Detecting the Absence of Lung Sliding in Ultrasound Videos Using 3D Convolutional Neural Networks

Michal Kolárik, Martin Sarnovský, Ján Paralič

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 04001 Košice, Slovakia; E-mail: michal.kolarik@tuke.sk, martin.sarnovsky@tuke.sk, jan.paralic@tuke.sk

Abstract: During recent years, deep learning models proved to be very effective in multiple tasks in medicine and frequently outperformed other traditional machine learning methods. Especially in tasks where the processing of image or video data is necessary, deep networks present a very popular tool. Especially in medicine, image and video data present a frequent source of data. The work presented in this paper focuses on the use of deep learning models to detect specific phenomena from lung ultrasonography data. We focused on the detection of lung sliding, which can be observed from such data and used by clinicians in the diagnostic process of evaluating of patient's health condition. Previous research in this area mostly focused on processing a sequence of static images obtained from ultrasonography. In our work, we focused on the development of deep learning models able to process short video sequences. We used different architectures of a Resnet model and experimentally evaluated them on a real-world dataset. Then, we compared the results of best-performing architectures with the more traditional approach based on static image processing.

Keywords: deep learning; neural networks; image classification; video classification; medicine; ultrasonography

1 Introduction

Lung diseases are among the most common pathological conditions worldwide. Therefore, accurate diagnosis and monitoring of lung diseases is of general interest. Also, for the reason of the pandemic caused by the SARS-COV-2 virus, there is a great motivation to collect clinical data. This also results in increasing interest in the academic field of artificial intelligence to create intelligent solutions using machine learning [1]. Standard scanning techniques for diagnosing and monitoring lung pathologies include chest X-rays, computed tomography (CT), and magnetic resonance imaging (MRI). However, these diagnostic approaches are based on ionizing radiation. This poses a risk to the patient's health, especially in the case of

a high dose or frequent exposure in the context of more frequent monitoring. Ultrasonic technologies can overcome the limitations of the mentioned technologies and thus provide a safer, portable, and cost-effective alternative [2].

Deep learning (DL) is a subfield of machine learning (ML) that attempts to learn abstractions at a higher level using hierarchical neural network architectures [3]. It is currently a very popular field, widely applied in traditional tasks of artificial intelligence (AI), such as text mining and natural language processing [4], [5], image processing, and computer vision [6]-[8], or speech recognition [9]. In medicine, DL proved to be very effective for multiple diagnostic tasks, including medical image analysis [10], [11]. With the increasing availability of sufficient computational resources, DL methods are becoming more applicable for video processing tasks as well. In comparison to images, video can provide important information about the evolution of the studied area over time. Therefore, we can track the movement of objects or the temporal appearance of an object, which can be challenging to obtain from static images. Short video sequences are also a frequent data source in medicine.

One of the sources of video data in medicine is ultrasonography. Our work focuses on the short videos obtained from the lung ultrasound. Our primary motivation is to train the deep learning model to detect the lung sliding phenomenon in the lung ultrasound data. Contrary to other published research, which uses mostly transformed static images (see Section 2), we used a 3D convolutional neural network capable of video processing. We implemented the Resnet model, experimentally evaluated multiple topologies of the model using differently preprocessed data and compared its performance to previous research.

This article is organized as follows. Section 2 provides an overview of the current state of the art in the analysis of ultrasound images using deep neural networks. The following (third) section describes the used data and models, and the fourth section summarizes the experiments and results. The final (fifth) section summarizes the main conclusions and sketches future work.

2 Related Work

When analyzing ultrasound (USG) images, the diagnostician observes certain signs (artifacts) that help to evaluate the patient's state of health. In the case of the lungs, these are mainly artifacts such as *lung sliding*, *A-lines,* and *B-lines*. Movement or *lung sliding* is a natural artifact that occurs in humans. The parietal and visceral pleura tightly adhere in a healthy individual with a minimal amount of fluid between them. This fact allows them to slide over each other during breathing. During the diagnostic process, a clinician may want to examine the patient for lung sliding (or its absence), which can be a symptom of one or more lung diseases. This backward

and forward movement of the pleura can be observed on images in *B-mode* or *M-mode*.

