), support vector machines (SVM), and k-nearest neighbors (KNN). These models were implemented following their binary realizations, and the multi-class classification problem was handled through a one-versus-all error-correcting output code (ECOC) scheme.

2.5.1. Decision trees (DT)

Decision tree (DT) learning is one of the most easily intelligible prediction models in the field of machine learning. In classification contexts, this approach adopts a tree structure consisting of root nodes, internal leaf nodes, and branches. Each root node in this structure represents a feature, while the leave nodes correspond to class labels. Correspondingly, the branches represent the conjunctions leading features to class labels. The performance of DT models mainly depends on how well the tree structure is constructed from the training data. Once identified, the model can be used to predict the class labels of new samples. In this work, the Gini's diversity index was used as the root node split criteria [46]. Parameters controlling the depth of the model included the maximum number of splits (branches) and the minimum number of leaf nodes. For hyperparameter tuning, structures with 1, 2, or 3 leaf nodes and 1 or 2 branches were considered.

2.5.2. Linear discriminant analysis (LDA)

For a binary classification problem, the linear discriminant analysis (LDA) projects the multidimensional feature vector onto a dimensionality reduced hyperplane. The most appropriate projection direction can be determined based on different criteria to best separate or characterize the two classes present within the training data. This hyperplane is then used as a decision boundary to classify new objects. As employed in this study, the fisher's linear discriminant approach simultaneously considers two criteria to estimate the separation hyperplane: (1) maximizing the distance between the class means (between-class scatter) and (2) minimizing the variance within classes (within-class scatter) [47].

2.5.3. Support vector machines (SVM)

Support vector machine (SVM) is one of the most popular and robust supervised prediction algorithms. Being mainly developed to solve binary classification problems, SVM adopts a linear and non-probabilistic approach to find a hyperplane that maximizes the separation between two classes [47]. Linear SVM models can be effectively tweaked to perform non-linear classification through the usage of kernel functions that map the input features into a higher dimensional space. Herein, an SVM classifier with a radial basis function was employed [48]. The sequential minimal optimization algorithm was used to train the model [49]. For fine parameter tuning, we explored different variations of the kernel scale and soft margin constants, with positive values over the interval [− 3, 3], with a step of 1.

2.5.4. K-nearest neighbors (KNN)

The K-nearest neighbors (KNN) is considered one of the simplest and commonly used distance-based algorithms to solve supervised classification problems. Using a lazy case-based learning approach, KNN models only store feature vectors and their corresponding class labels at the training stage. Subsequently, new objects can be classified based on the attributes of the nearest K objects in the training set [50]. In this work, the K closest training samples were selected in the feature space based on the Euclidean distance measure [51]. The classification decision was then made based on the majority voting of class labels among the identified neighbors. Since the number of nearest neighbors (K) is considered a key tuning parameter, variations of its value over the range of [1, 10] with a step of 1 were considered for hyperparameter optimization.

2.6. Performance evaluation criteria

To get a robust estimation of the overall classification performance, the models were trained and tested using 10-folds cross validation. The folds were split in a random stratified manner to consider the imbalanced property of the data. Each fold had the same number of sub-samples with a sub-class distribution matching the distribution in the complete dataset. For each of the 10 validation iterations, a single fold was retained for testing, and the remaining nine folds were used for hyperparameter tuning and training. The confusion matrices and the performance evaluation metrics were obtained independently for each validation fold. Hyperparameter tuning was performed using Bayesian optimization with a cross-validation loss cost function. For each ensemble or baseline classification model, all eligible parameters were optimized. Each of the weak learners in the ensemble models was also optimized independently in each fold. This approach is a standard in machine learning application that allows testing and training on the overall dataset. It results in a generalized and less biased estimate of the model performance in comparison to the simple train-test split.

The performance evaluation metrics included accuracy, sensitivity, specificity, F1-score, and Cohen's kappa coefficient (κ). Provided below are the confusion matrix-based definitions for each of these metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{13}$$

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{15}$$

$$\kappa = \frac{Po - Pe}{1 - Pe}, \tag{16}$$

where the true positives (*TP*) and the true negatives (*TN*) represent the count of correctly classified audio signals, while the false positives (*FP*) and false negatives (*FN*) represent the number of signals incorrectly classified. *Po* is the relative agreement between raters, and it is equivalent to the classification accuracy, while *Pe* is the hypothetical probability of agreement by chance and can be calculated as [52]:

$$Pe = \frac{(TP + FP)(TP + FN) + (TN + FN)(TN + FP)}{(TP + TN + FP + FN)^2} \quad (17)$$

3. Experimental results

The investigated models were trained tested using a computer with an Intel(R) Core TM i7-8750H (Intel Corporation, Santa Clara, USA), a 2.20 GHz CPU, 16 GB of RAM, and an NVIDIA GeForce GTX 1050 TI GPU (NVIDIA, California USA). The training process for all the ensemble models lasted around 15.38 min, while the baseline models needed around 2.02 min to train in total. The complete analysis was performed via Matlab software (R2020a, Natick, Massachusetts, USA).

