ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION

UDC 004.93'1; 004.932
DOI: 10.37661/1816-0301-2025-22-3-7-24

Original Article
Оригинальная статья

Algorithm for lung pathology detection in X-ray images using binary classification with emphasis on preprocessing

Dzmitry A. Paulenka✉, Aleksandra A. Kosareva, Eduard V. Snezhko, Vassili A. Kovalev

The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
✉*E-mail: dmitri.pavlenko@gmail.com*

Abstract

Objectives. The purpose of the work is automatic detection of lung lesions: cavities, infiltrates, and nodules on chest X-ray images. Also, the possibility of spatial localization of these lesions on the image is investigated.

Methods. Binary classification using deep convolutional neural networks and the Grad-CAM method are used.

Results. For the Xception model, the binary classification accuracy on the test dataset is 73.1% for cavities, 71.9% for infiltrates, and 72.8% for nodules. Heat maps with true positive outcomes for cavities and nodules are mostly understandable to radiologists. More research is needed to get heat maps for infiltrates that are understandable to experts.

Conclusion. The average classification accuracy of the Xception model for three lesion types (cavities, infiltrates, and nodules) is equal to 72.6%. Heat maps associated with pathological processes in the lungs and lesion localization were constructed. Obtained results are good, but not excellent. Thus, further investigation should be done to improve the classification accuracy and quality of the heat maps.

Keywords: medical image processing, image analysis, lung lesions, deep learning, binary classification, computer-aided diagnosis, chest X-ray, Grad-CAM

Acknowledgements. The work was carried out with the financial support of the ISTC-PR150 "Belarus TB Database and TB Portal" project and BRFFR project No. Ф22КИТГ-001 "Targeted cyclic peptide-mediated calcification therapy and early imaging diagnosis of lung cancer". The work was carried out with the funding of the State Program of Scientific Research "Digital and Space Technologies, Human, Society and State Security 2021-2025", subprogram "Digital Technologies and Space Informatics", task 1.1.3. Special thanks to the radiologist expert Oleg V. Tarasov for his valuable expertise during the research. Special gratitude to Mr. Andrei Gabrielian for the helpful questions and feedback.

For citation. Paulenka D. A., Kosareva A. A., Snezhko E. V., Kovalev V. A. *Algorithm for lung pathology detection in X-ray images using binary classification with emphasis on preprocessing*. Informatika [*Informatics*], 2025, vol. 22, no. 3, pp. 7–24. DOI: 10.37661/1816-0301-2025-22-3-7-24.

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

Received | Поступила в редакцию 02.07.2025
Accepted | Подписана в печать 18.07.2025
Published | Опубликована 30.09.2025

Алгоритм обнаружения патологии легких на рентгеновских снимках с использованием бинарной классификации с акцентом на предварительную обработку

Д. А. Павленко✉, А. А. Косарева, Э. В. Снежко, В. А. Ковалев

Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
✉*E-mail: dmitri.pavlenko@gmail.com*

Аннотация

Цели. Осуществляется автоматическое обнаружение поражений легких: полостей, инфильтратов и узелков – на рентгеновских снимках грудной клетки. Также исследуется возможность пространственной локализации этих поражений на изображении.

Методы. Используются бинарная классификация при помощи глубоких сверточных нейронных сетей и метод Grad-CAM.

Результаты. Для модели Xception точность бинарной классификации на тестовом наборе данных составляет: 73,1 % для полостей, 71,9 % для инфильтратов и 72,8 % для узелков. Тепловые карты с истинно положительными результатами для полостей и узелков в основном понятны радиологам. Чтобы получить понятные экспертам тепловые карты для инфильтратов, необходимо провести дополнительные исследования.

Заключение. Средняя точность классификации модели Xception для трех типов поражений (полости, инфильтраты и узелки) равна 72,6 %. Были построены тепловые карты, связанные с патологическими процессами в легких и локализацией поражений. Полученные результаты являются хорошими, но не отличными. Таким образом, необходимо провести дальнейшие исследования для повышения точности классификации и качества тепловых карт.

Ключевые слова: обработка медицинских изображений, анализ изображений, поражения легких, глубокое обучение, бинарная классификация, автоматизированная диагностика, рентгенография грудной клетки, метод Grad-CAM

Благодарности. Работа выполнена при финансовой поддержке: проекта ISTC-PR150 «Белорусская база данных по туберкулезу и туберкулезный портал»; проекта БРФФИ № Ф22КИТГ-001 «Целевая терапия кальцификации, опосредованной циклическими пептидами, и ранняя визуализирующая диагностика рака легких»; ГПНИ «Цифровые и космические технологии, безопасность человека, общества и государства 2021–2025», подпрограмма «Цифровые технологии и космическая информатика», задание 1.1.3. Особая благодарность выражается врачу-рентгенологу Олегу Владимировичу Тарасову за ценную экспертизу во время проведения исследований, а также Андрею Габриэляну за полезные вопросы и отзывы.

