research: a study guide]. Kyiv: Olimpiiska literatura; 2021. 216 p. Ukrainian.

19. Crumley C. Abdominal Negative Pressure Wound Therapy Devices for Management of the Open Abdomen: A Technologic Analysis. Journal of Wound, Ostomy and Continence Nursing. 2022;49(2):124-7. doi: https://doi.org/10.1097/WON.0000000000000862

20. Fernandez LG, Sibaja Alvarez P, Kaplan MJ, et al. Application of Negative Pressure Wound Therapy with Instillation and Dwell Time of the Open Abdomen: Initial Experience. Cureus. 2019;11(9):e5667. doi: https://doi.org/10.7759/cureus.5667

21. Kocaaslan FND, Ozkan MC, Akdeniz Z, Sacak B, Erol B, Yuksel M, et al. Use of abdominal negative pressure wound therapy in different indications: a case series. Journal of Wound Care. 2019;28:4. doi: https://doi.org/10.12968/jowc.2019.28.4.240

22. Mansoor J, Ellahi I, Junaid Z, Habib A, Ilyas U. Clinical evaluation of improvised gauze-based negative pressure wound therapy in military wounds. International Wound Journal. 2013;12(50):559-63. doi: https://doi.org/10.1111/iwj.12164

23. Alhan D, Şahin I, Güzey S, Aykan A, Zor F, Öztürk S, et al. Staged repair of severe open abdomens due to high-energy gunshot injuries with early vacuum pack and delayed tissue expansion and dual-sided meshes. Turkish Journal of Trauma and Emergency Surgery. 2015;21(6):457-62. doi: https://doi.org/10.5505/tjtes.2015.05942

Стаття надійшла до редакції 26.03.2024;
затверджена до публікації 17.09.2024

UDC 616.24-073.7:004.8:004.032.26 https://doi.org/10.26641/2307-0404.2024.3.313569

D.V. Panaskin *,
S.H. Stirenko,
D.S. Babko

RESPIRATORY DISEASE DETECTION IN LUNG AUSCULTATION WITH CONVOLUTIONAL NEURAL NETWORKS AND CVAE AUGMENTATION

Ihor Sikorsky Kyiv polytechnic institute
Peremohy ave., 37, Kyiv, 03056, Ukraine
Київський політехнічний інститут ім. Ігоря Сікорського
пр. Перемоги, 37, Київ, 03056, Україна
**e-mail: denys_panaskin@edu.cn.ua*

Цитування: Медичні перспективи. 2024. Т. 29, № 3. С. 96-107
Cited: Medicni perspektivi. 2024;29(3):96-107

Key words: *pulmonary diseases, lung sounds, deep learning, convolutional neural network, conditional variational auto encoder*
Ключові слова: *легеневі захворювання, легеневі звуки, глибоке навчання, згорткова нейронна мережа, умовно-варіаційний автокодер*

Abstract. Respiratory disease detection in lung auscultation with convolutional neural networks and CVAE augmentation. Panaskin D.V., Stirenko S.H., Babko D.S. *The main purpose of this work was to investigate the possibility of detecting respiratory diseases in audio recordings of lung auscultation using modern deep learning tools, as well as to explore the possibility of using data augmentation by generating synthetic spectral representations of audio samples. The ICBHI (International Conference on Biomedical and Health Informatics) dataset was used for training, validation and augmentation. The dataset includes lung auscultations of 126 different subjects, there are a total of 920 sounds, of which 810 have signs of chronic diseases, 75 of non-chronic diseases and 35 with no pathology. The stage of data preprocessing includes discretization to 4kHz frequency, as well as filtering of frequency bands that do not carry*

information value for the task. In the next step, each sample was transformed into a frequency spectrum and Melspectrograms were generated. To solve the problem of class imbalance, the required number of synthetic spectrograms generated by convolutional variation autoencoders was added. At the stage of building the model, the methods of classical convolutional neural networks were used. The quality of the obtained algorithm was evaluated using a 10-fold cross-validation. Also, to assess the generalization of the proposed method, experiments were performed with the split of audio recordings into training and test sets using patient grouping. Qualitative evaluation of the model was performed using sensitivity, specificity, F1-score and Cohen's kappa. A score of 98.45% F1-score was achieved for the 5-class classification problem which can contribute to the development of ways to synthesize and augment sensitive medical data. In addition, a cons of existing methods in the generalization of the obtained predictions were revealed, which opens the way for further research in the direction of clinical respiratory diseases detection.

Реферат. Виявлення респіраторних захворювань при аускультації легень за допомогою згорткових нейронних мереж та доповнення CVAE. Панаскін Д.В., Стіренко С.Г., Бабко Д.С. *Основною метою цієї роботи було дослідити можливість виявлення респіраторних захворювань в аудіозаписах аускультації легень за допомогою сучасних інструментів глибокого вивчення, а також дослідити можливість використання аугментації даних шляхом генерації синтетичних спектральних представлень аудіозразків. Для вивчення, валідації та доповнення використовувався набір даних ICBHI (Міжнародна конференція з біомедичної та медичної інформатики). Набір даних включає аускультації легень 126 різних суб'єктів, загалом 920 звуків, з яких 810 мають ознаки хронічних захворювань, 75 – нехронічних захворювань і 35 – без патології. Етап попередньої обробки даних включає дискретизацію до частоти 4 кГц, а також фільтрацію частотних смуг, які не несуть інформаційної цінності для поставленої задачі. На наступному етапі кожна вибірка була перетворена в частотний спектр і згенеровані мелоспектрограми. Для вирішення проблеми дисбалансу класів було додано необхідну кількість синтетичних спектрограм, згенерованих згортковими варіаційними автокодерами. На етапі побудови моделі використовувалися методи класичних згорткових нейронних мереж. Якість отриманого алгоритму оцінювали за допомогою 10-кратної перехресної перевірки. Також для оцінювання узагальненості запропонованого методу були проведені експерименти з розбиттям аудіозаписів на навчальну та тестову множини з використанням групування пацієнтів. Якісну оцінку моделі проводили за допомогою чутливості, специфічності, F1-рахунку та каппи Коена. Для 5-класової задачі класифікації було досягнуто 98,45% F1-рахунку, що може сприяти розробці способів синтезу та доповнення чутливих медичних даних. Крім того, виявлено недоліки наявних методів в узагальненні отриманих прогнозів, що відкриває шлях для подальших досліджень у напрямку клінічної діагностики респіраторних захворювань.*

Respiratory diseases are now a significant problem and widespread worldwide. Pathologies such as asthma, chronic obstructive pulmonary disease (COPD), bronchiolitis, lung cancer, tuberculosis and COVID-19 account for a significant proportion of deaths each year. It is well known that early diagnosis of diseases is crucial to limit the spread of respiratory diseases, as well as their treatment and prevention [4].

