Turkish Journal of Science & Technology
17(1), 79-88, 2022

Research Paper
https://doi.org/10.55525/tjst.1063039

A Lung Sound Classification System Based on Data Augmenting Using ELM-Wavelet-AE

Berna ARI[1*], Ömer Faruk ALÇİN[2], Abdulkadir ŞENGÜR[1]
[1] Department of Electrical and Electronics Engineering, Firat University, Elazig, Turkey
[2] Department of Electrical and Electronics Engineering, Malatya Turgut Ozal University, Malatya, Turkey
[*1] bernagurler.a@gmail.com, [1] ksengur@firat.edu.tr, [2] omer.alcin@ozal.edu.tr

(Geliş/Received: 25/01/2022; Kabul/Accepted: 08/02/2022)

Abstract: The method is of great importance in systems that include machine learning and classification steps. As a result, academics are constantly working to improve the process. However, the data pertaining to the methodology's performance is equally as valuable as the methodology's creation. While the data is utilized to show the result of the modeling process, it is critical to consider the proper labeling of the data, the technique of acquisition, and the volume. Obtaining data in certain sectors, particularly medical fields, can be costly and time consuming. Thus, data augmenting via classical and synthetic methods has recently gained popularity. Our study uses synthetic data augmentation since it is newer, more efficient, and produces the desired effect. Our study's goal is to classify a data collection of lung sounds into four groups using data augmenting. Obtaining and standardizing the wavelet scatter transformation of each cycle of lung sounds, splitting the transformed data into test and training, augmenting and classifying the training data. In the augmenting stage, we utilized ELM-AE, then ELM-W-AE, with six wavelet functions (Gaussian, Morlet, Mexican, Shannon, Meyer, Ggw) added. The SVM and EBT classifiers improved performance by 4% and 3% in ELM-W-AE compared to the original structure.

Key words: Lung sound, wavelet scatter, data augmentation, ELM-Auto Encoder.

ELM- Dalgacık-AE Kullanılarak Veri Çoğullama Tabanlı Bir Akciğer Sesleri Sınıflama Sistemi

Öz: Makine öğrenmesi ve sınıflandırma adımlarını içeren sistemlerde yöntem büyük önem arz etmektedir. Bu sebeple araştırmacılar genellikle yöntemin iyileştirilmesi üzerinde çalışmalar yapmaktadır. Ancak metodolojinin geliştirilmesi kadar performansını etkileyen veri de bir o kadar değerlidir. Veri, modelleme sürecinde sonucu gözler önüne serebilmek için kullanılırken; verinin doğru etiketlenmesi, elde edilme yöntemi ve hacmi dikkat edilmesi gereken diğer önemli noktalardır. Medikal alanlar başta olmak üzere bazı alanlarda veri elde etmek maliyetli ve zor olabilmektedir. Bu sebeple klasik ve sentetik yöntemlerle veri çoğullama yaklaşımları son zamanlarda popüler olmaya başlamıştır. Sentetik veri çoğullama teknikleri daha yeni, verimli ve istenebilen sonuca yönelik olduğundan çalışmamızda tercih edilmiştir. Çalışmamızın amacı akciğer seslerine ait veri setini dört kategoride sınıflandırırken seçtiğimiz veri çoğullama yönteminin başarımını göstermektir. Önerdiğimiz yöntemin adımları şu şekildedir: akciğer seslerine ait her bir saykılın Dalgacık saçılım dönüşümünü elde edilmesi ve normalizasyonu, dönüşümden elde edilen verinin test ve eğitim olarak bölünmesi, eğitim için ayrılan verinin çoğullanması ve sınıflandırılmasıdır. Veri çoğullama aşamasında Aşırı Öğrenme Makinesi Oto Kodlayıcı (ELM-AE) ve sonrasında bu modele altı farklı dalgacık fonksiyonun (Gaussian, Morlet, Mexican, Shannon, Meyer, Ggw) eklenmesiyle ELM-W-AE yapısını kullandık. Orijinal yapıya kıyasla sınıflandırmada kullanılan SVM ve EBT sınıflandırıcıları için ELM-W-AE'de sırasıyla yaklaşık %4 ve %3 oranında başarım artışı gözlemledik.

Anahtar kelimeler: Akciğer sesleri, dalgacık saçılımı, veri çoğullama, ELM-Oto Kodlayıcı

1. Introduction

Data is a critical structure that plays a critical part in the modelling phase and enables us to monitor the outcome through system testing. While the volume and consistency of data collected throughout the process of artificial intelligence impact the end, it also results in noticeable variances in the result step. However, approaches such as Convolutional Neural Network (CNN) and (Long Short-term Memory) LSTM, which have lately gained popularity, have been shown to improve performance as the amount of huge data in the system increases. When the studies are examined, it is clear that the approaches prioritize modeling over data-driven approaches, and that when the anticipated performance response is not obtained, the modeling is revised [1]. Data collection procedures in a variety of fields, including engineering, medicine, and education, can be time consuming and costly. Recent studies on data augmenting technologies come to mind at this time. Along with traditional data augmentation techniques, new study topics have been identified for synthetic data augmentation techniques such as (Generative Adversarial Network (GAN), AE (Auto Encoder) and Variational Auto Encoder (VAE) [2-5]. It has been

[*] Corresponding author: bernagurler.a@gmail.com. ORCID Number of authors: [*1] 0000-0003-1000-2619, [2] 0000-0002-2917-3736, [1] 0000-0003-1614-2639

demonstrated that better performance may be obtained by augmenting medical images and sounds when creating decision support systems for time and cost savings [6-9].

