Separation of Cardiopulmonary Sound Signals for Classification of Respiratory Diseases

Ruxin Zheng* and Saeid Sanei†
* Imperial College London, London, United Kingdom
E-mail: 18272020090@163.com Tel: +86-15729321697
† Imperial College London, London, United Kingdom
E-mail: s.sanei@imperial.ac.uk Tel: +44-7919098002

***Abstract*—Due to having the mixture of heart and lung sounds, the traditional automatic cardiopulmonary sound signal analyses have some challenges in achieving sufficient classification accuracy and robustness. The aim of this study is to improve the performance of respiratory disease classification model by cardiopulmonary sound separation. We used 920 auscultated audio signals involving 127 subjects who were healthy or had different respiratory conditions. For separation of lung sounds from heart sounds and other interferences we have developed an adaptive subspace analysis which iteratively maximizes the kurtosis. The separated lung sound is then classified using random forest classifier, and the classification performance is evaluated by cross-validation. The results show that the classification performance of cardiopulmonary sound for different age groups (infants and adults) significantly improves.**

I. Introduction

Lung diseases mainly including chronic obstructive pulmonary disease (COPD), lung cancer and asthma are the major global public health problem. These diseases not only affect the quality of life of patients, but also place an enormous burden on global healthcare systems. Pulmonary sound classification plays an important role in clinical diagnosis, personalized treatment, public health monitoring, medical research and teaching, automation and telemedicine. By improving the accuracy and efficiency of pulmonary sound classification, the management and treatment of respiratory diseases can be significantly improved, and the quality of patients' lives can be improved.

Lung sounds are usually heard or recorded using stethoscope, which is also used for heart sound auscultation. The sounds heard by the stethoscope from the chest wall consist mainly of heart sounds, lung sounds, and other underlying sounds, such as thorax and muscle frictions. Heart sounds are mainly caused by the mechanical activity of the heart, including closing of the heart valves and the vibration of the heart.

A parallel source separation system is proposed to extract heart and lung sounds from single-channel mixed signals based on non-negative matrix decomposition and clustering strategies [1]. In addition, a hard threshold is used in the wavelet transform domain to separate the non-stationary part of the input signal (heart sound) from the stationary part (lung sound). By placing two microphones on the left and right chest walls to collect signals, the fast ICA algorithm is applied to separate the heart and lung sound signals [2]. Blind source separation is performed using non-negative matrix factorization (NMF). By constructing spectral mask and using NMF to decompose spectral features, heart and lung sound signals are separated [3]. A new period-encoding depth-using autoencoder (PC-DAE) method is proposed to isolate mixed cardiopulmonary sounds in an unsupervised manner by assuming different periodicity between the heart and respiratory rates [4].

Singular spectrum analysis (SSA) reveals the intrinsic structure of the data by breaking down the time series into interpretable components [5][6][7][8][9]. Therefore, in this paper, the collected cardiopulmonary sound signals are separated by SSA algorithm advance to extract purer lung sound signals, so as to improve the accuracy of subsequent classification of lung diseases.

VGG16 deep learning model has been used to detect and classify lung diseases [10]. The study has a large, publicly available dataset containing X-ray images of COVID-19, pneumonia, and pneumothorax. Classification of the lung diseases using deep learning and interpretable artificial intelligence (XAI) techniques has been discussed [11]. It mainly uses a variety of deep learning models, including convolutional neural networks (CNNs), hybrid model, integrated model and transformer model, and combines XAI technology to improve the interpretability of the model. A multi-class classification method based on CNN to learn from pulmonary disease images has been proposed in [12]. The pre-processing has been performed using center clipping, the fine-tuning learning performed using the EfficientNet B7 model, and the Multi GAP structure used to maximize the features of each layer.

Random forest is an ensemble learning method based on decision trees that is widely used in classification and regression tasks. It constructs multiple decision trees by randomly selecting samples and features during the training process, significantly improving the stability and generalization ability of the model, while being able to assess the importance of features and provide an interpretation of the classification results [13].

In this paper, we propose a method combining SSA and random forest classifier to separate cardiopulmonary sound signals from infants and adults, and classify the extracted lung sounds into various pulmonary diseases. The method is implemented through three main steps: signal separation, feature extraction, and classification including model training

and validation.