3.1. Feature characterization and statistical analysis

Fig. 3 visualizes the distribution of the Shannon entropy, logarithmic energy entropy, and spectral entropy features for different RD conditions. The Kruskal–Wallis statistical test showed a statistically significant difference in the Shannon entropy feature between the RD conditions, chi-square $x^2(4) = 47.32$, $p = 4.89 \times 10^{-9}$. The effect of the RD factor on the Log energy entropy feature was also significant: $x^2(4) = 62.18$, $p = 4.31 \times 10^{-12}$. Similarly, the spectral entropy feature exhibited significant differences between different disease conditions: $x^2(4) = 47.08$, $p = 5.48 \times 10^{-9}$. The statistical test results of the pair-wise comparisons between different conditions are given in Table 3. The tabulated values represent the P-values obtained after applying Dunn–Sidák correction for multiple comparisons. Generally, results show that the chosen entropy features are highly distinctive of the different disease conditions, and thus, provide a solid ground for the classification analysis.

3.2. Performance of the ensemble classifiers

Table 4 summarizes the performance of the bagged decision tree (Bagged-DT), bagged linear discriminant analysis (Bagged-LDA), boosted decision tree (Boosted-DT), and boosted linear discriminant analysis (Boosted-LDA). The provided values represent the averaged performance metrics across validation folds for each respiratory disease class. To elaborate more on the obtained results, Fig. 4 illustrates the confusion matrices obtained through validating the bagging and boosting ensemble models. The represented matrices correspond to the sum of the matrices obtained from the ten validation folds. For the bagging method, the results show that Bagged-DT model outperformed Bagged-LDA model for all the metrics, with an average classification accuracy of 96.34%. The Bagged-LDA model provided a relatively small overall accuracy of 86.77%.

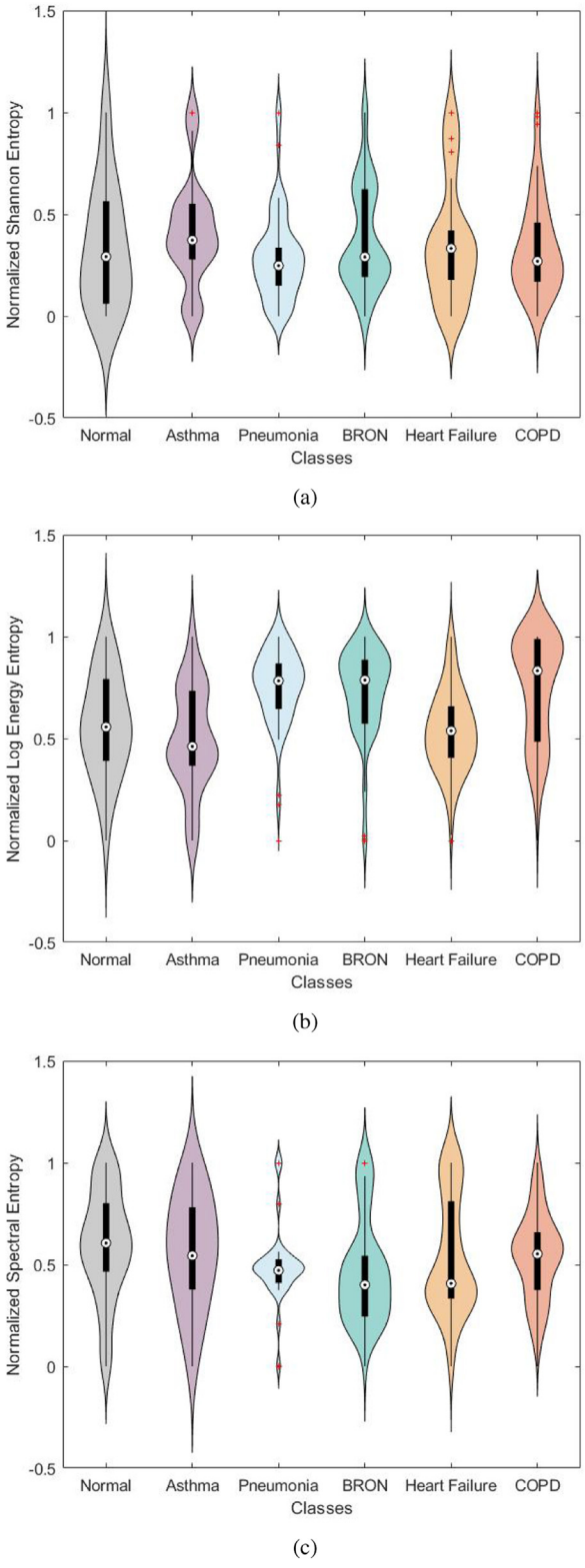

Fig. 3 – Distribution of the: (a) Shannon entropy; (b) logarithmic energy entropy; (c) spectral entropy features among different respiratory disease conditions.