M-mode presents images that are a 2D time-slice of a video frame in the lung region. Two patterns can be observed in these images: a *barcode* or a *seashore pattern*. These patterns are an essential feature in the classification of *M-mode* images, whether the sliding movement of the lungs occurs on the given video (*seashore pattern*) or not (*barcode pattern*) [12]. Lichtenstein et al. in work [13] classified the sliding movement of the lungs in patients arriving at the intensive care unit (ICU) with acute respiratory problems while also using the mentioned patterns in *M-mode* images. Subsequently, in combination with other monitored symptoms on ultrasound, they could diagnose lung diseases with high accuracy. In Figure 1, we can see a comparison of these two patterns and an example of the selected slice of the *B-mode* image on the *M-mode* image over time.

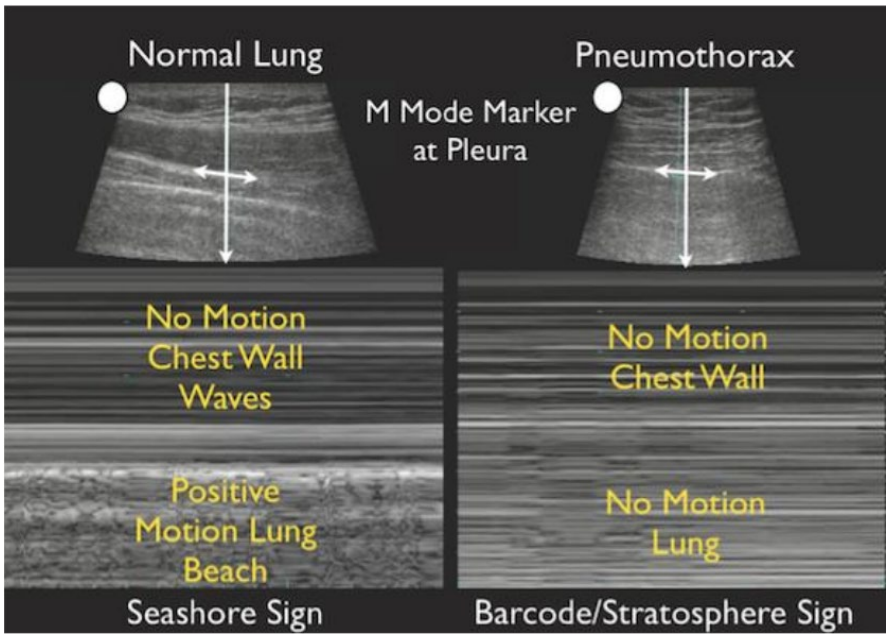

Figure 1
Example of B-mode and M-mode with seashore and barcode sign

Jaščur et al. [14] used deep learning methods for lung sliding classification on *M-mode* images. The result of their work is a system for automatic detection of *lung sliding*, which frees the doctor from the need to perform manual cuts, while with their neural network model with 64-frame architecture, they achieved 89% classification accuracy, 82% sensitivity level and 92% specificity.

Movahedi et al. [15] also devoted themselves to analyzing videos from a USG on the automatic classification of breast lesions based on videos obtained from an ultrasound examination. Their goal was to design an automated method to assist

radiologists in breast screening and cancer detection. The proposed algorithm was able to reduce the vibrations that arose during the creation of USG images by the doctor in such a way that the given algorithm evaluates every tenth image, regardless of what results in the previous image were achieved. The results of the proposed algorithm achieved better results than some existing breast lesion detection methods. Following works, even if are from the different field (not specifically USG video analysis). In the paper [16], the authors devoted themselves to the classification of atherosclerotic disease based on the segmentation of the boundaries of the middle part of the vessel wall while using the Kalman snake filter, achieving better results than similar works using dynamic programming. Sonka et al. in [17] performed automatic classification of ultrasound videos with the aim of early detection of cardiovascular disease. Their goal was the analysis of brachial ultrasound images while evaluating their performance.

Ouyang et al. in [18] presented an algorithm called EchoNet-Dynamic based on echocardiography videos that outperforms human experts in the tasks of left ventricular segmentation, ejection fraction estimation and cardiomyopathy assessment. Based on repeated measurements, the paper's authors found that the model achieves comparable or smaller variance than human experts, who need years of practice for reliable and correct decision making. The given model can quickly and accurately diagnose cardiovascular diseases in real-time based on several heart cycles, is more reproducible, and quickly identifies subtle changes in the ejection fraction. This published work also includes an annotated echocardiography dataset containing 10,030 videos.