II. Description of Database

In this study we use the ICBHI 2017 Challenge dataset as the baseline dataset for pulmonary auscultation sounds for the detection of respiratory diseases. It includes 920 recordings ranging in length from 10s-90s from 127 subjects. In total, there are 5.5 hours of recordings, containing 6898 breathing cycles. The data includes both clean breathing sounds and noisy recordings in real life. The data are from either healthy subjects or those having one of the seven categories of respiratory diseases.

III. Methodology

A. Singular Spectrum Analysis

SSA is a powerful time series analysis and processing tool, which is often used for single-channel signal separation, noise reduction, and pattern recognition [9]. The basic principle is to decompose and reconstruct the one-dimensional time series signal by embedding it into the high-dimensional trajectory space and extracting meaningful subspace structure from it.

1) Embedding: First, the one-dimensional time series is transformed into a high-dimensional trajectory matrix by stacking overlapping signal segments. Given a time series $X = \{x_1, x_2, \ldots, x_N\}$ and window length L. In general, L satisfies $2 \leq L < N/2$, where N is the length of the time series. In this research, L takes the value of 250. Then, we construct a trajectory matrix X by the Hankel transform of the signal, where each column is an overlapping segment of the time series, defined as:

$$X = \begin{bmatrix} x(1) & x(2) & \cdots & x(K) \\ x(2) & x(3) & \cdots & x(K+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(L) & x(L+1) & \cdots & x(N) \end{bmatrix} \tag{1}$$

where the trajectory matrix X is a matrix of L rows and K columns, where each column of X is part of the original signal, L is the window length or embedding dimension, and K is the number of columns of the trajectory matrix, equal to $K = N - L + 1$.

2) The Covariance Matrix and Eigenvalue Decomposition: Next, we compute the covariance matrix S of the trajectory matrix X. The covariance matrix S is defined as

$$S = XX^T \tag{2}$$

Then, the EVD is performed, and the eigenvalues are sorted, and the relevant principal components are extracted for the separation of subsequent signals.

3) Adaptively Grouping the Desired Signal Components: From the morphology of the signals it is evident that the heart sound has a higher kurtosis than the lung sound. This property is exploited as a criterion for our adaptive single-channel source separation system. To select the desired eigentriples, we multiply a diagonal matrix W by EVD factors and optimize it in each iteration. The W matrix is initialized as an identity matrix, and it is iteratively optimized to maximize the kurtosis of the signal. Kurtosis measures the sharpness of the signal distribution. In each iteration, the W matrix is updated by evaluating the gradient of kurtosis with respect to W and using the gradient descent method. Let the kurtosis function be estimated as follows:

$$Kurt(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^4 \Bigg/ \left(\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2\right)^2 \tag{3}$$

where X is the separated signal, $X_i = W_{ii}\lambda_i^{1/2}U_i$, N denotes the total number of data points, $Kurt$ stands for kurtosis, and μ is the average value of the signal X. In the optimization process, W is then updated as:

$$W(i+1) = W(i) - \alpha \cdot \nabla_W(Kurt) \tag{4}$$

where α is the learning rate and $\nabla_W(Kurt)$ is the gradient concerning W. λ_i and U_i are respectively, the ith eigenvalue and eigenvector of S in eq.(2).

4) Reconstruction: According to the optimized W matrix, the covariance matrix S is decomposed into the lung sound interferences subspaces including the heart sound subspace from which each signal can be reconstructed.

Then, the time domain signal is reconstructed by diagonal averaging [9], transforming the matrix back into a one-dimensional signal while retaining the main characteristics of the desired lung sound signal. Through these steps, SSA can effectively separate lung sound and heart sound signals, providing accurate data for subsequent signal analysis and diagnosis.

5) Kurtosis: Kurtosis is a statistic that measures the peakedness of the distribution of a signal's probability distribution. By maximizing the kurtosis, the accuracy of signal separation can be improved, especially when separating non-Gaussian signal components [14]. In this study, we propose to regularize the SSA-based signal reconstruction by maximizing the kurtosis of the reconstructed signal component to optimize the separation effect of heart and lung sounds. The overall is shown in Fig. 1.

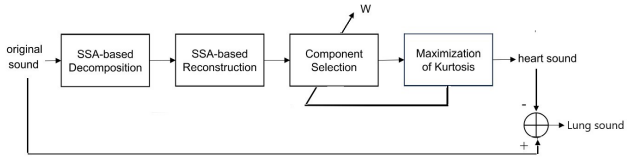


Fig. 1: The adaptive SSA algorithm regularized by kurtosis maximization criterion.