Synthetic data augmenting is a more sophisticated and recent technique than traditional data augmenting. There is research on the synthesis of AE using the Extreme Learning Machine (ELM) approach, one of the favored methods for producing synthetic data [3, 7]. AE has become popular due to its efficiency compared to other data duplication methods [2, 10]. On the other hand, the ELM paired with AE called ELM-AE is a single hidden layer feedforward neural network and has been shown to be a high-performing model in numerous investigations. The ELM-AE structure can be modified by altering the number of cells and activation functions in the hidden layer. The issue of data quantity, which is particularly acute in deep architectures, was attempted to be investigated in the categorization of lung disorders, the focus of this work.

Since lung disorders are the third leading cause of death worldwide, identification and follow-up are critical [11]. Following the Covid-19 outbreak, a rise in lung illnesses has been observed [12]. As a result, professionals should develop new approaches for characterizing these disorders. It is well established that lung sounds and their features significantly influence the development of pulmonary disease [13]. Auscultation can be used to discern between these sounds, which are roughly classified as normal and pathological [14]. However, classical auscultation is not without faults, since it is highly dependent on the physician's skill, hearing capacity, expertise, and experience [14, 15, 16].

The non-stationarity of lung sound waves is the primary element that challenges classic technique analysis [17]. Lung sounds are deemed normal when they range between 100 and 1000Hz and lack prominent peaks on the signal [18]. However, in the case of unusual (adventitious) breath sounds, the situation is reversed. These noises contain additional sounds in addition to the typical sounds, and they are classified as continuous-discrete [19]. Wheeze refers to continuous sounds produced by the lungs; crackle refers to isolated sounds produced by the lungs. Wheeze sounds, which contain a tonal structure, feature periodic waveforms with a frequency more than or equal to 100Hz and duration greater than or equal to 100ms [18, 20]. Crackle, on the other hand, has a more complex structure in terms of frequency content, although it runs in less than 20ms [21]. The diagnosis of asthma, pneumonia, and bronchitis is guided by wheeze sounds, whereas crackling sounds are usually encountered in cardiovascular illnesses [22]. Automatic recognition studies have risen to prominence in expert systems assessing the amounts mentioned above, assisting in disease diagnosis and guiding disease interpretation.

When studies were analyzed, it was discovered that lung sounds were classified and substantial results were discovered [23, 24, 25]. The time-frequency domain features of normal and abnormal sounds have been the focus of studies analyzing normal and abnormal noises [26]. In numerous research, Mel frequency cepstral coefficients (MFCC) and estimated entropy have been used [27]. Additionally, recent research using empirical mode decomposition (EMD) and intrinsic mode functions (IMFs) demonstrate these methods' superior efficacy in classifying lung sounds as normal-adventive [23, 25].

As indicated previously, another critical part of methodological preparation for lung sound analysis was the data set. In terms of the debatability of the investigations, the ICHBI 2017 dataset was assessed, which included studies with dual and multiple classifications [28]. When we consider the research that comprises various classes of normal, wheeze, crackle, and both wheeze and crackle, we see that in the study [29], the objective was to convert lung sound data into images using the short-term Fourier transform (STFT). Classification of signals extracted from spectrogram images was performed using a pre-trained CNN. The study [13] began by converting the audio signals in the dataset to spectrograms. Following that, a CNN model is suggested that tries to improve performance by parallelizing the average pooling and maximum pooling layers. Linear Discriminant Analysis (LDA) and Random Subspace Ensembles (RSE) were used to classify the deep features produced in this manner (71.15%). In [22], lung sound signals were transformed to spectrogram pictures and five-fold augmentation was achieved synthetically. It is stressed that the images fed into the multi-layered CNN model and the enhanced data have a beneficial effect on performance. The study [30] used MFCC coefficients as features and tested binary and multi-class classification using Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM). The study [20] feeds the radial-based SVM classifier with the features acquired using wavelet decomposition and STFT.

This study aimed to investigate multi-class discrimination using lung sound signals. The following are the study's contributions to the literature:

- In the feature extraction stage of lung sound analysis, the wavelet scatters transform method was used.
- The ICHBI data set is augmented with the ELM-AE.
- Comparison is accomplished through the use of numerous wavelet functions during augmenting.

2. Categories of Lung Sounds

Fig. 1 shows the methodology that was developed. In begin, lung sound files were extracted from the 2017 ICBHI Respiratory Sound Database [28]. Table 1 contains the cycle details for this dataset, which contains 6898 cycles in total. These breath sounds were collected from various locations on the chest using various instruments.