Using the Bagged-DT model, the highest Cohen's kappa coefficient of 84.81% and was observed for the normal class. The highest F-score of 9.1.9 was observed for the COPD class,

Table 3 – Statistical analysis of pair-wise feature differences between different respiratory disease conditions. Provided are the *p*-values after Dunn–Sidák correction for multiple comparisons.

Conditions	Feature		
	Shannon entropy	Log energy entropy	Spectral entropy
Normal/Asthma	0.250	0.982	0.965
Normal/heart failure	<0.001	1.000	1.000
Normal/pneumonia	1.000	0.003	0.808
Normal/BRON	1.000	0.058	0.006
Normal/COPD	0.475	<0.001	0.007
Asthma/heart failure	0.089	0.999	0.525
Asthma/pneumonia	0.032	<0.001	0.053
Asthma/BRON	0.183	0.001	0.303
Asthma/COPD	1.000	<0.001	0.334
Heart failure/pneumonia	<0.001	0.001	0.998
Heart failure/BRON	<0.001	0.025	<0.001
Heart failure/COPD	0.034	<0.001	<0.001
Penumonia/BRON	1.000	1.000	<0.001
Penumonia/COPD	0.085	0.975	<0.001
BRON/COPD	0.374	0.445	1.000

Note: Underlined values indicate significant differences between disease conditions ($p < 0.05$).

while the highest specificity was observed for the asthma class. Pneumonia and BRON classes showed the highest sensitivity and specificity values of 94.63 and 89.38, respectively. Using the boosting ensemble approach, the overall performance of the decision tree-based model notably improved. In specific, the overall accuracy, sensitivity, specificity, F1-score, and Cohen's kappa coefficient metrics improved to 98.27%, 95.28%, 98.9%, 93.61%, and 92.28%, respectively. On the contrary, boosting provided a slightly dropped performance for the linear discriminant analysis-based model. In terms of computational time, the Boosted-LDA model has the lowest training time of 3.13 min, followed by the Bagged-LDA model at 3.35 min, then

Table 4 – Performance evaluation metrics of different classifiers using the features derived from the stride signals (average ± standard deviation values across the 10 validation folds).

Model	Class	Performance evaluation				
		Accuracy	Sensitivity	Specificity	F1-Score	Kappa (κ)
Bagged decision tree	Normal	95.62%	92.92%	96.14%	87.44%	84.81%
	Asthma	97.71%	78.33%	99.06%	81.46%	80.26%
	Heart failure	97.84%	60.00%	99.30%	67.12%	66.05%
	Pneumonia	97.03%	94.63%	97.34%	87.76%	86.09%
	BRON	98.38%	88.59%	99.26%	89.89%	89.01%
	COPD	91.44%	90.03%	93.11%	91.90%	82.84%
	Average	**96.34%**	**84.08%**	**97.37%**	**84.26%**	**81.51%**
Bagged linear discriminant	Normal	79.99%	24.17%	90.76%	27.82%	16.67%
	Asthma	91.65%	15.22%	97.04%	19.98%	16.01%
	Heart failure	95.36%	28.00%	97.97%	30.92%	28.57%
	Pneumonia	88.82%	52.65%	93.33%	51.57%	45.35%
	BRON	91.64%	01.67%	99.78%	03.08%	02.48%
	COPD	73.17%	89.41%	54.06%	78.31%	44.54%
	Average	**86.77%**	**35.19%**	**88.82%**	**35.28%**	**25.61%**
Boosted decision tree	Normal	97.64%	97.08%	97.75%	93.17%	91.76%
	Asthma	99.06%	94.00%	99.42%	93.08%	92.58%
	Heart failure	99.06%	91.00%	99.37%	88.18%	87.70%
	Pneumonia	98.78%	98.79%	98.78%	94.95%	94.27%
	BRON	99.40%	97.56%	99.56%	96.36%	96.03%
	COPD	95.69%	93.27%	98.53%	95.89%	91.36%
	Average	**98.27%**	**95.28%**	**98.90%**	**93.61%**	**92.28%**
Boosted linear discriminant	Normal	78.17%	28.33%	87.78%	29.42%	16.62%
	Asthma	92.19%	13.11%	97.77%	18.10%	14.82%
	Heart failure	95.83%	21.00%	98.74%	26.74%	24.86%
	Pneumonia	87.81%	37.50%	94.08%	40.61%	34.07%
	BRON	91.37%	3.21%	99.33%	5.84%	4.37%
	COPD	73.11%	89.78%	53.51%	78.30%	44.41%
	Average	**86.41%**	**32.15%**	**88.54%**	**33.17%**	**23.19%**