3 Methods

In the presented work, we analyse *B-mode* videos from the USG examination of the lungs to classify the presence or absence of *lung sliding*. USG examination of the lungs is a relatively new but effective lung examination technique, the main advantage of which is the absence of unnecessary transport of the patient as well as the absence of radiation to the patient, which occurs when using a conventional X-ray examination of the chest. The analysis of such images from an ultrasonographic examination needs a certain level of expertise of the radiologist operating such a system [14].

In Figure 2, we can see what a video image from an ultrasonographic examination of the lungs looks like. The green marked area represents the ribs, blue represents the lung region, and red represents the pleura.

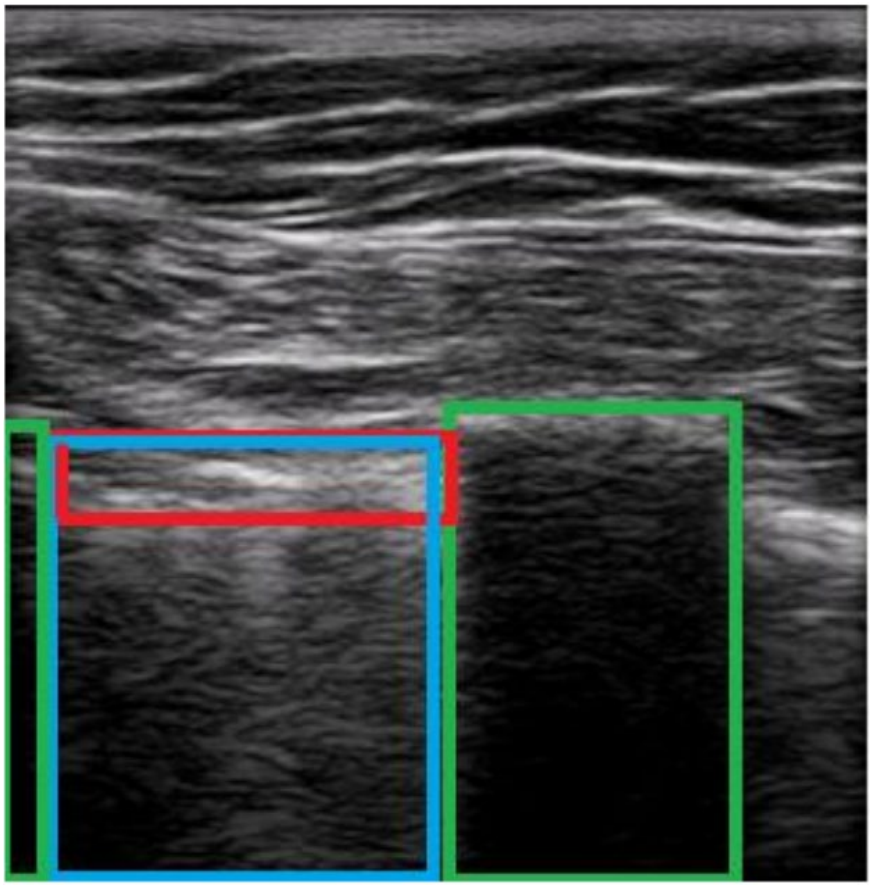

Figure 2
Image from USG video with marked areas of the lung

In this work, we focused on classifying videos containing *lung sliding* and where *lung sliding* is absent. Based on the previous results of works using *M-mode* images, we understood that for the model's decision, an image is needed in time, which creates a *Seashore* or *Barcode* image that is visible in the entire lung area. Therefore, any area of the lung should be suitable for its classification. Moreover, an image of 64 frames in a lung cross-section is already a sufficient sample. We decided to use the information about lung movement directly from the video. In contrast, for classification, we will use only a tiny area of the video in which the lungs are located, which we consider sufficient for the classification of movement or absence of lung movement. The models we will use are 3D convolutional neural networks.