B. Random Forest Classifier

A random forest is an ensemble learning method that builds multiple decision trees for classification or regression. Each tree is trained on a different sample set and feature subset,

and the final prediction is voted on by the results of all trees [12].The process of building and evaluating a random forest classifier is shown in Fig. 2.

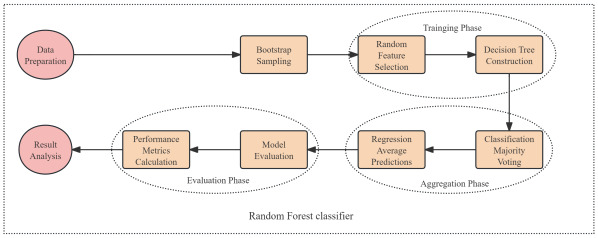


Fig. 2: The process of building and evaluating a random forest classifier.

For the original dataset, random forest generates multiple training sets through Bootstrap sampling [15]. Each training set is obtained by randomly drawing samples from the original data set, and the size of each training set is the same as the original dataset. Assume T is a training set generated through Bootstrap sampling: $\{x_i\}_{i=1}^{n}$ of which each x_i is randomly drawn from the original data set, there are put back, so each sample may be repeated in the training set.

For each of the Bootstrap sampling to generate the training set, build a decision tree. Split at each node, randomly select m characteristics (namely feature subset), and then select the best split point from these features [11]. If the total number of features is p, it is usually $m \ll p$ (classification problem) or $m \approx p$ (regression problem). For each node, select the split point that maximizes information gain or minimizes Gini impurity in the feature subset. If the feature subset is $\{x_1, x_2, \ldots, x_m\}$, the optimal split point is selected as s, where s represents the point of split, the information gain brought about by s, or the reduction of Gini impurity [15].

After all the trees are built, the random forest makes a final prediction by integrating the results of all the tree predictions. For the classification tasks, the random forest determines the final category through a majority voting mechanism. Let N be the number of trees in the forest and y_j the prediction of the j-th tree for sample x, then the final prediction category $\hat{y}$ is:

$$\hat{y} = \text{mode}(y_1, y_2, \ldots, y_N)$$

where mode represents the mode function, that is, the class that returns the most occurrences. For the regression task, the random forest determines the final predicted value by averaging the predicted values of all trees. Let $\hat{y}_j$ be the predicted value of the j-th tree for a new sample x, then the final predicted value $\hat{y}$ is:

$$\hat{y} = \frac{1}{N} \sum_{j=1}^{N} \hat{y}_j$$

IV. Results and Discussion

According to the characteristics of cardiopulmonary sound data selected in this study, the embedding dimension (window length) of SSA was set to 250 for constructing the trajectory matrix. For the W matrix optimization part, the learning rate is set to 0.01 for the gradient descent optimization of diagonal W matrix. At the same time, the convergence threshold is set to 1e-6 and the maximum number of iterations is set to 100, which is used to determine the convergence condition of the gradient descent algorithm and control the number of iterations of the gradient descent algorithm. The following takes one of the speech signals as an example to illustrate the effectiveness of using SSA and Kurtosis algorithm to separate the cardiopulmonary sound signals.

A. Results of separation

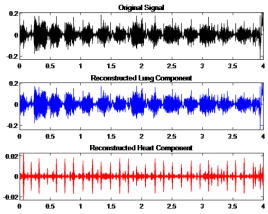
From the signal waveform and the histogram of Fig. 3, we can see that the amplitude of the original signal varies greatly at different time points, and the amplitude value presents a symmetric distribution between -0.2 and 0.2, and a wide frequency band. The time domain waveform and histogram of separated lung sound signal are similar to the original signal, including a wide frequency range and complex time domain changes. It is consistent with the characteristics that lung sounds usually contain high and low frequency components and vary with the respiratory cycle. The time domain waveform of separated heart sound signal is regular, the amplitude is relatively small and the periodicity is consistent. The amplitude values of the histogram are concentrated between -0.02 and 0.02, and the distribution is narrow. This is consistent with the physiological characteristics of heart sounds, which are mainly concentrated in the lower frequency range and produce a distinct pulse with each heartbeat.