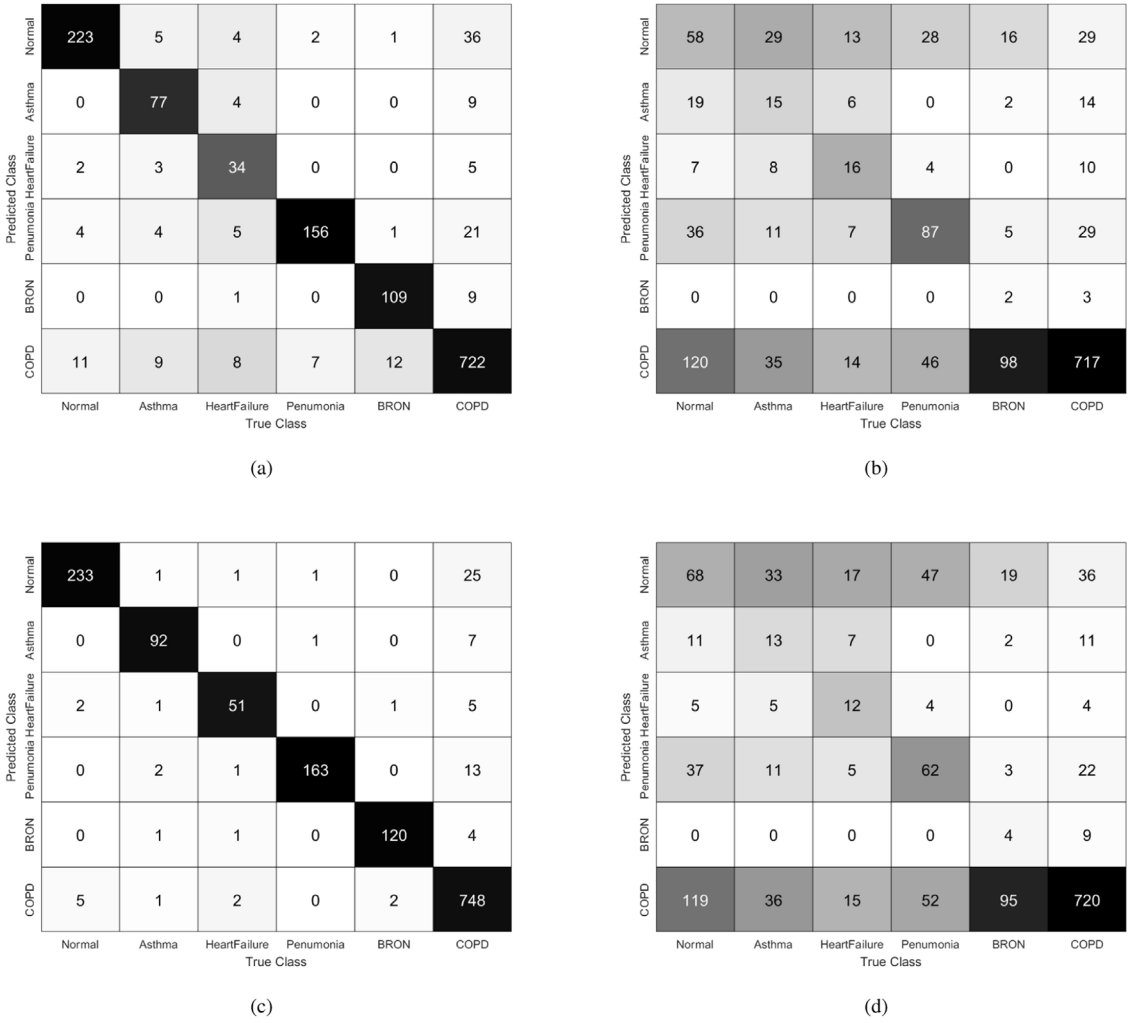

Fig. 4 – Confusion matrices for the ensemble classification models: (a) bagged decision trees; (b) bagged linear discriminant analysis; (c) boosted decision trees; (d) boosted linear discriminant analysis.

the Bagged-DT model at 3.60 min. Despite its superior performance, the Boosted-DT model required the highest training time of 5.32 min.

3.3. *Performance of the baseline classifiers*

Table 5 summarizes the performance of the baseline classification methods. The tabulated values represent the average percentage of the classification accuracy, sensitivity, specificity, F-score, and Cohen's kappa coefficient across validation folds and classes. Fig. 5 illustrates the corresponding confusion matrices for these models. The represented matrices correspond to the sum of the matrices obtained from the ten validation folds. The results show that the SVM model provided the highest detection accuracy at an average of 98.20%, followed by KNN at 97.04%, then DT at 93.49%. Moreover, the highest of Cohen's kappa coefficients, F1-Score, and specificity were achieved by the SVM model at an average of 91.5%, 92.89%, and 98.55%, respectively. On the contrary, the highest sensitivity of 95.49% was achieved by the KNN model. The LDA model produced the lowest accuracy of 86.41%. In addition, all other metrics significantly dropped to 32.77%, 88.01%, 33.31%, and 23.05% for the sensitivity, specificity, F1-score, and Cohen's kappa coefficients respectively. Providing an overall superior performance in comparison to the other baseline methods, the SVM model had the highest training time of 1.32 min. The training time required by the remaining models was significantly lower. Particularly, 0.28, 0.2, and 0.18 min, training duration where required for KNN, DT, LDA.

4. Discussion

Computer-aided detection of respiratory diseases can expedite diagnostic and treatment decisions and support the study of physiological patterns associated with various respiratory pathologies. In this work, we propose to combine entropy-based features and homogeneous ensemble classifiers to perform multi-class classification of a wide range of respiratory diseases.