3.1 Dataset Description

In our work, we used the dataset presented in [14], where the given data were collected from an ultrasound examination of the lungs by doctors from the Clinic of Thoracic Surgery of the Jessenius Faculty of Medicine in Martin, Slovakia. The obtained videos are from a pneumothorax examination in which the patient lies on his back, and the doctor scans the front and side of the chest with a linear probe and focuses on three main signs: *sliding lung*, *B-lines,* and *lung point*. All of these videos were recorded in *B-mode*, which captures the impedance of a two-dimensional tissue cross-section. Subsequently, the doctors divided this dataset into two classes, namely the class in which there are videos with the presence of lung sliding movement (28 videos) and the absence of lung sliding movement (20 videos). The dataset contains 48 videos from 48 patients in various stages of postoperative care. The given videos differ from each other in several aspects, specifically in:

- Resolution: 854 x 480 and 640 x 480
- Format: *MP4, AVI*
- Frames per second (FPS): 30 FPS and 54 FPS
- Video length:
  - Lung sliding presence (LSP): 3 – 18 seconds
  - Lung sliding absence (LSA): 2 – 11 seconds

3.1.1 Data Preprocessing

When preprocessing the dataset, we initially focused on unifying the characteristics of the videos. Since the videos contain two different video formats, namely *AVI* and *MP4*, in the first step, we processed these videos to the same type, specifically to the *MP4* format, for easier handling and further work with these videos. We then edited both types of videos to a uniform resolution of 640x480 and 30 FPS, keeping the length of the video in seconds.

In the next step, we cut the videos to a length of 2 seconds (60 frames), which should be sufficient for classifying the presence of lung movement based on the work [5]. In this way, we extended the dataset to 115 videos with the presence of lung movement and 48 videos with the absence of lung movement.

The next step of data preprocessing was to extract from the given videos the areas in which the sliding movement of the lungs occurs. This area is located between the ribs, the entire pleura width (the lung tissue between the two ribs). Since the given videos are shot in different zooms, and there is movement, it is impossible to crop the videos globally.

Based on this need, we created a function that allowed us to manually mark the area of interest in the given videos and extract from this marked area the coordinates we

will need to process the dataset further. In Figure 3, we can see how the marking of the pleural areas took place.