Table 1. Class and cycle knowledge regarding the dataset

Class	Cycles
Crackles	1864
Wheezes	886
Crackles+ wheezes	506
Normal	3642
Number of all	6898

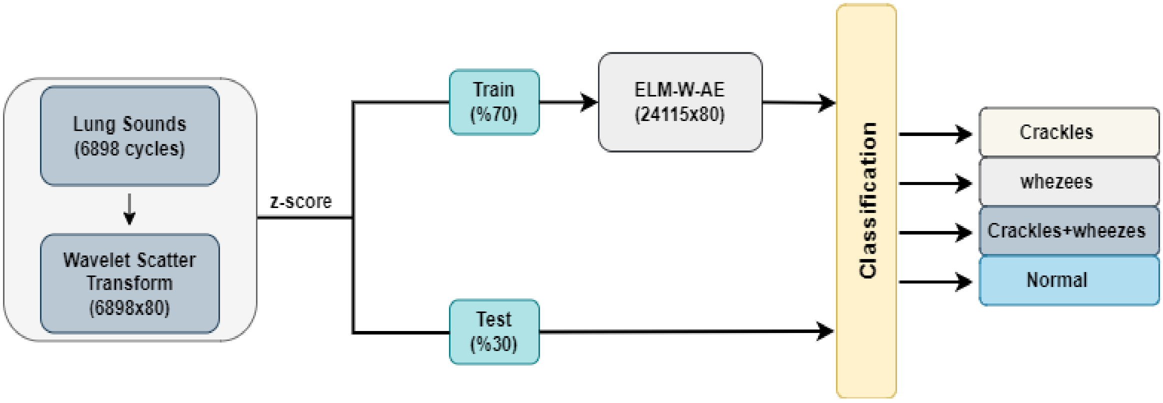

Fig.1. Flowchart of Methodology

2.1. Transform for Wavelet Scattering

Wavelet techniques, which are preferred for data representation and feature extraction, are advantageous because they can be used in conjunction with a wide variety of classification algorithms [31]. Additionally, it is one of the mathematical methods used when time-frequency domain feature extraction is insufficient for more complex signal feature extraction [32]. The Wavelet Scattering Transform (WST) is a structure proposed by Mallat that enables the generation of reliable features and their use in conjunction with a deep neural network structure [31, 33]. The convolution, nonlinearity, and averaging steps illustrated in Fig. 2 describe the primary steps involved in producing the wavelet scattering transform of the time series input signal. In this case, Ψ_1 denotes the wavelet function and φ_J denotes an average low-pass filter.

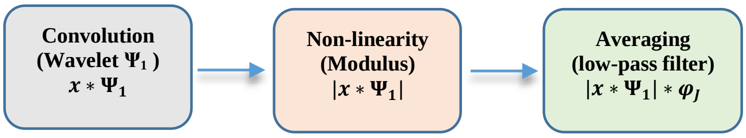

Fig.2. Representation of the wavelet scattering transform process with an x input.

WST defines a deformation-resistant representation. WST has been demonstrated to be capable of extending MFCC by processing modulation spectrum coefficients via wavelet convolutions and module operators [34]. Additionally, it has been demonstrated that WST outperforms MFCC for classification solutions with time scales greater than 25ms in audio representations. Using a set of wavelet decomposition and modulus operators, the scattering transform recovers information lost during Mel-frequency averaging. A wavelet transform is computed using constant-Q filter banks. A wavelet φ_J is a low-pass filter with $\tilde{\varphi}\,(0)$ equal to zero and is denoted by the center frequency ω in equation (1):

$$\varphi_{\omega}(\mathrm{j}) = \omega \cdot \varphi(\omega\, \mathrm{j}),\ \breve{\varphi}_{\omega}(\mathrm{s}) = \breve{\varphi}\left(\frac{s}{\omega}\right) \tag{1}$$

The frequency of $\breve{\varphi}$'s center has been normalized to 1 in this case. $\omega = 2^{k/Q}$. Q denotes the octave wavelets. $\breve{\varphi}$ is on a Q^{-1}-scale.

2.2. Synthetic data augmentation

In the study, WST was used to extract the features of the data in the ICHBI 2017 dataset. We applied data augmentation to these image representations containing information from four classes in the training process. Before we proceed with the steps, we utilized the z-score normalization (ZN), a straightforward feature-level transformation that can provide an effective solution for normalization. More precisely, when we speak of ZN, we subtract the mean of all components from each component and then divide by the standard deviation of all components [35]. The ELM-AE structure was investigated in the first model. The wavelet functions were then integrated into the designed structure to reveal the change. The abbreviation ELM-W-AE will be used to refer to wavelet functions. This section will detail each stage.