Table 5 – Performance evaluation metrics of the baseline classification models for different respiratory disease classes.

Model	Class	Performance evaluation				
		Accuracy	Sensitivity	Specificity	F1-Score	Kappa (κ)
Decision tree	Normal	91.64%	82.92%	93.33%	76.18%	71.19%
	Asthma	95.49%	66.33%	97.55%	65.97%	63.60%
	Heart failure	97.57%	71.00%	98.6%	68.20%	66.95%
	Pneumonia	95.62%	84.26%	97.05%	81.37%	78.91%
	BRON	95.15%	75.58%	96.91%	71.82%	69.18%
	COPD	84.91%	81.55%	88.84%	85.39%	69.84%
	Average	**93.49%**	**77.44%**	**95.42%**	**75.50%**	**70.68%**
Linear discriminant	Normal	82.41%	20.42%	94.37%	26.34%	17.88%
	Asthma	92.32%	13.22%	97.91%	19.02%	15.71%
	Heart failure	95.49%	31.33%	97.97%	34.92%	32.69%
	Pneumonia	88.21%	12.65%	97.65%	19.31%	14.93%
	BRON	91.71%	00.00%	100.0%	00.00%	00.00%
	COPD	65.63%	94.52%	31.65%	74.82%	27.43%
	Average	**86.41%**	**32.77%**	**88.01%**	**33.31%**	**23.05%**
Support vector machine	Normal	98.45%	96.25%	98.87%	95.24%	94.32%
	Asthma	98.79%	84.78%	99.78%	90.14%	89.50%
	Heart failure	98.99%	79.33%	99.79%	85.32%	84.82%
	Pneumonia	99.33%	97.61%	99.54%	97.02%	96.64%
	BRON	98.72%	94.29%	99.12%	92.49%	91.79%
	COPD	94.95%	95.89%	93.86%	95.36%	89.82%
	Average	**98.20%**	**91.52%**	**98.55%**	**92.89%**	**91.50%**
K-Nearest neighbors	Normal	97.31%	98.33%	97.11%	92.31%	90.69%
	Asthma	98.66%	96.00%	98.85%	90.68%	89.97%
	Heart failure	99.06%	94.67%	99.23%	88.57%	88.09%
	Pneumonia	97.57%	99.41%	97.34%	90.32%	88.96%
	BRON	97.64%	97.56%	97.65%	87.46%	86.18%
	COPD	91.71%	85.17%	99.41%	91.73%	83.53%
	Average	**97.04%**	**95.49%**	**98.3%**	**90.41%**	**88.17%**

As imperative to all machine learning frameworks, the feature extraction stage aims at providing a better representation of the pattern underlying the data. Thus, optimized feature extraction is key to building effective classification models and improving the model's predictive accuracy. Respiratory sounds are random and non-linear signals that are highly complex in nature; particularly, due to the changing lung volume. These properties are evident for both healthy and pathological subjects but become more apparent in pathological lung sounds [4]. Thus, this study employed different variations of the concept of entropy as discriminating features to model non-linearities and randomness in both temporal and spectral domains. The results showed that the investigated disease conditions exhibited distinct distributions for the Shannon entropy, logarithmic energy entropy, and spectrogram-based spectral entropy features. Moreover, our statistical analysis results showed that the selected entropy features positively represented characteristic changes in the audio signals among different disease conditions. For respiratory disease detection, we employed four types of ensemble classifiers: Bagged-DT, Bagged-LDA, Boosted-DT, and Boosted-LDA. In order to highlight the significance of ensembled predictions, we also considered four baseline models, namely DT, LDA, SVM, and KNN, for comparison. Within the ensemble models, Boosted-DT achieved the highest performance, with an average classification accuracy of 98.27%. At the same time, the SVM model provided the best performance among the baseline methods and the secondary performance overall. For all the explored performance metrics, the performance of the DT-based ensembles notably outperformed that of the baseline DT model. On average, Boosted-DT showed a 23.99% and improvement in the F1-score and 30.56% improvement in Cohen's kappa coefficient. The Bagged-DT model exhibited a slightly decreased improvement of 11.60% and 15.32% for the F1-score and Cohen's kappa metrics, respectively. On the contrary, the LDA classifier showed the worst performance for the bagging, boosting, and baseline methods, with an approximate accuracy slightly above 86%. This could be attributed to the sensitivity of the LDA classifiers to the imbalanced property of the data, as evident from the relatively higher number of false positives associated with the COPD class in Figs. 4 and 5.