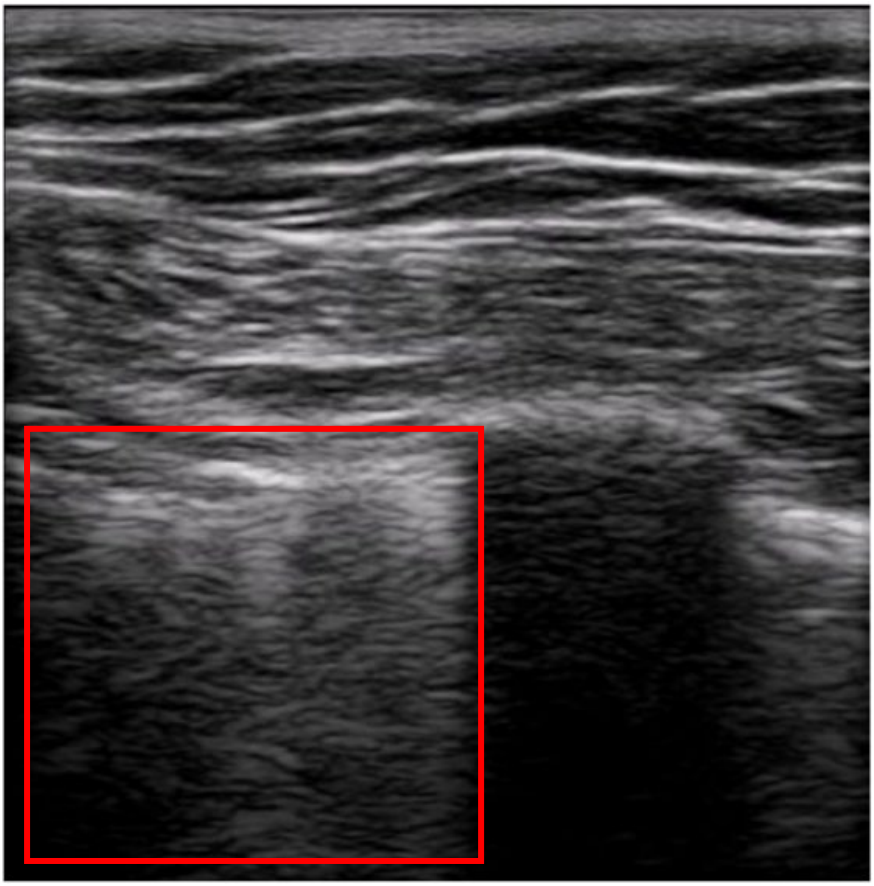

Figure 3
The marked pleural area on the image from the USG lung video

In the selected area of interest, we further divided the video into smaller videos with page dimensions of three alternative sizes: 30x30 pixels, 60x60 pixels, and 90x90 pixels with a length of 60 frames. We created two sets. In the first set of videos, we focused on the upper part of the section only, in which the pleura is located, where the movement of the lungs is most visible. We used the first two rows of videos for this.

In the second set, we focused on the entire selected area and created videos of the entire area. In this way, we obtained six different datasets (Figure 4 visualizes the class distributions in the datasets):

- 2_30x30 (60 frames) – containing 1928 videos (1338 pos., 590 neg.)

- all_30x30 (60 frames) – containing 7 774 videos (5435 pos., 2339 neg.)
- 2_60x60 (60 frames) – containing 890 videos (620 pos., 270 neg.)
- all_60x60 (60 frames) – containing 1647 videos (1155 pos., 492 neg.)
- 2_90x90 (60frames) – containing 663 videos (230 pos., 96 neg.)
- all_90x90 (60frames) – containing 663 videos (470 pos., 193 neg.)

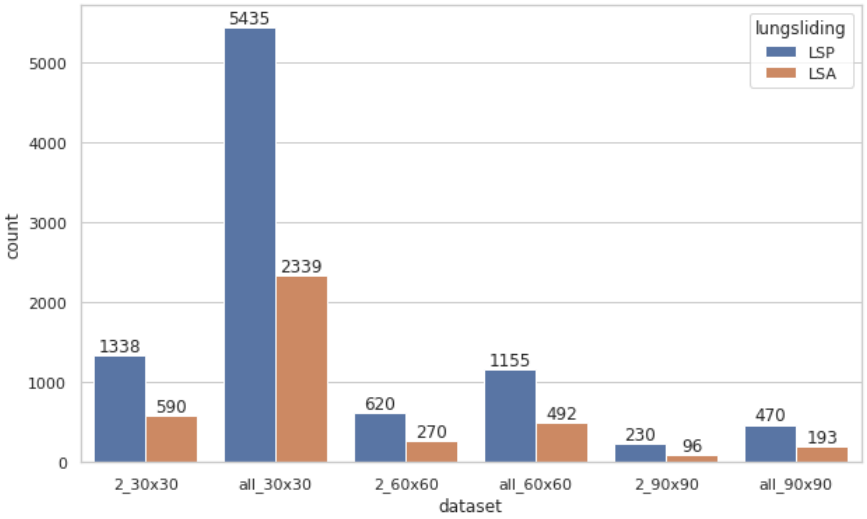

Figure 4

Class distributions in the datasets

These datasets were split into training and test sets in a ratio of 70:30, maintaining a patient-wise split to prevent videos from the same patient from appearing both in the training and test sets.

3.2 Proposed Model

Since lung sliding is a sign that appears over time, it is necessary to work with the time component of the video. One way is to use an *M-mode* image (a slice of one-pixel width and the entire video height in time) or to use a model that can process the whole video as input. In our work, we decided to investigate using the 3D CNN model directly for video processing and motion classification on video.

Considering the success of the Resnet network [19] in work [14], we decided to use Resnet 3D implementation from the freely available library keras-resnet3d[1]. This library allows the processing of 3D image data.

[1] https://github.com/JihongJu/keras-resnet3d

The input dimensions of the model were *N * H * W * 3*, where *3* is the number of channels, *H* represents the video height, *W* is the video width, and *N* is the number of video frames. In our experiments, we decided to compare three architectures, *Resnet3D-18*, *Resnet3D-34,* and *Resnet3D-50*, with 18, 34, and 50 layers respectively. The input goes through a 3D convolution layer, batch normalization, and ReLU activation, followed by a max-pooling 3D layer. They are followed by a series of residual blocks with 3D convolutional layers followed by a 3D average pooling layer. The output of the pooling layer is fully connected to the output of the model. Loss is calculated as binary cross-entropy, and the output is a binary prediction of the model.