In Fig. 3, the spectrum diagram shows the frequency distribution of signals at different time points, and power spectral density diagram shows the power distribution of signals at different frequencies. The reconstructed pulmonary sound signal has a significant energy distribution in the frequency range of 0 to 10 kHz, especially with a strong component above 1 kHz, and has a high power over a wide frequency range.

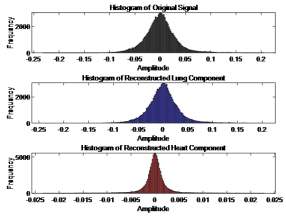
The reconstructed heart sound signals are mainly concentrated in the low frequency range (0 to 1 kHz), have low power, and show periodic changes on the timeline. This is consistent with the low-frequency nature of the heart sound signal and the pulses produced by each heartbeat. The spectral diagram and the power spectrum density analysis further quantified the distribution of the separated signal in the frequency domain, and proved the efficiency of the separation process.

In addition, in terms of numerical value, the dot product of reconstructed lung sound signal and heart sound signal is 1.480311, and Pearson correlation coefficient is 0.172788, indicating that the linear correlation between the two signals is weak, and there is a certain independence in the time domain. This further illustrates that the separation effect is better and the correlation between the signals is low.

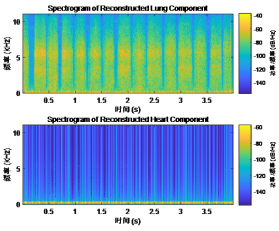
In addition, in terms of numerical value, the dot product of reconstructed lung sound signal and heart sound signal is 1.480311, and Pearson correlation coefficient is 0.172788, indicating that the linear correlation between the two signals



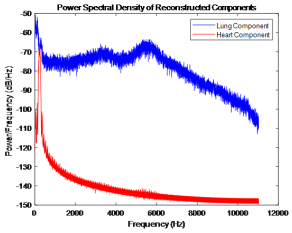
(a) Time domain waveform of original signal (left top), reconstructed lung sound signal (left middle) and heart sound signal (left bottom).



(b) Histogram of the original signal (right top), reconstructed lung sound signal (right middle), and heart sound signal (right bottom).



(c) Spectrograms of reconstructed lung sound signals (left top) and heart sound signals (left bottom).



(d) Power spectral density (PSD) of reconstructed lung sound signals and heart sound signals (right).

Fig. 3: Analysis of Time Domain, Histogram, Spectrogram, and Power Spectral Density of Cardiopulmonary Sound Signals.

are weak, and there is a certain independence in the time domain.

The overall classification accuracy with balanced dataset of infant samples before and after separation was 0.3333 and 0.5. The classification performance of the cardiopulmonary sound after separation has different performance in different categories.

B. Results of lung disease classification

Here is a summary of the respiratory diseases in the database, including the breakdown for children aged 7 years or younger in Table I. Because the number of samples of LRTI and Asthma patients is too small to be accurately classified, these two types of samples were not considered in the actual experiment.

TABLE I: Summary of the respiratory diseases in the database

Disease Category	Number of All Patients	Number of All Samples	Number of Infants ($\leq$ 7 years)	Number of Samples ($\leq$ 7 years)
Healthy	26	35	16	21
COPD	64	793	0	0
URTI	14	23	12	21
Bronchiectasis	7	16	0	0
Pneumonia	6	37	1	3
Bronchiolitis	6	13	6	12
Total	123	917	35	57

Table II and Table III illustrate respectively the performance of the classifier performance before and after applying SSA with balanced dataset. Precision represents the percentage of all samples correctly classified. Recall represents the proportion of all samples. F1 Score is the harmonic average of accuracy rate and recall rate,and is a comprehensive evaluation of the two.

TABLE II: The classification performance before and after lung sound separation (infants).

Disease Category	Precision (After)	Precision (Before)	Recall (After)	Recall (Before)	F1 Score (After)	F1 Score (Before)
Bronchiolitis	0.66667	NAN	0.66667	0	0.66667	NAN
Healthy	0	0	0	0	NAN	NAN
Pneumonia	0.33333	0.5	0.66667	1	0.44444	0.66667
URTI	1	0.25	0.66667	0.3333	0.8	0.28571

TABLE III: The classification performance before and after lung sound separation (infants and adults).