ELM: Huang's ELM is described as a simple one-hidden-layer neural network model [36]. Due to the random initialization of the input weights and single hidden layer thresholds, and the analytical calculation of the output weights, the ELM has a high learning rate. $\{(x_i, y_i) | x_i = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in R^n$ is the representation of the $y_i = [y_{j1}, y_{j2}, \dots, y_{jm}]^T \in R^m\}_{j=1}^{N}$ input-output structure for training pairs when the size of the training dataset is N, the number of input attributes is n, and the number of class labels is m.

$$\sum_{i=1}^{L} \beta_i f(w_i . x_j + b_i) = o_j,\ j = 1, \dots, N \tag{2}$$

In equation (2), L denotes the number of hidden layer neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ denotes the weight vector connecting the *ith* hidden node to the output node, f(·) indicates the activation function, $w_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{in}]^T$ denotes the weight vector of the input layer, b_i denotes the hidden layer thresholds, and $w_i . x_j$ denotes the values of w_i and x_j. It is used to denote the inner product of the output vector o_j. Equation (2) may not always produce the desired output o_j, and this new output may produce the desired output y_j , as illustrated in (3).

$$\sum_{i=1}^{L} \beta_i f(w_i . x_j + b_i) = y_j,\ j = 1, \dots, N \tag{3}$$

To optimize the performance of the Single Layer Feedforward Network (SLFN), the error should be $\sum_{j=1}^{N} \| o_j - y_j \| = 0$ or less. Equation (3) can be expressed straightforwardly in the matrix form specified in equation (4) [1].

$$Y = H\beta \tag{4}$$

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m},\ \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \tag{5}$$

where Y denotes the output vector, W denotes the weights of the output layer, and H represents the output layer matrix in the equation (6). Calculate the output weights by solving the equation in (7).

$$H = \begin{bmatrix} f(w_1 . x_1 + b_1) & \cdots & f(w_L . x_1 + b_L) \\ \vdots & \cdots & \vdots \\ f(w_1 . x_N + b_1) & \cdots & f(w_L . x_N + b_L) \end{bmatrix}_{NxL} \tag{6}$$

Here, H is the H matrix's generalized Moore-Penrose inverse.

$$\beta = H^{\dagger} Y \tag{7}$$

ELM-AE: The ELM-based Auto-Encoder (ELM-AE) is used to build an ELM-based multi-layer perceptron that is capable of learning new data representations. In contrast to ELM, which is used for classification, ELM-AE aims to minimize the reconstruction error associated with the input X. In other words, ELM-input AE's and output are both X. As a result, the objective function of ELM-AE with *L2* norm is as follows [37]:

$$\text{min:}\ \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\|X - H\beta\|^2 \tag{8}$$

where *C* is the factor of regularization. Then the gradient of Eq. 9 in terms of β is,

$$\beta = \left(\frac{1}{C} + H^T H\right)^{-1} H^T X \tag{9}$$

We obtain the optimal output weight β by setting the gradient to zero. The representation of the new data obtained is shown as in the Eq. (10).

$$X_{new} = G(X\beta^T) \tag{10}$$

where *G* is the function of activation. Notably, if the total number of hidden nodes in an ELM-based MLP is equal, *G* should be a linear activation function [37].

ELM-W-AE: Wavelet theory is a field of study that includes critical and constructive, including mathematics, physics, and engineering. The term "wavelet theory" translates as "little wave." In continuous form, the wavelet transform behaves similarly to a spanning elastic time-frequency window. It is classified into two types: continuous and discrete wavelet transforms. ELM is well-established as a superior method for learning Single Layer Feed-Forward Networks when compared to traditional methods. However, with careful consideration of parameter initialization and function selection, it is possible to achieve superior performance [38]. As stated in Eq.10, *G*, Table 2 contains six kernel types suitable for use as wavelet activation functions.

Table 2. Wavelet activation functions and mathematical representations used in the study

Wavelet Kernel Type	Function
Morlet	$\psi(t) = \cos(1.75t)e^{(-\frac{t^2}{2})}$
Gaussian	$\psi(t) = \frac{1}{\sqrt{2\pi}} exp^{(-\frac{t^2}{2})}$
Mexican	$c = \frac{2}{\sqrt{3}}\pi^{(-\frac{1}{4})}$ $\psi(t) = c(1 - t^2)exp\left(\frac{t^2}{2}\right)$
Shannon	$\psi(t) = \frac{sin\pi(t - 1/2) - sin2\pi(t - 1/2)}{\pi(t - 1/2)}$
Meyer	$\psi(t) = 35t^4 - 84t^5 + 70t^6 - 20t^7$
GGW	$\psi(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t)$

2.3. Classifiers

The original and augmented data sets were evaluated on classifiers, and this part explains the two classifiers that produced the best results.

SVM: Support Vector Machine (SVM) is a powerful technique for classifying data that works by creating a line in the plane between the members of two groups. It is favorable in that it applies to both linear and nonlinear data, has a high degree of precision, is capable of modeling complex decision boundaries, and works with many independent variables.