Respiratory medicine is based on both clinical information and additional laboratory results. However, despite current technological capabilities, a proper medical history and thorough physical examination cannot be replaced as an initial step in making a correct diagnosis and, consequently, in prescribing appropriate treatment [12].

One of the proven and effective ways to detect pathology of the respiratory tract is auscultation of the lungs. This method of monitoring and diagnosis is considered safe, fast and cost-effective in the medical community. But there are several disadvantages of stethoscope diagnosis: the requirement for a qualified health care professional to interpret auscultation signals and the subjectivity of the interpretation, which causes variability between listeners. These constraints are exacerbated in poverty and pandemics due to a shortage of health professionals. Automated analysis of respiratory sounds can address these shortcomings, as well as help telemedicine programs monitor patients outside the clinic by health workers' community.

At present, tools for automatic analysis of lung auscultation sounds are being developed and implemented in order to simplify the processes of detecting respiratory cycles, recognizing additional lung sounds, as well as diagnosing respiratory diseases. Most diseases have specific symptoms that can be heard by auscultation of the lungs.

Thus, chronic obstructive pulmonary disease (COPD) is a chronic pathology that is difficult to detect. One of the main symptoms of COPD, as well as asthma, is the presence of high continuous sounds in the frequency range 100-2500 Hz and duration up to 80 ms [12, 15].

Bronchiectasis is a chronic condition in which the airways of the lungs dilate abnormally. These damaged airways allow bacteria and mucus to accumulate and accumulate in your lungs. This leads to frequent infections and airway obstruction. All these symptoms can be interpreted as bronchiolitis or just a cold. The main difference between the two diseases is that bronchiolitis most often affects young

children and can be cured, while bronchiectasis is a chronic disease [12].

Upper respiratory tract infection is a non-chronic disease that can occur at any time, but is more common in the fall and winter. The vast majority of upper respiratory tract infections are caused by viruses. The symptoms of this disease can be confused with the symptoms of pneumonia. Most people with pneumonia can recover in a short time, but for some people it can be extremely serious and even life-threatening, so a diagnosis is crucial [7, 12].

In addition, a large number of normal lung sounds are heard during auscultation of the lungs of healthy subjects. Acoustically normal tracheal sounds cover a wide range of frequencies, from less than 100 Hz to 5000 Hz with a sharp drop in power at frequencies above 800 Hz and low energy above 1500 Hz. The sounds heard over the chest wall, normal lung sounds, are generally only heard during inspiration and early part of expiration. Normal lung sounds present higher inspiratory frequency and intensity values than expiratory sounds [13]. Most studies have been focusing on analysing lung sound frequencies in healthy people and in people with respiratory conditions.

As you can see, distinguishing the symptoms of diseases from each other and from normal lung sounds is a non-trivial task. Diagnosis depends not only on the professional skills of the doctor and the patient's condition, but also on the appropriate technical equipment. Due to this, there is a need to create auxiliary algorithms based on lung auscultation in the diagnosis of respiratory diseases.

Issues of analysis and use of lung auscultation have been studied for a long time since the invention of the stethoscope. From auscultation data, doctors receive extremely important information about the condition of the upper internal organs and respiratory tract. Assisting practitioners in developing effective algorithms for analysing respiratory sounds, detecting respiratory cycles, recognizing extraneous noises and pathologies, and in diagnosing common and little-known diseases and their dependence on airway conditions is an important part of medical research and encourages the development of new medical decisions.

Abbas et al. [1] proposed the first automatic auscultation system, which overcomes the limitations of traditional auscultation, uses several levels of input signal filtering and adaptive neural network elimination of erroneous signals of the environment and internal extraneous sounds. This system was one of the first to use modern algorithmic methods of lung auscultation.

There is a huge amount of work focused on the study of the analysis and classification of extraneous noises in the respiratory tract, such as wheezes, crackles, rchonchi etc. Rocha et al. [14] describe the largest available open International Conference on Biomedical and Health Informatics (ICBHI) database of 920 patient audio recordings containing 6898 annotated breaths. This dataset is used in most open studies to classify respiratory cycles and determine patient diagnoses.

The papers contain a wide variety of approaches. Monaco et al. [10] presented a framework for respiratory sound classification that was based on two different kinds of features: short-term features which summarize sound properties on a time scale of tenths of a second and long-term features which assess sounds properties on a time scale of seconds using the publicly available dataset provided by ICBHI. The proposed model reached an accuracy of 85%±3% and a precision of 80%±8%, which compare well with the body of literature.

Minami et al. [9] also use the dataset to implement a respiratory sound classification model based on convolutional neural networks. Firstly, transformation of one-dimensional signals into two-dimensional time-frequency representation images using short-time Fourier transform and continuous wavelet transform was implemented. Secondly, classification of transferred images using convolutional neural networks was applied. It achieved score of 28%, harmonic score of 81%, sensitivity of 54% and specificity of 42%.

Kim et al. [5] utilize deep learning convolutional neural network (CNN) to categorize non-public 1918 respiratory sounds (normal, crackles, wheezes, rhonchi) recorded in the clinical setting. The predictive model was developed for respiratory sound classification combining pretrained image feature extractor of series, respiratory sound, and CNN classifier. It detected abnormal sounds with an accuracy of 86.5% and the area under the ROC curve (AUC) of 0.93. It further classified abnormal lung sounds into crackles, wheezes, or rhonchi with an overall accuracy of 85.7% and a mean AUC of 0.92.