After the initial tests of the models, we chose the hyperparameters *batch_size = 32* and the *number of epochs = 16*, *optimizer = Adam* for the experiments, and used the *ModelCheckpoint* function to save the model weights while reducing the error function. Subsequently, the model with the smallest error function was used for evaluation.

4 Experimental Results and Evaluation

We used commonly used classification metrics such as accuracy, sensitivity, specificity, and F1 score to evaluate the results.

4.1 Experimental Results on Datasets 30x30

In the first experiment, we worked with a dataset with a resolution of 30x30 pixels for two rows and all rows from the lung region, respectively. We compared three Resnet3D models on each of the datasets. The results of the first experiment can be seen in Table 1. The best results were achieved by the Resnet3D-18 model on the dataset from the whole lung region (all_30x30). This model has the smallest depth among the compared models but achieved the best results. This case resulted in an accuracy of 89.13%, a sensitivity of 93.60%, and F1 score of 92.38%. Interestingly, this model performed the worst for the dataset of two lines selected around the pleura region.

Table 1
Models' performance on the 30x30 videos

Model	Dataset	Accuracy	Sensitivity	Specificity	F1
Resnet3D-18	2_30x30	0.7047	0.5985	0.9344	0.7349
Resnet3D-34	2_30x30	0.7953	0.7083	**0.9836**	0.8256
Resnet3D-50	2_30x30	0.8420	0.8788	0.7624	0.8838
Resnet3D-18	all_30x30	**0.8913**	0.9360	0.7853	**0.9238**
Resnet3D-34	all_30x30	0.8752	**0.9863**	0.6113	0.9175
Resnet3D-50	all_30x30	0.8849	0.9470	0.7376	0.9205

4.2 Experimental Results on Datasets 60x60

For the experiments on 60x60 pixel videos, we followed the same procedure as with the previous datasets. When using datasets with 60x60 pixel videos, we found that the models used on data containing only the pleura surroundings performed better. However, our results deteriorated when using the entire lung height. The *Resnet3D-34* architecture achieved the best classification results with accuracy of 85.96%, specificity of 89.85%, and F1 score of 94.34%. The best-performing architecture that used the full lung height dataset was Resnet3D-50. It achieved accuracy of 79.09%, sensitivity of 87.22%, and F1 score of 85.16%. The results of the experiments can be seen in Table 2.

Table 2
Models' performance on the 60x60 videos

Model	Dataset		Accuracy	Sensitivity	Specificity	F1
Resnet3D-18	2_60x60		0.7640	0.7143	0.8643	0.8019
Resnet3D-34	2_60x60		**0.8596**	0.8403	**0.8985**	**0.8889**
Resnet3D-50	2_60x60		0.7978	0.9076	0.5765	0.8571
Resnet3D-18	all_60x60		0.6848	0.7048	0.6407	0.7547
Resnet3D-34	all_60x60		0.6970	0.6256	0.8544	0.7396
Resnet3D-50	all_60x60		0.7909	**0.8722**	0.6118	0.8516

Compared to the best model for 30x30 pixel videos, the best model for 60x60 videos performed worse across all metrics.

4.3 Experimental Results on Datasets 90x90

This experiment ran on datasets containing 90x90 pixel videos and full lung height. The experiment confirmed that the simpler neural network architecture achieved the best results, and the deeper the architecture, the worse the classification results, as shown in Table 3. Our best model - *Resnet3D-18*, achieved classification accuracy of 88.15%, specificity of 89.40%, and F1 score of 87.68%. Overall, the models performed better than the previous experiment on the entire 60x60 dataset.

Table 3
Models' performance on the 90x90 videos

Model	Dataset	Accuracy	Sensitivity	Specificity	F1
Resnet3D-18	all_90x90	**0.8815**	0.8768	**0.8940**	**0.9150**
Resnet3D-34	all_90x90	0.8419	0.8518	0.8154	0.8870
Resnet3D-50	all_90x90	0.8267	**0.9875**	0.3960	0.8925

4.4 Experimental Results Comparison

Finally, we decided to compare the results of all our experiments. Comparing the experiments in Table 4, we can see that the best results were obtained by the *Resnet3D-18* model on the 30x30 pixel dataset from the whole lung region, by the *Resnet3D-34* on the 60x60 data of the first two rows, and by *Resnet3D-18* on the 90x90 pixel dataset from the entire lung region. Interestingly, according to the doctors, the area around the pleura is important for deciding lung movement, and the movement should be visible just at the pleura.