Decision-making function for SVM; x_i i is the data point, x_* a test vector, a_i is the Lagrangian multiplier associated with the training example x_i, y_i is the class of data point i (-1 or +1), and with b being the bias value, they are defined as[39,40]:

$$f(x_*) = sign\left[\sum_{i=1}^{N} a_i \, y_i \varphi\left(x_{*,} x_i\right) + b\right] \tag{11}$$

In the expression of the Quadratic (2nd order) optimization problem, ρ is the width of the separator between the support vector classes, w is the normal of the multi-plane (weight vector), $\parallel w \parallel$ w is the Euclidean representation of w for each $\{(x_i, y_i)\}$:

It is maximized with ρ=2/ (‖w‖). If $y_i = 1$ then; $w^T x_i + b \geq 1$. If $y_i = -1$ is equal to $w^T x_i + b \geq -1$.

EBT: The Ensemble Bagging Tree (EBT) classifier, which is a form of community learning, is intended to boost success rates through collaborative classification techniques. Rather than using a single learner, a decision tree is constructed using many copies of the primary learner's output, and the classifier output is coupled with the voting method [41]. Bagging is a term that refers to a group of decision trees that is utilized in regression. Community techniques employ numerous models to improve prediction performance by bagging together many lousy learner results into a high-quality community predictor. Bagging a community of decision trees is a variance reduction approach used to enhance decision trees' prediction performance. The bagged community power is calculated by estimating out-of-bag observations for each tree and averaging them across the entire community. Each observation's estimated out-of-bag response is compared to its actual value. The average out-of-bag error is computed by comparing expected to actual responses for all observations utilized in education. The average error extracted from this bag is a non-negative estimator of the genuine union error [42].

3. Experiments and Results

We present a series of experimental results relating to multi-class discrimination using lung sound signals in this study. The classification results were validated using the architectures described in Section 2.3. The effects of data augmentation were then analyzed using synthetic data augmentation methods. Synthetic data were generated using the methods described in Section 2.2. Because the Mexican kernel type performed better in the ELM-W-AE method during the experiments, the results were heavily weighted toward this section.

Table 3 compares the original, ELM-AE and ELM-W-AE results according to accuracy, specificity, sensitivity, precision, F1, MCC, Kappa [43]. Using the WST method, features were extracted from lung sounds (Crackle/ Whezees /Crackle+ Whezees / Normal) taken as 6898 cycles in four classes. 6898×80 features were extracted as a result of feature extraction. After normalization with ZN, the obtained features were divided into 70% training and 30% testing using hold-out cross-validation. A feature vector from the training set was augmented four times and formed into 5×80 dimensions using ELM-AE. A 24145×80 dimensional feature matrix was created using this approach for training features. The test data was not intended to be included in the augmented process, preventing memorization and ensuring the study's reliability. By testing synthetic lung data for three distinct phases, we examined the effects of data augmentation. We first evaluated the classifiers on the original images for the three major stages mentioned previously. In the second step, we created and analyzed synthetic images with the specified ELM-AE structure. The final stage involved integrating the six wavelet kernel listed in Table 2 into the ELM-W-AE structure and evaluating those using classifiers.

SVM and EBT methods were used to classify augmented training data and non-augmented test data, and the results are summarized in Table 3. The study's summary diagram is shown in Fig. 1. When Table 3 is examined, it is clear that the Mexican wavelet function provides the best performance. According to the classification of the original data, it was observed that EBT improved performance by approximately 3% and SVM improved performance by approximately 4.5 %. The complexity matrices for the original and Mexican wavelet functions are shown in Figures 3-6 in light of these data. When wavelet functions are viewed in general, it is clear that they perform better than the results obtained with the original data. Given the prevalence of lung diseases, it is clear that this percentage increase is significant. The authors concluded that wavelet kernel functions are worth experimenting with for this and similar classifications.

Table 3. Performance comparison of wavelet functions on original, ELM-AE and ELM-W-AE structure

Category		Method	Accuracy	Sensitivity	Specificity	Precision	F1	MCC	Kappa
Original		EBT	69.599	52.446	85.947	68.675	0.565	0.462	0.189
		DVM	56.404	31.284	78.440	57.059	0.291	0.157	0.140
ELM-AE		EBT	69.212	51.318	86.324	64.475	0.544	0.439	0.179
		DVM	59.981	37.673	81.214	52.318	0.382	0.247	0.063
ELM-W-AE	**Gaussian**	EBT	70.130	54.357	86.514	67.645	0.581	0.474	0.203
		DVM	59.884	37.400	80.966	51.589	0.381	0.245	0.065
	Morlet	EBT	71.242	54.828	86.884	68.990	0.587	0.488	0.233
		DVM	59.836	37.306	80.940	51.414	0.380	0.243	0.066
	Mexican	EBT	72.692	56.318	87.617	71.514	0.605	0.513	0.272
		DVM	60.609	38.969	81.200	56.848	0.407	0.277	0.048
	Shannon	EBT	69.357	50.862	85.915	65.721	0.544	0.441	0.183
		DVM	55.582	30.319	78.003	47.320	0.280	0.137	0.156
	Meyer	EBT	70.420	53.705	86.807	65.667	0.569	0.465	0.211
		DVM	59.401	36.948	80.721	50.565	0.376	0.236	0.076
	Ggw	EBT	70.904	55.978	87.052	68.783	0.597	0.493	0.224
		DVM	60.029	37.652	81.267	50.712	0.381	0.244	0.062