Tasar et al. [16] aims to introduce a high accurate sound classification model using a nonlinear histogram-based generator. To reach this aim, piccolo pattern (it uses S-box of the piccolo cipher as a pattern), statistical moments, tunable q-factor wavelet transform (TQWT), iterative neighbourhood component analysis (INCA), and conventional classifiers are used together. The model uses TQWT to create levels. Piccolo pat-tern is employed to generate textural features (it is the main feature generator of this model), and statistics are deployed to extract statistical features. INCA chooses the relevant features. Decision tree (DT), support vector machine (SVM), and k nearest neighbours (KNN) classifiers

are applied to the selected feature vectors to calculate results. The model reached 99.45%, 99.31%, and 99.19% accuracies on three cases.

Another part of the studies is aimed at helping to diagnose existing diseases or their absence. Wu et al. [17] aims to use the output signals of a stethoscope and classify them through deep learning models automatically. In this research, the dataset consists of four classes, normal, wheezing, crackles, and unknown are used. To effectively classify each signal, we use the spectrogram generated by the short-time fast Fourier transform as the feature value of each lung sound signal and found the best parameters to do model selection. Besides, we also adopt Depthwise separable (DS) convolution technic, and refer to the architecture of Mobile-Net, to achieve the purpose of high accuracy and low model parameters.

Pham et al. [11] propose a new framework to classify anomalies in respiratory cycles and detect diseases, from respiratory sound recordings. The framework begins with front-end feature extraction that transforms input sound into a spectrogram representation. Then, a back-end deep learning network is used to classify the spectrogram features into categories of respiratory anomaly cycles or diseases. Experiments, conducted over the ICBHI benchmark dataset of respiratory sounds, confirm three main contributions towards respiratory-sound analysis.

Due to lack of open annotated clinical data, some researches are focused on setting frameworks with data augmentation processes. Ma et al. [8] propose an adventitious lung sound classification model, LungRN+NL, which has demonstrated a drastic improvement compared to our previous work and the state-of-the-art models. The model has incorporated the non-local block in the ResNet architecture. To address the imbalance problem and to improve the robustness of the model, the mix up method was incorporated to augment the training dataset. As a result, LungRN+NL has achieved a performance score of 52.26%.

Fraiwan et al. [2] investigate 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. 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.

Also, in another work Fraiwan et al. [3] conduct to explore the ability of deep learning in recognizing pulmonary diseases from electronically recorded lung sounds. Initially, all signals were checked to have a sampling frequency of 4 kHz and segmented into 5 s segments. Then, several preprocessing steps were undertaken to ensure smoother and less noisy signals. These steps included wavelet smoothing, displacement artifact removal, and z-score normalization. The deep learning network architecture consisted of two stages; convolutional neural networks and bidirectional long short-term memory units. The developed algorithm achieved the highest average accuracy of 99.62% with a precision of 98.85% in classifying patients based on the pulmonary disease types using CNN + BDLSTM.

So, a lot of works have been devoted to the study and improvement of methods for working with lung auscultation sounds. Some of them focused on data preprocessing, while others explored new algorithms. Most of the work was done with one of the few relatively large ICBHI datasets available, but some work also used private clinical datasets.

Unfortunately, the problem of the availability of medical data for independent researchers is very significant and does not allow the field to develop at a rapid pace. On the other hand, most known methods of data augmentation can degrade the quality of sensitive medical data. In this work, we plan to explore the possibility of using augmentation by synthesizing data according to the existing distribution, rather than standard methods of augmentation and further generalization of the resulting model.

The research was conducted in accordance with the principles of bioethics set out in the "Universal Declaration on Bioethics and Human Rights" (UNESCO).

MATERIALS AND METHODS OF RESEARCH

Dataset

The main data source for our study was an open dataset from the International Conference on Biomedical and Health Informatics [14]. At the time of the study, this dataset is one of the largest publicly available sources with audio recordings of lung auscultation. The Respiratory Sounds Database contains sound samples collected independently by two research teams in two different countries over several years.

The database consists of a total of 5.5 hours of recordings containing 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes, in 920 annotated audio samples from 126 subjects.

The cycles were annotated by respiratory experts as including crackles, wheezes, a combination of them, or no adventitious respiratory sounds. The recordings were collected using heterogeneous equipment and their duration ranged from 10 s to 90 s. In addition, for each of the 126 subjects, a corresponding association of lung audio recordings with the diagnosis was established. The dataset contains subjects with such diagnosis: COPD, healthy, URTI, bronchilectasis, pneumonia, bronchiolitis, LRTI and asthma.

Auscultation data for each patient were collected at different locations and on different types of recording equipment: AKG C417L Microphone, 3M Littmann Classic II SE Stethoscope, 3M Litmann 3200 Electronic Stethoscope and Welch Allyn Meditron Master Elite Electronic Stethoscope. In this work, 917 records were selected and records of patients with asthma and LRTI were removed as unrepresentative within the task.

Preprocessing

As with any other electronically recorded sound, audio lung auscultations have a set of distortions, additional noises, ambient sounds, and so on. This directly affects the quality of the information content of the sample. Audio is also affected by the device used for recording, each of which has its own amplitude-frequency response and recording environment.

Raw audio available in the dataset has different sampling rates from 4 kHz to 44.1 kHz. The frequency band in which useful lung noise can be localized is 100-1500 Hz. After that, the useful band was filtered out of the 5-th order Butterworth bandpass filter and the extra frequencies that did not carry useful information were cut off.

In this study, experiments were performed with spectral characteristics obtained after preliminary standardization, and the most appropriate for the task was the use of meltspectrograms. A melspectogram is a visual representation of the spectrum of a sound on the Mel scale. The conversion of frequencies into Mels proceed by this formula (1):

$$\boldsymbol{m = 2595 \cdot log_{10}\left(1 + \frac{f}{700}\right)}, \qquad (1)$$

where f is frequency in Hz.

The melspectrogram choice may be justified by the presence of the vast majority of continuous lung sounds with melodic characteristics, rather than discontinuous sounds. The number of frames of 2048 with a hop of 256 was used to generate meltspectrograms.