Для цитирования. Алгоритм обнаружения патологии легких на рентгеновских снимках с использованием бинарной классификации с акцентом на предварительную обработку / Д. А. Павленко, А. А. Косарева, Э. В. Снежко, В. А. Ковалев // Информатика. – 2025. − Т. 22, № 3. – С. 7–24. – DOI: 10.37661/1816-0301-2025-22-3-7-24.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Introduction. In recent years, artificial intelligence-based clinical decision support systems (AI-CDSS) have gained substantial traction in the medical domain, driven by advancements in deep learning and the increasing availability of annotated medical imaging datasets. These systems are increasingly recognized for their potential to enhance diagnostic precision, reduce inter-observer variability, and streamline clinical workflows. Among various applications, AI-driven solutions for radiological image analysis have emerged as a particularly active area of research and development. Modern convolutional neural networks and vision transformers are being widely employed to perform

tasks such as lesion detection, segmentation, and classification across modalities including computed tomography (CT), magnetic resonance imaging (MRI), and X-ray. Such tools are especially valuable in scenarios requiring rapid and reliable interpretation of complex images, where human expertise may be limited or subject to fatigue-induced errors. Despite promising results, challenges remain regarding model generalizability, interpretability, and integration into clinical practice.

Addressing the challenges of tuberculosis diagnosis is crucial due to its significant impact on public health and the need for timely, reliable interventions. Several approaches have been proposed in the literature for the AI-CDSS aimed at tuberculosis diagnostics.

One of the key challenges in developing systems for delineating pathological areas is the need for labor-intensive manual segmentation, which places a substantial burden on radiologists and hinders the efficient annotation of large-scale medical imaging datasets. There is a clear need for the development of methods capable of identifying pathological regions in the absence of segmentation annotations.

The investigation goal is *to automatically detect lesions and construct heat maps associated with pathological processes in the lungs*. To achieve the goal we trained convolutional neural networks for the automatic detection of lung lesions: *cavities*, *infiltrates*, and *nodules*. Also, we investigated the possibility of spatial localization of these lesions in the chest X-ray (CXR) image. In the paper, we'll describe in detail the dataset preparation, training process, and the results obtained. The main concept is illustrated in fig. 1.

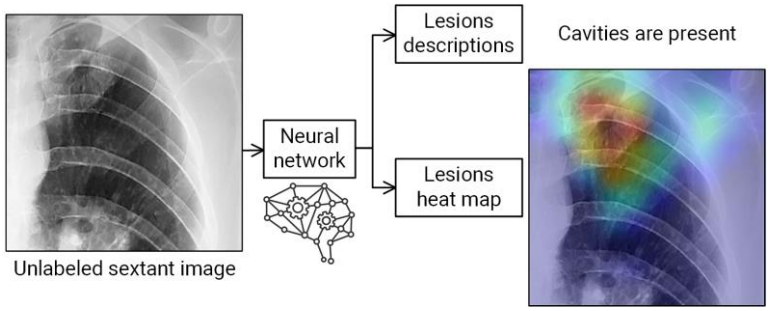


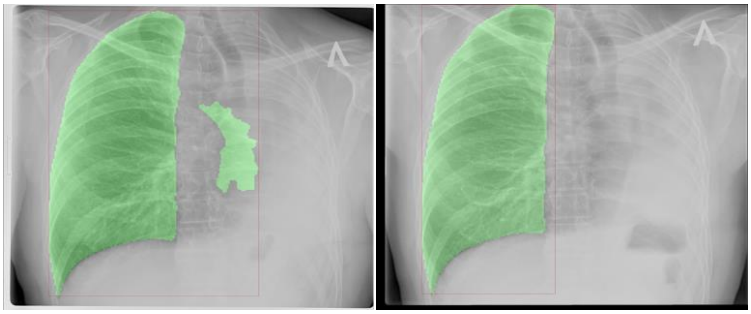
Fig. 1. Scheme of AI-CDSS to help a specialist

As shown in fig. 1, an unlabeled lobe (lung sextant) of the CXR image is input to the experimental software. The experimental AI-CDSS results in lesion descriptions and corresponding heat maps with spatial localization of these lesions.

An overview of existing solutions was conducted to identify relevant methodologies and approaches. In paper [1], the authors present a detailed literature review of existing solutions for detecting tuberculosis on chest X-rays. Existing publications present several conceptually similar approaches to the problem under consideration [2–4]. The authors [2] achieved a high classification accuracy of 96% in distinguishing tuberculosis from non-tuberculosis cases. However, their study exhibits several notable differences. First, the authors utilized images characterized by high contrast. Second, their approach did not include differential diagnosis of pathologies – specifically, the classification of lesion types. Activation maps have been derived during the training of CNN-based classifiers [3]. The authors developed an ensemble of neural networks for the classification of tuberculosis, achieving an overall accuracy of 82.93%. Notably, the model demonstrated higher performance – reaching 90% accuracy – in identifying specific pathological patterns such as effusion, infiltration, mass, and nodule (when dividing the data set into training and test samples in a ratio of 8:2). However, the accuracy was lower for other pathologies, such as fibrosis. A distinctive feature of our work is the classification of sextants, which was employed to enhance the accuracy of pathological area localization in the generation of activation maps.