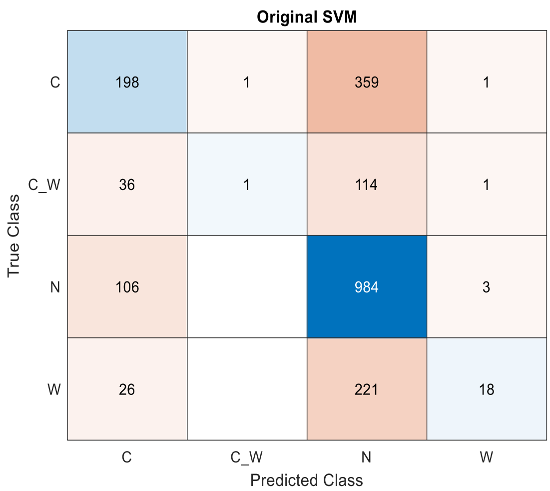

Fig.3. The complexity matrix of classification with SVM for the original data

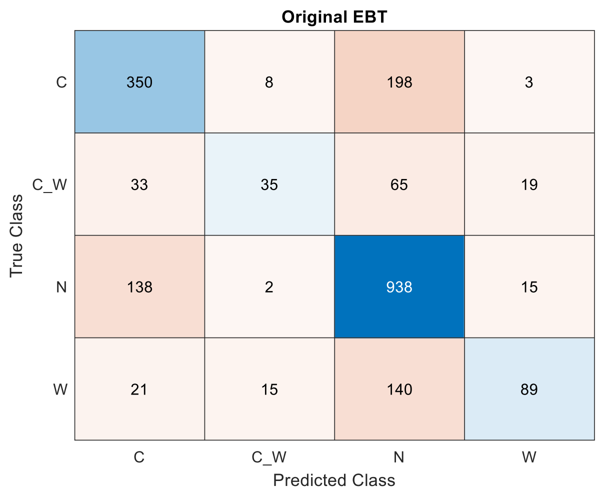

Fig.4. The complexity matrix of classification with EBT for the original data

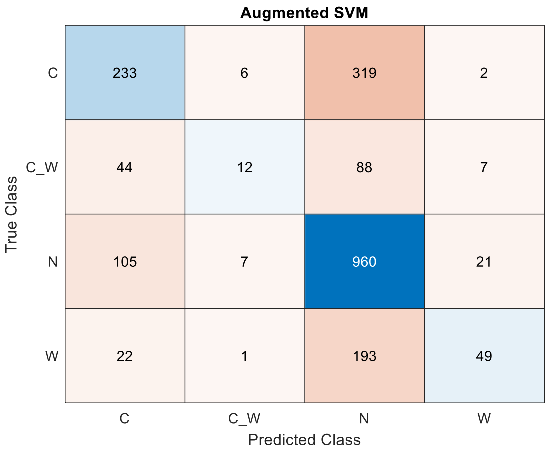

Fig.5. The complexity matrix of classification with SVM for data multiplexed with Mexican structure

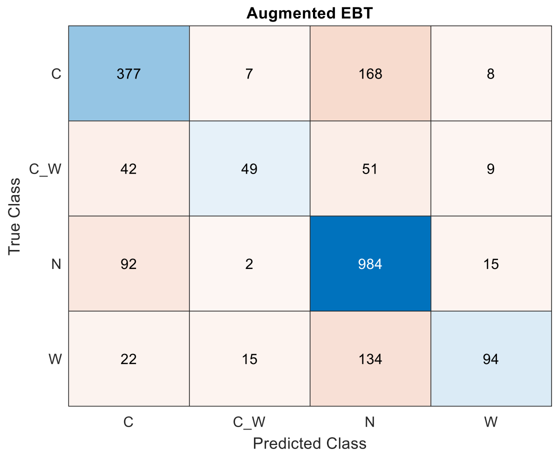

Fig.6. Complexity matrix of classification with SVM for multiplexed with Mexican structure

3.1. Comparison with other studies

As mentioned previously, we performed classification using synthetic amplification. The ELM-W-AE construct outperformed synthetic augmentation with ELM-AE alone. We compared our classification results for synthetic augmentation to the state-of-the-art studies using our best result, ELM-Mexican-AE. We chose it as a comparison because the studies we chose used the same dataset and classes. Table 4 compares the performance of the aforementioned studies in terms of preprocessing, feature extraction, classification, and classification accuracy.