The peculiarity is that for further application of data augmentation using conditional variational autoencoder (CVAE) and the use of convolutional neural networks, the input data must be reduced to the same dimension, but audio of different lengths are transformed into spectrograms of different sizes. To solve this issue, the obtained spectrograms were linearly interpolated to the same size, based on the average size of the resulting spectrogram.

The obtained spectrograms were normalized in the range from 0 to 1 using min-max normalization (2):

$$\boldsymbol{x' = \frac{x - x_{min}}{x_{max} - x_{min}}}, \qquad (2)$$

Data Augmentation

Like most publicly available medical datasets, we face the problem of imbalance. In this case, the set of lung auscultation data is formed in such a way that for one disease subjects and audio recordings are much larger than others. Thus, the number of samples of the COPD class is 86.5% of the total.

There are many solutions to the problem of unbalanced classes. Limited medical data and the need for their rational use does not allow the use of oversampling and undersampling methods. It may also be inappropriate to use a weighted target variable, given the size of the dataset. The most obvious option for dataset balancing is data augmentation.

For the task of classification of respiratory diseases, data enrichment can be used within time and frequency domains. The paper proposes the use of a full conveyor of work with the frequency representation of lung auscultation.

The generated normalized spectrograms can be perceived as single-channel images of fixed size, where each pixel reflects the power of the frequency spectrum at a certain point in time. Melspectrograms of each class have common characteristics that can be used to generate synthetic data of little represented classes. This method is more stable to retraining for oversampling.

To generate synthetic data, the paper proposes to use the architecture of autoencoders, namely convolutional variational autoencoders (VAE) [6].

An autoencoder network is a pair of two connected networks, an encoder and a decoder. An encoder network takes in an input, and converts it into a smaller, dense representation, which the decoder network can use to convert it back to the original input. VAEs have one fundamentally unique property

that separates them from vanilla autoencoders, and it is this property that makes them useful for generative modelling: their latent spaces are continuous, allowing easy random sampling and interpolation. Convolutional vibrational autoencoders differ in that convolution layers are used as starting input and output layers for the formation and subsequent coding to the latent space of spatial characteristics and subsequent decoding of the spatial tensor. Main scheme of CVAE model shows in Figure 1.

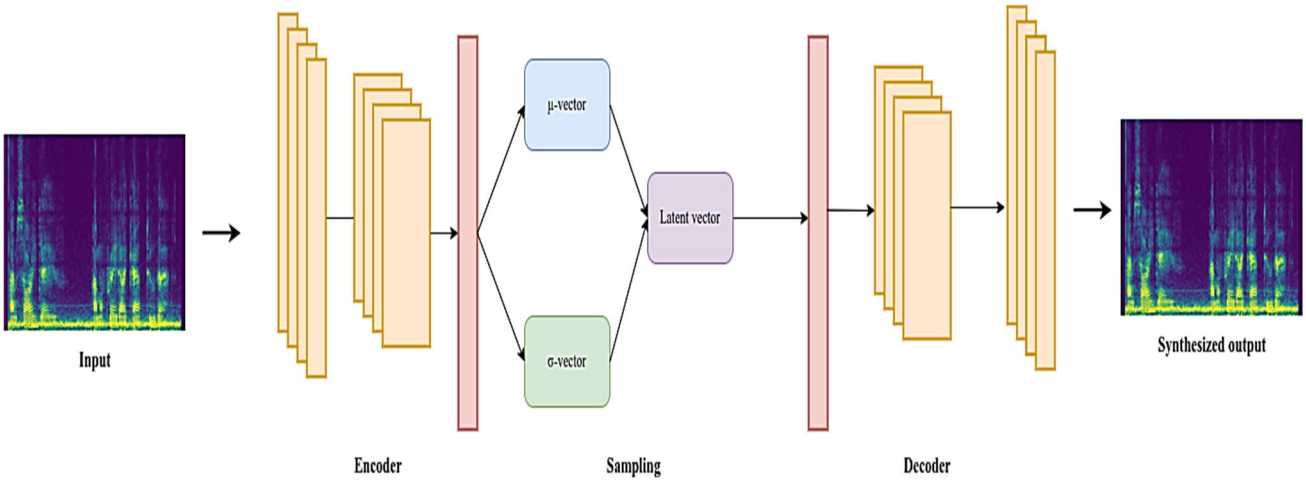


Fig. 1. Illustration of Convolutional Variational Autoencoder pipeline and its main structure

Variation autoencoders can be positioned as a probabilistic model that aims to simulate the distribution of data for further generation of synthetic data that could belong to the original distribution. The loss function of VAEs consists of two weighted parts: reconstruction term, which is responsible for the conditional similarity of the input tensor to the output, and regularization term, which aims to reduce the distribution of latete space generated by encoders to normal distribution. MSE was used as the reconstruction term, and the Kulback-Leibler divergence was used as the regularization term. The loss function is as follows [15]:

$$\mathcal{L}_{VAE} = ||\boldsymbol{x} - \hat{\boldsymbol{x}}||^2 + \boldsymbol{D_{KL}}(\boldsymbol{P}||\boldsymbol{Q}), \quad (3)$$

where x^ is the reconstruction of x and DKL (P||Q), P is a distribution N(μx,σx) with original mean and standard deviation, and Q is a target normal distribution.

The Kulback-Leilber divergence defined below (4):

$$\boldsymbol{D_{KL}}(\boldsymbol{P}||\boldsymbol{Q}) = -\int \boldsymbol{p}(\boldsymbol{x})\boldsymbol{log}\boldsymbol{q}(\boldsymbol{x})\boldsymbol{dx} + \\ +\int \boldsymbol{p}(\boldsymbol{x})\boldsymbol{log}\boldsymbol{p}(\boldsymbol{x})\boldsymbol{dx}, \quad (4)$$

Thus, using convolutional variational autoencoders, it is possible to synthesize data of each class by encoding normalized melspectrograms to a normally distributed latent space with subsequent decoding of vectors to images that could belong to the distribution of images of the target class.