To the best of our knowledge, only three studies in literature have attempted to perform six-class classification of respiratory pathologies. These studies have used the ICBHI dataset for model training and validation. Table 6 presents a comparative summary of these studies. It is worthy to note that direct quantitative comparison is not applicable due to methodological variations in the employed feature extraction and classification algorithms, model validation approaches, and datasets (sample size and respiratory disease classes involved). However, the obtained classification results in this study were numerically comparable to or slightly higher better than the results reported in [28,29]

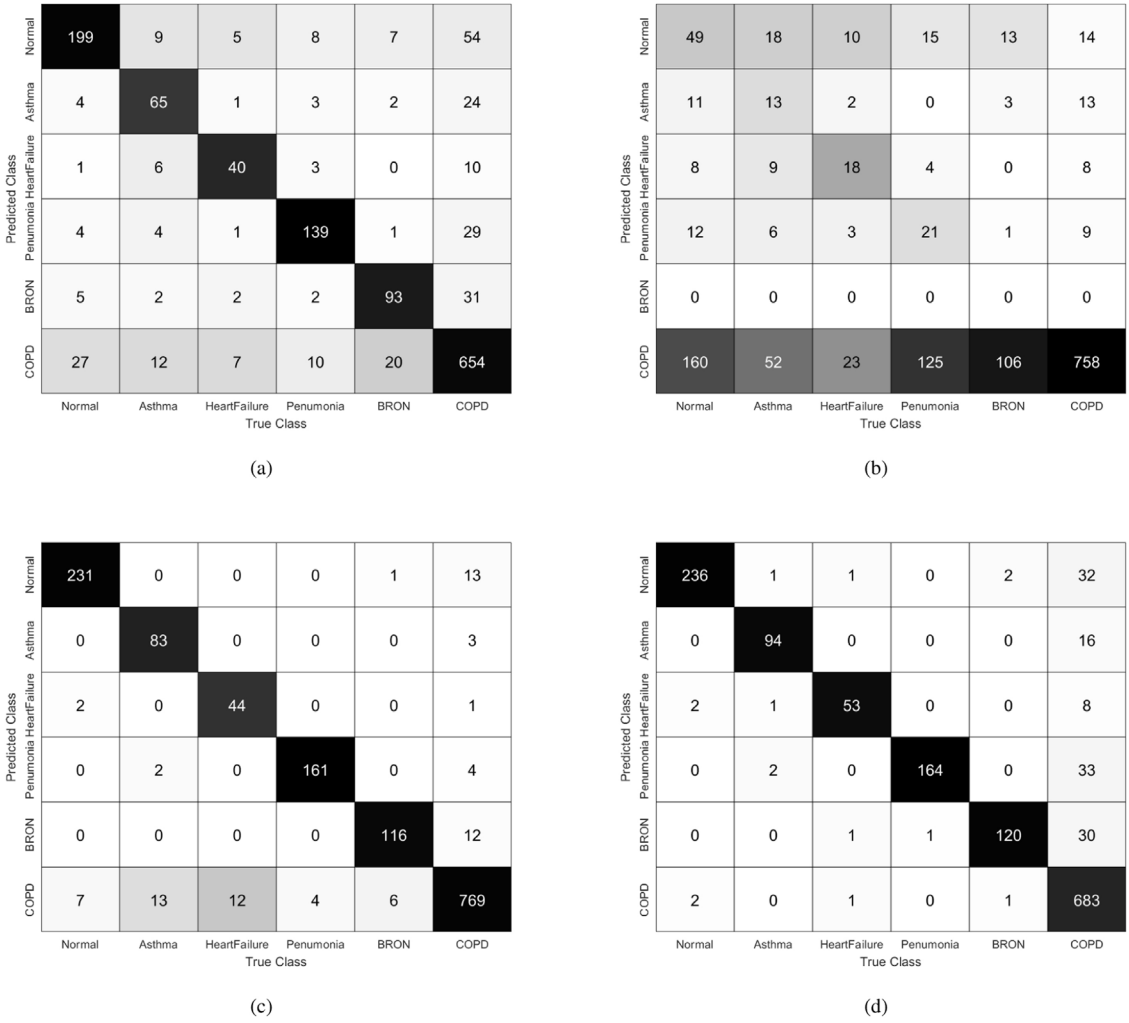

Fig. 5 – Confusion matrices for the baseline classification models: (a) decision tree; (b) linear discriminant analysis; (c) support vector machines; (d) k-nearest neighbors.

Table 6 – Comparative summary of multi-class respiratory disease classification literature.

Study	Database	Classes	Method	Performance		
				Accuracy	F1-Score	Kappa (κ)
M. Ordas et al. (2020) [28]	ICBHI Database (with data augmentation)	6 classes (normal, bronchiolitis, bronchiectasis, pneumonia, URTI, and COPD)	Mel-Spectrogram + CNN	99.00%	90.00%	–
V. Basu et al. (2020) [29]	ICBHI Database (with data augmentation)	6 classes (normal, bronchiolitis, bronchiectasis, pneumonia, URTI, and COPD)	MFCC Features + RNN	95.67%	95.66%	94.74%
Z. Tariq et al. (2019) [53]	ICBHI Database (with data augmentation)	6 classes (normal, bronchiolitis, bronchiectasis, pneumonia, URTI, and COPD)	Spectrogram + CNN	97%	–	–
Our Study	King Abdullah University Hospital + ICBHI Database	6 classes (Normal, BRON disorders, pneumonia, **asthma**, **heart failure**)	Entropy features + Boosted DT	98.27%	93.61%	92.28%