Disease Category	Precision (After)	Precision (Before)	Recall (After)	Recall (Before)	F1 Score (After)	F1 Score (Before)
Bronchiectasis	0.75	0.73333	0.92308	0.84615	0.82759	0.78571
Bronchiolitis	0.41667	0.4375	0.38462	0.53846	0.4	0.48276
COPD	0.92308	0.92308	0.92308	0.92308	0.92308	0.92308
Healthy	0.4	0.36364	0.30769	0.30769	0.34783	0.33333
Pneumonia	0.75	0.66667	0.92308	0.76923	0.82759	0.71429
URTI	0.36364	0.125	0.30769	0.076923	0.33333	0.095238

In general, the classification model of cardiopulmonary sound after separation has different performance in different categories. For infant data, the classification performance of the bronchitis and upper respiratory infection categories improved significantly. However, for the healthy and pneumonia categories, the model performance did not significantly improved or deteriorate. When the adult data were added, the performances of the model in the identification of bronchitis, healthy and pneumonia were different, and the overall performance of the identification of pneumonia and upper respiratory tract infection was significantly improved.

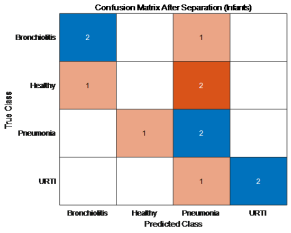
Classification of the separated single infant samples and the addition of the adult samples were generally better than classification of the unseparated samples. After separation, the model has higher accuracy and lower loss in training and validation data. The unseparated models showed obvious signs of overfitting, the verification accuracy and loss fluctuated greatly, and the classification performance was poor.

The overall classification accuracy with balanced dataset of infant samples before and after separation was 0.3333 and 0.5, and that of all samples before and after separation was 0.57692 and 0.62821, and the accuracy with unbalanced dataset of all samples before and after separation was 0.91385 and 0.91603, respectively. In conclusion, the separation of cardiopulmonary sounds is helpful to improve the accuracy and stability of respiratory disease classification.

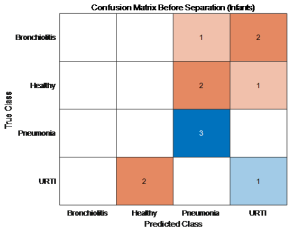
V. Conclusions

In this study, SSA algorithm has been regularized by maximization of kurtosis to enhance auscultated sound signal separation. The classification performance before and after separation was evaluated for a random forest classifier and used cross-validation and provided a comprehensive evaluation process and detailed analysis of the results.

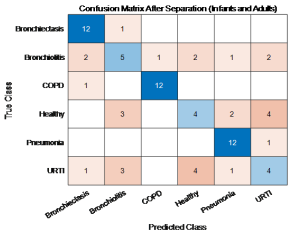
The separation of cardiopulmonary sound signals from infants and adults significantly improved the performance of the respiratory disease classification model. The results suggest that by incorporating kurtosis for SSA optimization, the model is better equipped to handle the complexities of cardiopulmonary sounds, leading to more accurate classifications. This



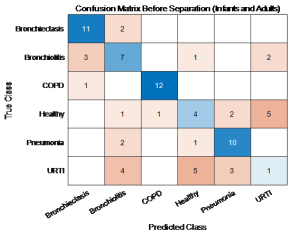
(a) Confusion Matrix After Separation (Infants).



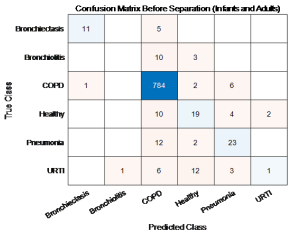
(b) Confusion Matrix Before Separation (Infants).



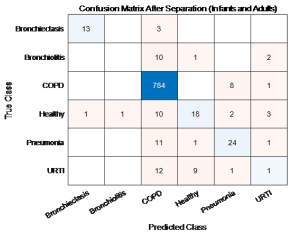
(c) Confusion Matrix After Separation (Infants and Adults).



(d) Confusion Matrix Before Separation (Infants and Adults).

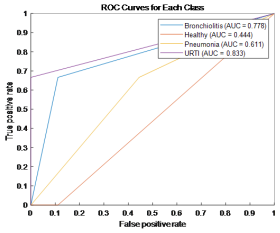


(e) Confusion Matrix After Separation (Unbalanced Dataset).