Training and Validation

Using the concept of spectral analysis of lung auscultation allows you to treat real and enriched spectrograms as images. Convolutional neural networks (CNN) are a common method of analysing the spatial characteristics of the image and generalizing the obtained representations. CNN are a deep learning algorithm for processing two-dimensional representations of input data that automatically process complex data dependencies using trainable kernel sets.

In this study, the standard architecture of convolutional neural networks is used. This architecture involves the use of conventional convolutional layers adjacent to the pooling layers. Convolutional layers allow to generalize the spatial characteristics of the spectrogram and find complex dependencies in the data, while pooling layers provide a combination of the found dependencies and the transition to higher-level analysis.

To ensure the stability of the neural network optimization algorithm, each convolutional layer is also bordered by a batch normalization layer, which ensures normal weight distribution and stabilizes the spread of gradients. The ReLU nonlinearity function is used to expand the hypothesis space.

After reaching the desired level of generalization of the characteristics, the original data of the convolutional neural network is transformed into a linear space and processed by linear layers. Next, the space of linear characteristics is reduced to a layer with the number of neurons corresponding to the number of classes, in our case 5, which calculates the

probability of belonging of the input data tensor to a particular class. During the experiment between the linear layers, Dropout layers with a p = 0.3 were also added to prevent overfitting. The basic model of the network used is shown in the figure.

Cross-entropy, which works with the initial probabilities of the model, was chosen as a function of losses for the problem of multiclass classification (5):

$$\mathcal{L} = -\Sigma_{c=1}^{M} y_{o,c} log(p_{o,c}), \quad (5)$$

where M is number of classes, y is binary indicator 0 or 1 if class label c is the correct classification for observation o, p is predicted probability observation o is class c.

As an algorithm for weight optimization, Adam optimizer was used as a proven solution with learning rate l=0.001. In addition, for better network convergence, an exponential learning rate reduction scheduler with γ=0.9 was used. The training followed a 10-fold cross-validation to ensure coverage of all possible combinations within the dataset. The full data pipeline shows in Figure 2.

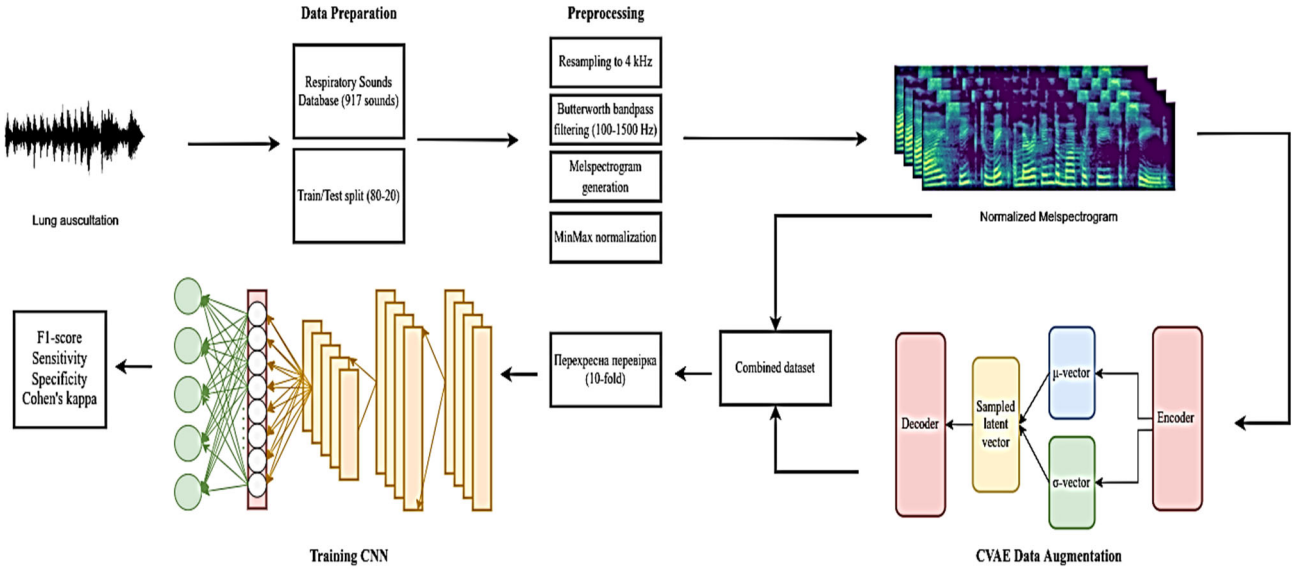


Fig. 2. Data pipeline of proposed disease classification method

RESULTS AND DISCUSSION

Dataset Preparation

First, we analyse the distribution of audio recordings of lung auscultation and their affiliation to subjects of a certain class of disease. The Table 1 shows the quantitative characteristics of the available audio according to the diagnosis:

Table 1

Number of audiosamples for each diagnosis

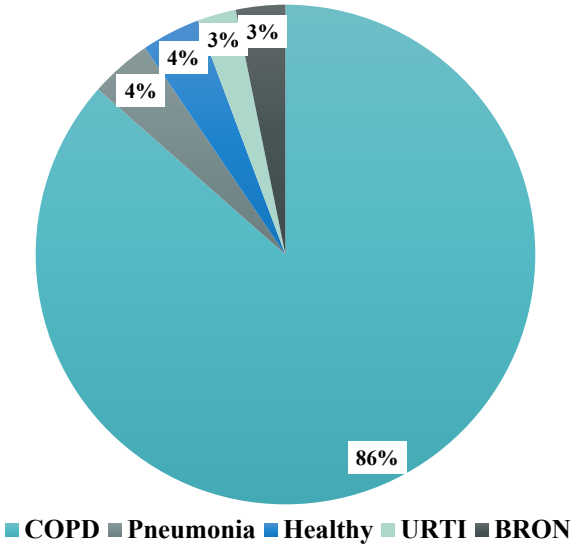
Disease	#
COPD	**793**
Pneumonia	**37**
Healthy	**35**
URTI	**23**
Bronchiectasis	**16**
Bronchiolitis	**13**
LRTI	**2**
Asthma	**1**

There is insufficient data for characterization and generalization, as well as for possible synthesis and augmentation for LRTI and asthma class records, so these records will not be used. Also, it was decided to combine bronchial diseases into one Bronchiectasis class as a potentially similar. In Figure 3a shows the distribution of classes with which the research will be performed.