5. Conclusion

To sum up, this paper investigated the use of ensemble classifiers with a dataset of lung sounds obtained via a stethoscope to perform multi-class classification. The dataset included a total of 215 subjects with 308 clinically acquired lung sound recordings, in addition to the 1176 recordings obtained from the ICBHI Challenge database. Entropy was the central feature representation used, more specifically Shannon entropy, logarithmic energy entropy, and spectrogram-based spectral entropy. Decision trees and linear discriminant

classifiers were employed as base learners to build bootstrap aggregation and adaptive boosting ensemble methods and were compared with the results of classical machine learning techniques. Bayesian optimization was applied to all aforementioned algorithms to find optimal training parameters. Experimental results showed that the ensemble classification models generally outperformed the baseline methods commonly employed in the literature. In specific, the boosted decision tree model exhibited the best performance as demonstrated by its highest accuracy (98.20%), sensitivity (91.5%), and specificity (98.55%).

Authors' contribution

Dr. Luay Fraiwan: project administration and supervision; fund acquisition, Abu Dhabi University Fund; conceptualization; methodology; writing. Eng. Omnia Hassanin: conceptualization, creating training and testing models; software: Matlab programming (ensemble classifier); methodology; formal analysis; writing. Dr. Mohammed Fraiwan: investigation: building the data acquisition system and measurement protocol; formal analysis: analyzing the recorded sound signal and clinical data verification; funding acquisition: JUST research grant (for data collection); resources management. Dr. Basheer Khaswaneh and Dr. Ali Bnian: data Curation (annotating clinical sounds); data collection and subjects' management. Eng. Mohanad Alkhodari: software: Matlab programming (data preparation and validation); formal analysis.

Funding

This research is supported by the Deanship of Scientific Research at Jordan University of Science and Technology, Jordan, grant no. 20180356, and the Office of Research and Sponsored Programs (ORSP) Abu Dhabi University's, UAE.

Conflict of interest

The authors declare no conflict of interest associated with this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.bbe.2020.11.003.

REFERENCES

[1] Bousquet JNGK. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. Geneva: World Health Organization; 2007. p. c2007.
[2] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLOS Med 2006;3:1–20.
[3] Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. Ann Thorac Med 2015;10:158.
[4] Andrès E, Gass R, Charloux 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;11:89.
[5] Pramono RXA, Imtiaz SA, Rodriguez-Villegas E. Evaluation of features for classification of wheezes and normal respiratory sounds. PLOS ONE 2019;14:e0213659.
[6] Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: advances beyond the stethoscope. Am J Respir Crit Care Med 1997;156:974–87.
[7] Reichert S, Gass R, Brandt C, Andrès E. Analysis of respiratory sounds: state of the art. Clin Med Circ Respir Pulmon Med 2008;2. CCRPM-S530.
[8] Shi L, Du K, Zhang C, Ma H, Yan W. Lung sound recognition algorithm based on vggish-bigru. IEEE Access 2019;7:139438–49.
[9] Naves R, Barbosa BH, Ferreira DD. Classification of lung sounds using higher-order statistics: a divide-and-conquer approach. Comput Methods Programs Biomed 2016;129:12–20.
[10] Acharya J, Basu A, Ser W. Feature extraction techniques for low-power ambulatory wheeze detection wearables. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2017. p. 4574–7.
[11] Gautam N, Pokle S. Wavelet scalogram analysis of phonopulmonographic signals. Int J Med Eng Informatics 2013;5:245–52. http://dx.doi.org/10.1504/IJMEI.2013.055700
[12] Kahya YP, Yeginer M, Bilgic B. Classifying respiratory sounds with different feature sets. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society 2006;2856–9.
[13] Orjuela-Ca nón Alvaro D, Gómez-Cajas Diego F, Jiménez-Moreno Robinson. Artificial neural networks for acoustic lung signals classification. Ibero-American Congress on Pattern Recognition 2014;214–21.
[14] Serbes G, Sakar CO, Kahya YP, Aydin N. Pulmonary crackle detection using time-frequency and time-scale analysis. Dig Signal Process 2013;23:1012–21.
[15] Bahoura M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Comput Biol Med 2009;39:824–43.
[16] Jin F, Sattar F, Goh DY. New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. Neurocomputing 2014;123:362–71.
[17] Aykanat M, Kiliç Ö, Kurt B, Saryal S. Classification of lung sounds using convolutional neural networks. EURASIP J Image Video Process 2017;1–9.
[18] Messner E, Fediuk M, Swatek P, Scheidl S, Smolle-Juettner F, Olschewski H, et al. Multi-channel lung sound classification with convolutional recurrent neural networks. Comput Biol Med 2020;122:103831.
[19] Bokov P, Mahut B, Flaud P, Delclaux C. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. Comput Biol Med 2016;70:40–50.
[20] Oweis R, Abdulhay E, Khayal A, Awad A. An alternative respiratory sounds classification system utilizing artificial neural networks. Biomed J 2014;38.
[21] Lin BS, Wu HD, Chen S. Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. J Healthc Eng 2015;6:649–57.
[22] Riella R, Nohama P, Maia J. Method for automatic detection of wheezing in lung sounds. Braz J Med Biol Res 2009;42:674–84.