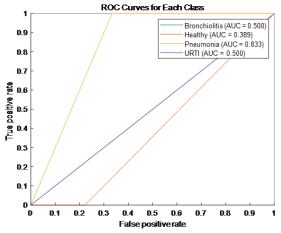


(f) Confusion Matrix Before Separation (Unbalanced Dataset).

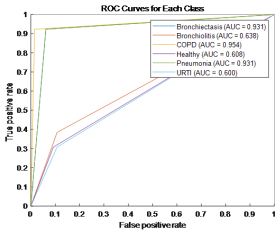
Fig. 4: Confusion Matrices for Different Conditions.



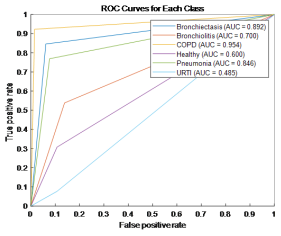
(a) ROC Curve After Separation (Infants).



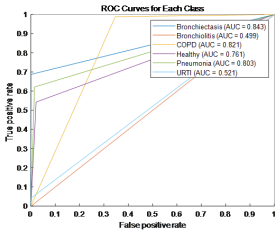
(b) ROC Curve Before Separation (Infants).



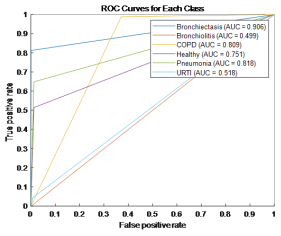
(c) ROC Curve After Separation (Infants and Adults).



(d) ROC Curve Before Separation (Infants and Adults).



(e) ROC Curve After Separation (Unbalanced Dataset).



(f) ROC Curve Before Separation (Unbalanced Dataset).

Fig. 5: ROC Curves After (left of top) and Before (middle of top) Separation of Infants and After (right of top) and Before (left of bottom) Separation of Infants and Adults with Balanced Dataset and After (middle of bottom) and Before (right of bottom) Separation of Infants and Adults with Unbalanced Dataset.

study provides empirical support for the practical application of cardiopulmonary sound separation technology in clinical settings, demonstrating its potential to enhance diagnostic accuracy and contribute to better patient outcomes.

REFERENCES

[1] A. J. Muñoz-Montoro, D. Suarez-Dou, and R. Cortina, "Parallel source separation system for heart and lung sounds," *J Supercomput*, vol. 77, pp. 8135–8150, 2021.

[2] J. C. Chien, M. C. Huang, Y. D. Lin, and F. C. Chong, "A study of heart sound and lung sound separation by independent component analysis technique," in *Conf Proc IEEE Eng Med Biol Soc*, 2006, pp. 5708–5711.

[3] C. Lin and E. Hasting, "Blind source separation of heart and lung sounds based on nonnegative matrix factorization," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, Naha, Japan, 2013, pp. 731–736.

[4] K.-H. Tsai and others, "Blind Monaural Source Separation on Heart and Lung Sounds Based on Periodic-Coded Deep Autoencoder," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 3203–3214, 2020.

[5] S. Sanei and H. Hassani, *Singular Spectrum Analysis of Biomedical Signals*, CRC, 2015.

[6] S. Sanei, T. K. M. Lee, and V. Abolghasemi, "A new adaptive line enhancer based on singular spectrum analysis," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 428-434, 2012.

[7] F. Ghaderi, H. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3360-3367, 2011.

[8] S. Sanei, M. Ghodsi, and H. Hassani, "A constrained singular spectrum analysis approach to murmur detection from heart sounds," *Elsevier Journal of Medical Engineering and Physics*, vol. 33, no. 3, pp. 362-367, 2011.

[9] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, Springer, New York, NY, 2013, pp. 11.

[10] A. Chatchaiwatkul and others, "Lung Disease Detection and Classification with Deep Learning Approach," in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2021, pp. 1–4.

[11] T. Taher, S. Ahmed, S. Shahriar, and T. Towhid, "Explainable Lung Disease Classification from Chest X-Ray Images Utilizing Deep Learning and XAI," 2024, pp. 1–5.

[12] M. Hong, B. Rim, H. Lee, H. Jang, J. Oh, and S. Choi, "Multi-Class Classification of Lung Diseases Using CNN Models," *Applied Sciences*, vol. 11, no. 19, p. 9289, 2021.

[13] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[14] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
[15] A. Liaw and M. Wiener, "Classification and Regression by RandomForest," *Forest*, vol. 23, 2001.