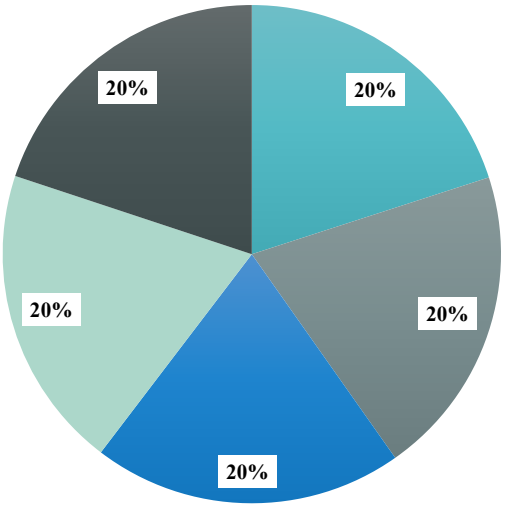
After defining and forming a working dataset, we will apply an additional 10-fold split into training and test sets. Two main types of partitioning were performed to study the influence of different patient audio recordings on the generalization of the model: with and without patient groupings.

For each group of selected training and test sets, the preprocessing steps described above are used to further obtain a melspectral representation. According to the defined parameters of the generation of meltspectrograms and their reduction to the same size, an image of 128x336 pixels with normalized values was obtained.

After performing the pre-processing step, it is necessary to augment the training set. For Melspectrograms of each target class, a copy of CVAE is trained with certain parameters for 100 epochs. After obtaining a correct copy of the augmentation model, synthetic meltspectrograms of lung auscultation are generated. After obtaining a correct instance of the augmentation model, synthetic meltspectrograms of lung auscultation are generated according to the percentage of unbalanced dataset so that the number of instances of each class is approximately equal (Fig. 3b).



(a)



(b)

Fig. 3. Class distribution in: (a) unbalanced dataset and (b) CVAE-augmented dataset

Finally, a joint dataset of real and synthetic data is formed, which is ready for further modelling.

Performance Evaluation

Several basic metrics were introduced to test the quality of the model and to analyse the confusion matrix. The confusion matrix was generated sequentially after every fold, and all evaluation metrics were calculated from the overall confusion matrix after the tenfold cross-validation of the training/classification scheme.

Standard valuation methods are formed using certain error values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). As we can see in the Formulas (6) – (9), the paper used traditional methods for assessing the quality of classification for medical datasets: sensitivity, specificity, f1-score and Cohen's kappa:

$$\boldsymbol{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}, \quad (6)$$

$$\boldsymbol{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}}, \quad (7)$$

$$\boldsymbol{F1_{score}} = \frac{\boldsymbol{2TP}}{\boldsymbol{2TP \mid FP \mid FN}}, \quad (8)$$

$$\boldsymbol{k} = \frac{\boldsymbol{P_o - P_c}}{\boldsymbol{1 - P_c}}, \quad (9)$$

where P_o is the observed agreements and P_c is the agreements expected by chance.

An additional metric that allows evaluating the quality of the model relative to the participants in the ICBHI competition is the score (10):

$$\boldsymbol{Score} = \frac{\boldsymbol{Sensitivity+Specificity}}{\boldsymbol{2}}, \quad (10)$$

Experiments

The proposed method was entirely implemented using Python 3.7 and PyTorch framework. The experiments were conducted on an Intel processor (i7-9700) with 32 GBs of RAM. The training process was performed on NVIDIA Tesla T4 graphics processing unit (GPU) of 16 GBs display memory (VRAM). Each fold was synthesised for 10 minutes and trained a total 25 minutes to complete the whole training/classification scheme. The prediction of per-patient class took less than a second under the aforementioned machine specifications.

All experiments were separated into two parts: with stratified split into training and test sets taking into account grouping by subjects and stratified division without audio grouping by belonging to a particular subject.

In the first case, the training and test sets have intersections on patient IDs. The division takes place using stratification, so the distribution of classes in the training and test sets remains unchanged. All audio recordings are pre-processed. Next is the process of synthesis of spectrograms of minor classes using the data augmentation phase described above. Demonstration of data synthesis based on trained CVAE for target classes is shown in Figure 4.

Using the above metrics, the class evaluation of the proposed model is shown in Table 2.

High class specificity indicates a sufficient quality of imitation of patterns of pathologies at the stage of augmentation. The highest specificity values for the URTI, Pneumonia and Bronchiectasis classes indicate them as the classes that differ most in the presented features. The highest F1-score and the sensitivity for the URTI class, which was not augmented, indicates the greatest variety of samples among other classes that provides generalization of characteristics within the class.

But rather mediocre general results on cross-checking indicate the existing shortcomings of this method. So, obviously, one of the disadvantages is the relative lack of generalization, which can be explained by the need to synthesize large amounts of data from the available usually quite a small amount of data.

The second stage of the experiments is carried out with small differences from the first, namely with the division into training and test sets so that the audio recordings of subjects from the test set do not intersect with the audio recordings of subjects of the training set. Data collected from a single entity is more likely to be recorded at the same time, with the same background noise etc., this can lead to uncontrolled data leakage, so special attention should be paid to the formation of datasets.

After completing the process of creating a training and test dataset in accordance with the goal, the experiments were repeated with a new distribution of data and using ten-fold cross-validation. The results are shown in Table 3.

The results are significantly different from those obtained in the first stage of the experiments, so we can assume that without using grouping there is a certain data leakage from the training to the test dataset, which is not corrected by using only cross-validation. The relatively high F1-score of 91.93% and 93.08% sensitivity for the COPD class also indicates that there were enough specimens of this class for a satisfactory COPD classification, while augmentation of other classes with relatively narrow distributions was insufficient to generalize characteristics and further classification.