Table 4
Summary table of all evaluated models

Model	Dataset	Accuracy	Sensitivity	Specificity	F1
Resnet3D-18	2_30x30	0.7047	0.5985	0.9344	0.7349
Resnet3D-34	2_30x30	0.7953	0.7083	0.9836	0.8256
Resnet3D-50	2_30x30	0.8420	0.8788	0.7624	0.8838
Resnet3D-18	all_30x30	0.8913	0.9360	0.7853	0.9238
Resnet3D-34	all_30x30	0.8752	0.9863	0.6113	0.9175
Resnet3D-50	all_30x30	0.8849	0.9470	0.7376	0.9205
Resnet3D-18	2_60x60	0.7640	0.7143	0.8643	0.8019
Resnet3D-34	2_60x60	0.8596	0.8403	0.8985	0.8889
Resnet3D-50	2_60x60	0.7978	0.9076	0.5765	0.8571
Resnet3D-18	all_60x60	0.6848	0.7048	0.6407	0.7547
Resnet3D-34	all_60x60	0.6970	0.6256	0.8544	0.7396
Resnet3D-50	all_60x60	0.7909	0.8722	0.6118	0.8516
Resnet3D-18	all_90x90	0.8815	0.8768	0.8940	0.9150
Resnet3D-34	all_90x90	0.8419	0.8518	0.8154	0.8870
Resnet3D-50	all_90x90	0.8267	0.9875	0.3960	0.8925

Our experiments' results were then compared with the study by Jaščur et al. [14], which used an M-mode approach to time-lapse video images and their classification using Resnet networks. Table 5 shows the results of comparing our best models with the Resnet M-mode model for image classification.

Table 5
Comparison of the best performing model with the Resnet M-mode model (Resnet-18)

Model	Dataset	Accuracy	Sensitivity	Specificity	F1
Resnet3D-18	all_30x30	**0.9193**	**0.9360**	0.7853	0.9238
Resnet3D-34	2_60x60	0.8596	0.8403	0.8985	0.8889
Resnet3D-18	all_90x90	0.8815	0.8768	0.8940	0.9150
Resnet-18	64-frame	0.8900	0.8600	**0.9200**	-

The best-performing model achieved better accuracy and sensitivity but worse specificity. In the case of sensitivity, we achieved more than 7.5% better. Our model learned better to recognize positive examples, i.e., those for which there was a

sliding lung movement in the video. The metrics are comparable to those achieved by the authors, while we had a smaller amount of data available for the training phase than the authors had for the 64-frame architecture.

4.5 Explainability

After the model training and evaluation, we wanted to explore the possibility of the model explanation. We decided to use the Vanilla saliency map method [20], which generates a heat map to show which part of the input image (in our case, which pixel in the video) affects the output classification. As a second method, we chose SmoothGrad [21], which is based on the saliency map method. It works on a similar principle, but adds noise at the input of its heat map calculations to reduce the noise at the output of this explainability. Thus, it provides more distinct regions on which to base the model's decisions. In Figure 5, we can see the applied explainability methods for a particular video over time. The given frames are six frames apart to show the given explainability for the entire video. The original video model (input) is in the first row. The second row depicts the frames from the video with Vanilla Saliency Map applied, and in the third row, SmoothGrad result is presented. In both cases, the highlighted areas visualize the pixels most significantly contributing to the classification.