Table 4. Comparison of lung sounds with four classes (Normal, crackles, wheezes, crackles+wheezes) with other studies

Reference	Pre-process	Feature Extraction and Method	Classification	Acc.
Ref[19]	resample to 4 KHz for standardization of all signals, 12th order Butterworth band pass filter	STFT, Q-factor wavelet	SVM	%54.15
Ref[22]	Band-pass filter in the frequency range 150-250 Hz, FFT	Spectrogram	CNN	%64.50
Ref[29]	-	Deep Feature with CNN model	SVM	%65.50
Ref[44]	Butterworth band pass filter, a non-linear resonance based wavelet decomposition	13 first mel-frequency cepstral coefficients	SVM	%49.86
Ref[45]	resample to 4 KHz for standarization of all signals	13 MFCCs coefficients STF and LTF parametres	SVM ANN RF	%72.1 %68.7 %68.7
The first proposed method	ZN	WST, ELM-AE	SVM EBT	%69.21 %59.98
The second proposed method	ZN	WST, ELM-W-AE	SVM EBT	%72.69 %60.61

4. Conclusion

The purpose of this study was to distinguish four classes based on voice recordings from the ICBHI 2017 respiratory dataset: normal, wheezing, wheezing, wheezing, and wheezing. Experiments were conducted in stages to evaluate the proposed method's classification performance. To begin, in contrast to previous research, audio signals are subjected to the wavelet scatter process for feature extraction. To offer a suitable classification approach, we first calculated and compared the classification accuracies of the original and augmented data. The train feature parts were augmented fivefold, including the original images with ELM-W-AE. SVM and EBT were used to classify the test features that were not included in the augmenting process and the augmented train features. Simultaneously, this classification process has been validated without the use of augmenting. The system was first compared to other architectures in terms of incremental and non-incremental classifiers. Both cases demonstrated unequivocally that the proposed incremental method performed better. The proposed architecture indicates that it is capable of providing a solution for disease detection while introducing a novel feature and experimental stage for the analysis of multi-class lung sounds.

References

[1] Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021, May). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

[2] Altan, G., & Kutlu, Y. (2018). Generative autoencoder kernels on deep learning for brain activity analysis. Natural and Engineering Sciences, 3(3), 311-322.

[3] Ferreira, J., Ferro, M., Fernandes, B., Valenca, M., Bastos-Filho, C., & Barros, P. (2017, November). Extreme learning machine autoencoder for data augmentation. In 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)(pp. 1-6). IEEE.

[4] Nishizaki, H. (2017, December). Data augmentation and feature extraction using variational autoencoder for acoustic modeling. In 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 1222-1227). IEEE.

[5] Cao, G., & Kamata, S. I. (2019, September). Data augmentation for historical documents via cascade variational auto-encoder. In 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 340-345). IEEE.

[6] Hussain, Z., Gimenez, F., Yi, D., & Rubin, D. (2017). Differential data augmentation techniques for medical imaging classification tasks. In AMIA annual symposium proceedings (Vol. 2017, p. 979). American Medical Informatics Association.

[7] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018, April). Synthetic data augmentation using GAN for improved liver lesion classification. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 289-293). IEEE.

[8] Zhao, A., Balakrishnan, G., Durand, F., Guttag, J. V., & Dalca, A. V. (2019). Data augmentation using learned transformations for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8543-8553).

[9] Eaton-Rosen, Z., Bragman, F., Ourselin, S., & Cardoso, M. J. (2018). Improving data augmentation for medical image segmentation.

[10] ŞENGÜR, D. EEG, EMG and ECG based Determination of Psychosocial Risk Levels in Teachers based on Wavelet Extreme Learning Machine Autoencoders. Politeknik Dergisi, 1-1.

[11] Shuvo, S. B., Ali, S. N., Swapnil, S. I., Hasan, T., & Bhuiyan, M. I. H. (2020). A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram. IEEE Journal of Biomedical and Health Informatics.

[12] ÇOLAK, M., BENLİ, Ş. G., & Müge, D. O. L. U. Akciğer Hastalıklarının Dalgacık Katsayıları Kullanılarak Karar Ağaçlarına Dayalı Sınıflandırılması. Avrupa Bilim ve Teknoloji Dergisi, (24), 463-468.

[13] Demir, F., Ismael, A. M., & Sengur, A. (2020). Classification of lung sounds with cnn model using parallel pooling structure. IEEE Access, 8, 105376-105383.

[14] N. Sengupta, M. Sahidullah and G. Saha, "Lung sound classification using cepstral-based statistical features", Comput. Biol. Med., vol. 75, pp. 118-129, Aug. 2016.

[15] G.-C. Chang and Y.-F. Lai, "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments", Comput. Methods Programs Biomed., vol. 97, pp. 141-150, Feb. 2010.

[16] S. Reichert, R. Gass, C. Brandt and E. Andrès, "Analysis of respiratory sounds: State of the art", Clin. medicine. Circulatory Respiratory Pulmonary Med., vol. 2, pp. 1-14, Jan. 2008.

[17] S. İçer and Ş. Gengeç, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," Digit. Signal Process., vol. 28, pp. 18–27, 2014, doi: 10.1016/j.dsp.2014.02.001.

[18] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," Comput. Biol. Med., vol. 34, no. 6, pp. 523–537, 2004, doi: https://doi.org/10.1016/S0010-4825(03)00092-1.