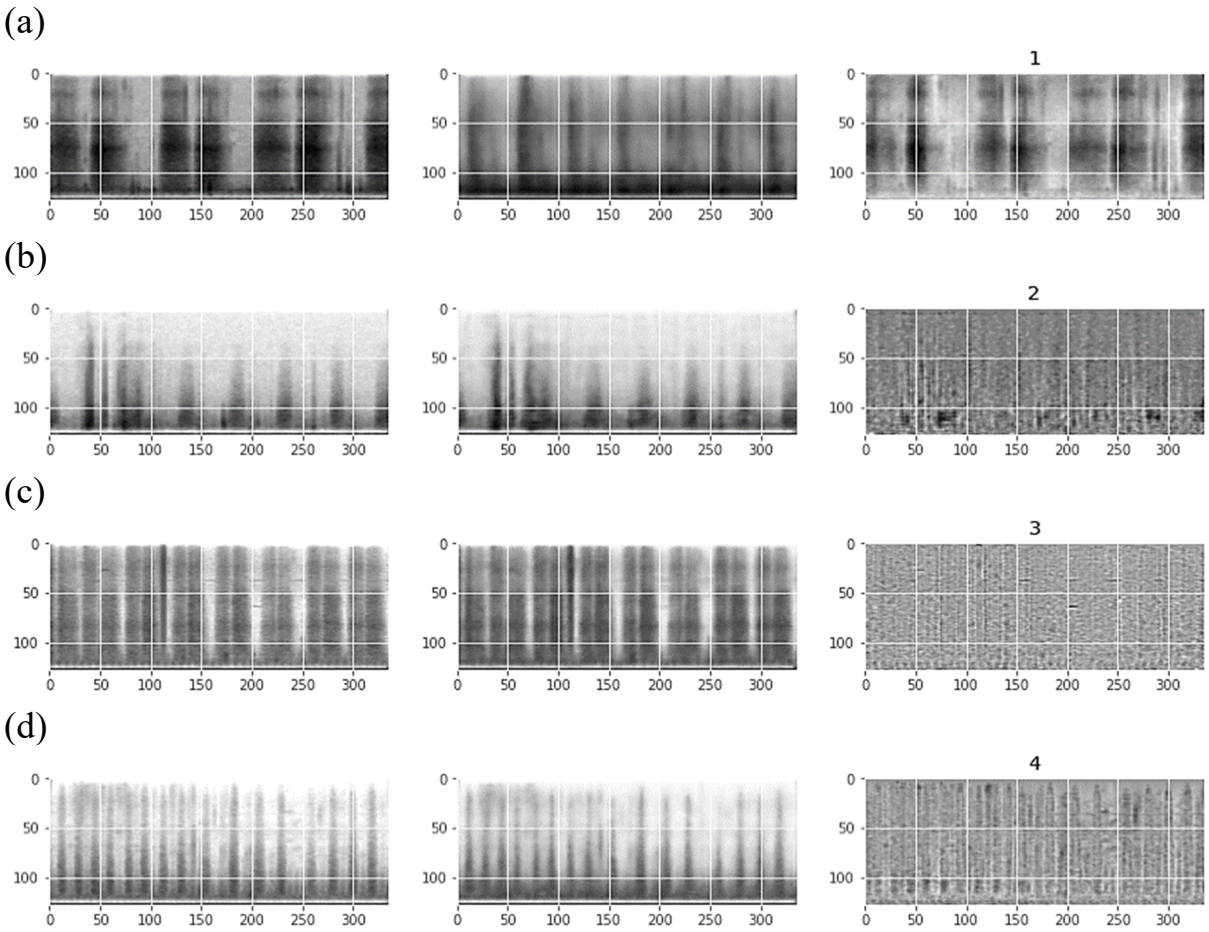


Fig. 4. Synthesized training samples in the process of date augmentation using CVAE. Left: original sample, middle: synthesized, right: difference. The samples are shown for classes (a) Bronchiectasis, (b) Pneumonia, (c) Healthy, (d) URTI

The generated synthetic data from the spectrograms of minor classes are of sufficient quality for use, but, unfortunately, the limitations of the data set and variations in the technical characteristics of recording devices in the available data set will not allow one to unambiguously learn the distribution and generalize on completely new data, which partly explains the results of the second stage of experiments.

Thus, with a careful approach to splitting the data into training and test sets, taking into account the belonging of spectrograms to certain patients, it prevents leakage of useful characteristics between the sets (Tab. 2, 3)

Table 2

Performance evaluation (in %), of proposed model without patient IDs grouping based on tenfold cross validation

Diagnosis	Sensitivity	Specificity	Accuracy	Cohen's kappa	F1-score
Healthy	**98.46**	**99.61**	**-**	**-**	**98.46**
COPD	**100.00**	**98.45**	**-**	**-**	**96.96**
Bronchiectasis	**95.31**	**100.00**	**-**	**-**	**96.06**
Pneumonia	**98.46**	**100.00**	**-**	**-**	**99.22**
URTI	**100.00**	**100.00**	**-**	**-**	**100.00**
Overall (avg)	**98.44**	**99.61**	**98.44**	**98.05**	**98.45**

Table 3

Performance evaluation (in %), of proposed model with patient IDs grouping based on tenfold cross validation

Diagnosis	Sensitivity	Specificity	Accuracy	Cohen's kappa	F1-score
Healthy	28.57	90.34	-	-	26.67
COPD	93.08	40.17	-	-	91.93
Bronchiectasis	16.67	100.00	-	-	28.57
Pneumonia	22.22	95.95	-	-	26.67
URTI	0.00	88.65	-	-	0.00
Overall (avg)	32.11	83.02	82.70	27.48	34.76

On the other hand, a similar approach to data augmentation worked with ten-fold cross-validation without explicit separation by patient and allowed us to generalize the characteristics for minor classes. This indicates a hypothetical application of the augmentation method using variational autoencoders with the possible use of other characteristics or in conjunction with other classifier architectures.