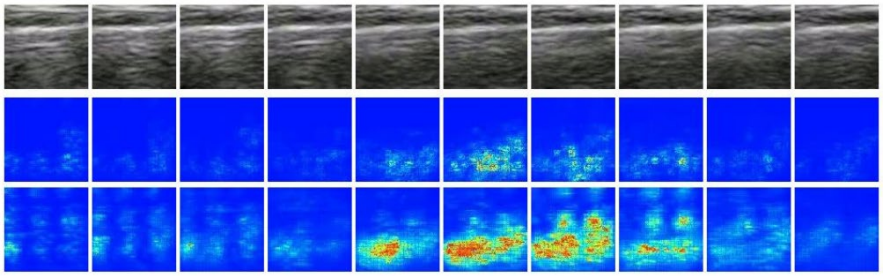

Figure 5
Example of processed frames (first row), and result of XAI methods Vanilla Saliency Map (second row) and SmoothGrad (third row) to each processed frame

Conclusion

This work's goal was to use 3D convolutional neural networks for lung motion classification from USG B-mode videos. We used only a part of the video, thereby obtaining a larger number of samples, reducing the demands on the neural network, and proving that it is possible to replace lung movement in the time domain using M-mode classification directly from the video. We also tried to explain the decisions of the network using XAI methods such as Vanilla Saliency Maps and SmoothGrad. Assuming that additional annotated data from the ultrasonographic examination of the lungs are obtained, further research on the given issue could focus directly on the pleural area and classify the movement of the lungs only from this particular area. This was not possible due to the small amount of input data.

Subsequently, the given experiments presented in the paper could be repeated with additional explainability methods, the use of other neural network architectures, or the use of transfer learning to improve the results of the used network.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-20-0232 and contract No. APVV-17-0550 and by the Slovak VEGA research grant No.1/0685/21.

References

[1] S. B. Desai, A. Pareek, and M. P. Lungren, "Deep learning and its role in COVID-19 medical imaging," *Intell. Med.*, 2020

[2] L. Demi, T. Egan, and M. Muller, "Lung ultrasound imaging, a technical review," *Applied Sciences (Switzerland)* 2020

[3] C. C. Aggarwal and others, *Neural networks and deep learning*. Springer, 2018

[4] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, "Comparison of Deep Learning Models and Various Text Preprocessing Techniques for the Toxic Comments Classification," *Appl. Sci.*, Vol. 10, No. 23, 2020

[5] V. M. Kresnakova, M. Sarnovsky, and P. Butka, "Deep learning methods for Fake News detection," in *IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, CINTI-MACRo 2019 - Proceedings*, 2019

[6] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*. 2018

[7] V. Maslej-Krešňáková, K. El Bouchefry, and P. Butka, "Morphological classification of compact and extended radio galaxies using convolutional neural networks and data augmentation techniques," *Mon. Not. R. Astron. Soc.*, 2021

[8] Š. Mackovjak, M. Harman, V. Maslej-Krešňáková, and P. Butka, "SCSS-Net: solar corona structures segmentation by deep learning," *Mon. Not. R. Astron. Soc.*, 2021

[9] Y. Dokuz and Z. Tufekci, "Mini-batch sample selection strategies for deep learning based speech recognition," *Appl. Acoust.*, 2021

[10] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, 2017

[11] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Lecture Notes in*

Computational Vision and Biomechanics, 2018

[12] S. Singh, H. Kaur, S. Singh, and I. Khawaja, "Basic Insights of Lung Ultrasonography in Critical Care Setting," *Cureus*, 2018

[13] D. A. Lichtenstein and G. A. Mezière, "Relevance of lung ultrasound in the diagnosis of acute respiratory failure the BLUE protocol," *Chest*, 2008

[14] M. Jaščur, M. Bundzel, M. Malík, A. Dzian, N. Ferenčík, and F. Babič, "Detecting the absence of lung sliding in lung ultrasounds using deep learning," *Appl. Sci.*, 2021

[15] M. M. Movahedi, A. Zamani, H. Parsaei, A. T. Golpaygani, and M. R. H. Poya, "Automated analysis of ultrasound videos for detection of breast lesions," *Middle East J. Cancer*, 2020

[16] S. Zhao, G. Li, W. Zhang, and J. Gu, "Automatical Intima-Media Border Segmentation on Ultrasound Image Sequences Using a Kalman Filter Snake," *IEEE Access*, 2018

[17] M. Sonka, W. Liang, and R. M. Lauer, "Automated analysis of brachial ultrasound image sequences: Early detection of cardiovascular disease via surrogates of endothelial function," *IEEE Trans. Med. Imaging*, 2002

[18] D. Ouyang *et al.*, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, 2020

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016

[20] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, 2014

[21] D. Omeiza, S. Speakman, C. Cintas, and K. Weldermariam, "Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models," Aug. 2019