[23] Zhang K, Wang X, Han F, Zhao H. The detection of crackles based on mathematical morphology in spectrogram analysis. Technol Health Care: Off J Eur Soc Eng Med 2015;23:89–94.
[24] Serbes G, Sakar CO, Kahya YP, Aydin N. Pulmonary crackle detection using timeâfrequency and time scale analysis. Dig Signal Process 2013;23:1012–21.
[25] Xavier G, Melo-Silva C, Gaio E, Amado V. Accuracy of chest auscultation in detecting abnormal respiratory mechanics in the immediate postoperative period after cardiac surgery. J Bras Pneumol 2019;45. http://dx.doi.org/10.1590/1806-3713/e20180032
[26] Perna D. Convolutional neural networks learning from respiratory data. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2018;2109–13.
[27] Islam MA, Bandyopadhyaya I, Bhattacharyya P, Saha G. Multichannel lung sound analysis for asthma detection. Comput Methods Programs Biomed 2018;159:111–23.
[28] García-Ordás M, Benítez Andrades J, García I, Benavides C, Alaiz Moreton H. Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. Sensors 2020;20.
[29] Basu V, Rana S. Respiratory diseases recognition through respiratory sound with the help of deep neural network. 2020 4th International Conference on Computational Intelligence and Networks (CINE) 2020;1–6.
[30] Rocha B, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, et al. A respiratory sound database for the development of automated classification. International Conference on Biomedical and Health Informatics 2017;33–7.
[31] Andrès E, Gass R, Charloux 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;11:89–106.
[32] Nuckowska M, Gruszecki M, Kot J, Wolf J, Guminski W, Frydrychowski A, et al. Impact of slow breathing on the blood pressure and subarachnoid space width oscillations in humans. Sci Rep 2019;9:1–13.
[33] Lapi S, Lavorini F, Borgioli G, Calzolai M, Fontana G. Respiratory rate assessments using a dual-accelerometer device. Respir Physiol Neurobiol 2014;191:60–6.
[34] Zhang W, Lei W, Xu X, Xing X. Improved music genre classification with convolutional neural networks. Interspeech 2016 2016;3304–8.
[35] Chen Q, Zhang W, Tian X, Zhang X, Chen S, Lei W. Automatic heart and lung sounds classification using convolutional neural networks. 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) 2016;1–4.
[36] Shannon CE. A mathematical theory of communication. Bell Syst Techn J 1948;27:379–423.
[37] Aydin S, Saraoglu HSK. Log energy entropy-based eeg classification with multilayer neural networks in seizure. Ann Biomed Eng 2009;37:2626–30.
[38] Helakari H, Kananen J, Huotari N, Raitamaa L, Tuovinen T, Borchardt V, et al. Spectral entropy indicates electrophysiological and hemodynamic changes in drug-resistant epilepsy a ‘‘a multimodal mreg study. NeuroImage: Clin 2019;22:1–12.
[39] Vanluchene AL, Vereecke H, Thas O, Mortier EP, Shafer SL, Struys MM. Spectral entropy as an electroencephalographic measure of anesthetic drug effect: a comparison with bispectral index and processed midlatency auditory evoked response. Anesthesiology 2004;101:34–42.
[40] He K, Meeden G. Selecting the number of bins in a histogram: a decision theoretic approach. J Stat Plan Inference 1997;61:49–59.
[41] Vakkuri A, Yli-Hankala A, Talja P, Mustola S, Tolvanen-Laakso H, Sampson T, et al. Time-frequency balanced spectral entropy as a measure of anesthetic drug effect in central nervous system during sevoflurane, propofol, and thiopental anesthesia. Acta Anaesthesiol Scand 2004;48:145–53.
[42] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55:119–39.
[43] Polikar R. Ensemble based systems in decision making. IEEE Circuits Syst Mag 2006;6:21–45.
[44] Breiman L. Bagging predictors. Mach Learn 1996;24:123–40.
[45] Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybern Part C Appl Rev 2012;42:463–84.
[46] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 1991;21:660–74.
[47] Kamruzzaman J, Begg RK. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. IEEE Trans Biomed Eng 2006;53:2479–90.
[48] Xu Y, Zomer S, Brereton R. Support vector machines: a recent method for classification in chemometrics. Crit Rev Anal Chem 2006;36:177–88.
[49] Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines. Adv Kernel Methods-Support Vector Learn 1998;208:1–21.
[50] Phan TN, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors 2017;18:1–18.
[51] Chen S, Shen B, Wang X, Yoo S. A strong machine learning classifier and decision stumps based hybrid adaboost classification algorithm for cognitive radios. Sensors (Basel Switzerland) 2019;19:1–15.
[52] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[53] Tariq Z, Shah SK, Lee Y. Lung disease classification using deep convolutional neural network. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2019;732–5. http://dx.doi.org/10.1109/BIBM47256.2019.8983071