CONCLUSION

1. This paper proposes a framework for processing and analysing lung auscultation sounds to facilitate the diagnosis of diseases. The framework involves working with an unbalanced set of sensitive medical data. A method for data preprocessing based on filtering non-informative frequencies, as well as converting audio recordings into a spectral form, is presented. Also presented is a method for synthesizing medical data using convolutional variational autoencoders. To perform the classification task, a standard convolutional neural network architecture was used. Frameworks like this are characterized by fast learning and prediction speed using GPUs, as well as the ability to quickly adapt the architecture to the tasks.
2. The experiment was conditionally divided into 2 parts: with stratified grouping of audio by subjects and only stratification. The best result in 10-fold cross-validation was demonstrated with stratification alone. 98.44% accuracy, 98.05% Cohen's kappa and 98.45% F1-score were achieved. At the same time, the idea of using CVAE as an element of minor class augmentation was tested.
3. The results of the second part of the experiments are significantly different. The reason for this may be the insufficient amount of data for a full-fledged modelling of the distribution of data of minor classes, as well as the imperfection of the approach used, which is based only on one type of features (Melspectrograms). On the other hand, some of works on this topic do not pay attention to the approach to the formation of stratified and grouped training and test sets, which can lead to unrepresentative results.
4. In our opinion, the ability to find a mechanism for augmenting sensitive medical data can help accelerate the progress of improving algorithms. One such possibility is to simulate a data distribution and synthesize new instances from that distribution. Future work may be aimed at improving the data augmentation mechanism, as well as using a different set of characteristics for this purpose and, as a result, other more advanced approaches to classification, and it is also necessary to take into account the different amplitude-frequency characteristics of recording devices.
5. In addition, in future works it is planned to test the proposed augmentation method on a wider data set, where, as a result of data processing and selection, the simulation of data distribution will be much closer to useful characteristics than to random forming circumstances.

Contributors:

Panaskin D.V. – conceptualization, methodology, writing – review & editing, visualization;

Stirenko S.H. – investigation, data curation, formal analysis;

Babko D.S. – software, resources, validation, visualization.

Funding. This research received no external funding.

Conflict of interests. The authors declare no conflict of interest.

REFERENCES

1. Abbas A, Fahim A. An automated computerized auscultation and diagnostic system for pulmonary diseases. J Med Syst. 2010;34:1149-55.
doi: https://doi.org/10.1007/s10916-009-9334-1

2. Fraiwan L, Hassanin O, Fraiwan M, Khassawneh B, Ibnian AM, Alkhodari M. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. Biocybern Biomed Eng. 2021;41(1):1-14.
doi: https://doi.org/10.1016/j.bbe.2020.11.003

3. Fraiwan M, Fraiwan L, Alkhodari M, Hassanin O. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. J Ambient Intell Humaniz Comput. 2021;13(10):1-13. doi: https://doi.org/10.1007/s12652-021-03184-y

4. Gibson GJ, Loddenkemper R, Lundback B, Sibille Y. Respiratory health and disease in Europe: The new European lung white book. Eur Respir J. 2013;42(3):559-63. doi: https://doi.org/10.1183/09031936.00105513

5. Kim Y, Hyon Y, Jung SS, Lee S, Yoo G, Chung C, et al. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Sci Rep. 2021;11(1):1-11.
doi: https://doi.org/10.1038/s41598-021-96724-7

6. Kingma DP, Welling M. Stochastic gradient vb and the variational auto-encoder. In: Second International Conference on Learning Representations. 2014. p. 121.

7. Lema GF, Berhe YW, Gebrezgi AH, Getu AA. Evidence-based perioperative management of a child with upper respiratory tract infections (urtis) undergoing elective surgery: A systematic review. Int J Surg Open. 2018;12:17-24. doi: https://doi.org/10.1016/j.ijso.2018.05.002

8. Ma Y, Xu X, Li Y. LungRN+ NL: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation. In: Interspeech; 2020.
doi: https://doi.org/10.21437/Interspeech.2020-2487

9. Minami K, Lu H, Kim H, Mabu S, Hirano Y, Kido S. Automatic classification of large-scale respiratory sound dataset based on convolutional neural network. In: 2019 19th International Conference on Control, Automation and Systems (ICCAS). 2019. p. 804-7.
doi: https://doi.org/10.23919/ICCAS47443.2019.8971689

10. Monaco A, Amoroso N, Bellantuono L, Pantaleo E, Tangaro S, Bellotti R. Multi-time-scale features for accurate respiratory sound classification. Appl Sci. 2020;10(23):8606.
doi: https://doi.org/10.3390/app10238606

11. Pham L, Phan H, Palaniappan R, Mertins A, McLoughlin I. Cnn-moe based framework for classification of respiratory anomalies and lung disease detection. IEEE J Biomed Health Inform. 2021;25(8):2938-47.
doi: https://doi.org/10.1109/JBHI.2021.3064237

12. Priftis KN, Hadjileontiadis LJ, Everard ML. Breath Sounds. Springer; 2018. 319 p.
doi: https://doi.org/10.1007/978-3-319-71824-8

13. Reyes B, Charleston-Villalobos S, Gonza´lez-Camarena R, Aljama-Corrales T. Assessment of time–frequency representation techniques for thoracic sounds analysis. Comput Methods Programs Biomed. 2014;114(3):276-90.
doi: https://doi.org/10.1016/j.cmpb.2014.02.016

14. Rocha BM, Filos D, Mendes L, Serbes G, Ulukaya S, Kahya YP, et al. An open access database for the evaluation of respiratory sound classification algorithms. Physiol Meas. 2019;40(3):035001.
doi: https://doi.org/10.1088/1361-6579/ab03ea

15. Sarkar M, Bhardwaz R, Madabhavi I, Modi M. Physical signs in patients with chronic obstructive pulmonary disease. Lung India. 2019;36(1):38-47.
doi: https://doi.org/10.4103/lungindia.lungindia_145_18

16. Tasar B, Yaman O, Tuncer T. Accurate respiratory sound classification model based on piccolo pattern. Appl Acoust. 2022;188:108589.
doi: https://doi.org/10.1016/j.apacoust.2021.108589

17. Wu YS, Liao CH, Yuan SM. Automatic auscultation classification of abnormal lung sounds in critical patients through deep learning models. In: 2020 3rd IEEE International Conference on Knowledge Innovation and Invention. 2020. p. 9-11.
doi: https://doi.org/10.1109/ICKII50300.2020.9318880

Стаття надійшла до редакції 04.04.2024;
затверджена до публікації 31.08.2024