[19] S. ULUKAYA, G. SERBES, İ. ŞEN, and Y. P. KAHYA, "Akciğer Solunum Seslerinin Spektral Öznitelikler ile Sınıflandırılması," Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg., vol. 22, no. 2, p. 711, 2017, doi: 10.19113/sdufbed.84471.

[20] A. Sovijärvi et al., “Characteristic of breath sounds and adventitious respiratory sounds,” Charact. Breath Sounds Adventitious Respir. Sounds, vol. 10, pp. 591–596, Jan. 2000.
[21] Rocha, B. M., Pessoa, D., Marques, A., Carvalho, P., & Paiva, R. P. (2021, January). Influence of Event Duration on Automatic Wheeze Classification. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 7462-7469). IEEE.
[22] ER, M. B. (2020). Akciğer Seslerinin Derin Öğrenme ile Sınıflandırılması. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 8(4), 830-844.
[23] Khan, S. I., Palodiya, V., & Poluboyina, L. (2021). Automated classification of human lung sound signals using phase space representation of intrinsic mode function.
[24] Nguyen, T., & Pernkopf, F. (2021). Crackle Detection In Lung Sounds Using Transfer Learning And Multi-Input Convolitional Neural Networks. arXiv preprint arXiv:2104.14921.
[25] Khan, S. I., & Pachori, R. B. (2021). Automated classification of lung sound signals based on empirical mode decomposition. Expert Systems with Applications, 184, 115456.
[26] Murphy, Raymond & Vyshedskiy, Andrey & Power-Charnitsky, Verna-Ann & Bana, Dhirendra & Marinelli, Patricia & Wong-Tse, Anna & Paciej, Rozanne. (2005). Automated Lung Sound Analysis in Patients With Pneumonia. Respiratory care. 49. 1490-7. 10.1378/chest.124.4_MeetingAbstracts.190S-b.
[27] Aras, S., & Gangal, A. (2017, July). Comparison of different features derived from mel frequency cepstrum coefficients for classification of single channel lung sounds. In 2017 40th International Conference on Telecommunications and Signal Processing (TSP) (pp. 346-349). IEEE.
[28] "Challenge", 2017, [online] Available: https://bhichallenge.med.auth.gr/.
[29] Demir, F., Sengur, A., & Bajaj, V. (2020). Convolutional neural networks based efficient approach for classification of lung diseases. Health information science and systems, 8(1), 1-8.
[30] Serbes, G., Ulukaya, S., & Kahya, Y. P. (2017, November). An automated lung sound preprocessing and classification system based onspectral analysis methods. In International Conference on Biomedical and Health Informatics (pp. 45-49). Springer, Singapore.
[31] Soro, B., & Lee, C. (2019). A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization. Sensors, 19(8), 1790.
[32] Sepúlveda, A., Castillo, F., Palma, C., & Rodriguez-Fernandez, M. (2021). Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning. Applied Sciences, 11(11), 4945.
[33] Mallat, S. Group invariant scattering. Commun. Pure Appl. Math. **2012**, 65, 1331–1398.
[34] Lauraitis, A., Maskeliūnas, R., Damaševičius, R., & Krilavičius, T. (2020). Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features. IEEE Access, 8, 96162-96172.Zscore
[35] Fei, N., Gao, Y., Lu, Z., & Xiang, T. (2021). Z-Score Normalization, Hubness, and Few-Shot Learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 142-151).
[36] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.
[37] Uzair, M., & Mian, A. (2016). Blind domain adaptation with augmented extreme learning machine features. IEEE transactions on cybernetics, 47(3), 651-660.
[38] Javed, K., Gouriveau, R., & Zerhouni, N. (2014). SW-ELM: A summation wavelet extreme learning machine algorithm with a priori parameter initialization. Neurocomputing, 123, 299-307.
[39] Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), 113-126.
[40] Ucuz, I., Ciçek, A. U., Ari, A., Ozcan, O. O., & Sari, S. A. (2020). Determining the probability of juvenile delinquency by using support vector machines and designing a clinical decision support system. Medical hypotheses, 143, 110118.
[41] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
[42] Berna, A. R. I., Ali, A. R. I., & ŞENGÜR, A. (2020). Suicide Prediction from Hemogram with Machine Learning. Avrupa Bilim ve Teknoloji Dergisi, 364-369.
[43] Berna, A. R. I., İlknur, U. C. U. Z., Ali, A. R. I., Özdemir, F., & SENGUR, A. (2020). Grafik Tablet Kullanılarak Makine Öğrenmesi Yardımı ile El Yazısından Cinsiyet Tespiti. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 32(1), 243-252.
[44] Chambres, G., Hanna, P., & Desainte-Catherine, M. (2018, September). Automatic detection of patient with respiratory diseases using lung sound analysis. In 2018 International Conference on Content-Based Multimedia Indexing (CBMI) (pp. 1-6). IEEE.
[45] Gómez, A. F. R., & Orjuela-Cañón, A. D. (2021, May). Respiratory Sounds classification employing a Multi-label Approach. In 2021 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI) (pp. 1-5). IEEE