We used image analysis of lung lobes (sextants) cropped by the lung mask rather than the whole CXR image. Image analysis of whole CXR images (cropped by the lung mask too) gave lower accuracy. Before being sent to the neural network the image is compressed to 299×299 or

224×224 pixels. This image compression makes small lesion foci invisible or faintly visible. Thus, processing lung lobes (sextants) instead of the whole lung preserves small lesion foci for processing by the neural network.



a) *b)*

Fig. 2. Example images with a missing lung:
a) image with an incorrectly generated one lung mask;
b) image (patient ID 248) with a correctly generated one lung mask

a) *b)*

Fig. 3. Example images for which no masks were generated:
a) poor quality image (patient ID 2841);
b) an image with too large black background area (patient ID 16325)

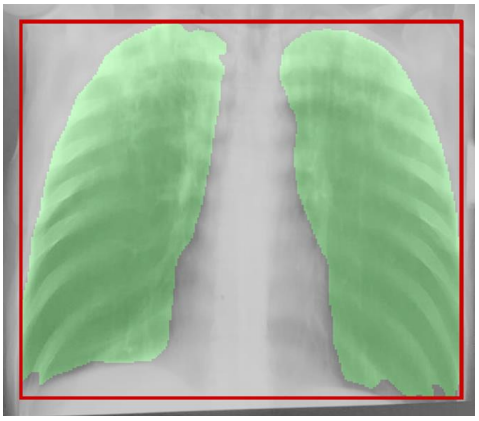


Fig. 4. Example of generated image for manual mask review

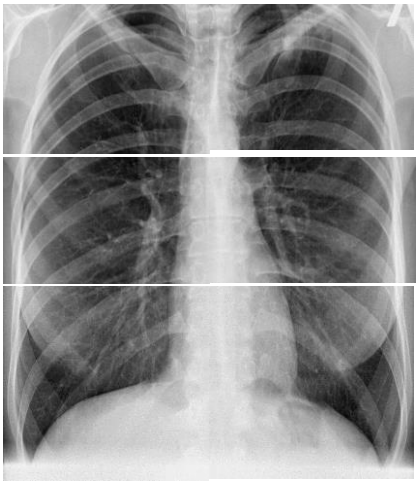


Fig. 5. Example of dividing lungs into six parts (sextants) by mask area

Experimental AI-CDSS to assist radiologists is based on a neural network and artificial intelligence techniques. The output of experimental software is lesion descriptions (for example, "cavities are present") and lesions heat map that highlights these lesions on CXR. AI-CDSS automatically produces results for three lesion types: cavities, infiltrates, and nodules.

To get lesion descriptors we used a binary classification method. Binary classification [1] is classifying the elements of a set into one of two groups (each called a class). For example, to determine if CXR has cavities or not.

To get heat maps we used the Gradient-weighted Class Activation Mapping (Grad-CAM) method [5]. Grad-CAM uses the gradients, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image for predicting the classes. That is, the Grad-CAM method allows one to visualize the areas of high attention of the neural network, based on which it makes the final decision.

To retrieve a dataset of sextants from the original CXR database of the TB Portals Case Browser [6] (URL: https://data.tbportals.niaid.nih.gov) and TB DEPOT [7] (Data Exploration Portal, URL: https://depot.tbportals.niaid.nih.gov) a lot of preliminary work that has to be completed:

1. manually exclude inappropriate images with lateral patient orientation, containing large white/black frames, with incorrect modality, containing other body parts, etc.;
2. automatically invert the image in color if it is inverted;
3. automatically flip the image if it is flipped horizontally (the heart is on the right side);
4. develop an algorithm for automatically cutting lungs into sextants;
5. conduct a quantitative analysis of the obtained data.

Steps 1–3 were described in the research paper [8]. Steps 4 and 5 are explained in this paper.

Algorithm for cutting lung into sextants. To work with lung sextants, it is necessary to correctly automatically slice lung images into parts. During a meeting with the radiologist, it was determined that the sextant separation when annotating images is interrelated with anatomical aspects of the position of veins, arteries, etc. In more detail, the division of the lung into lobes is made by a radiologist according to the following rules.

For CXR scans:

– upper sextants are above the lower edge of the aortic arch;

– middle sextants are between the lower edge of the aortic arch and the right inferior pulmonary vein;

– lower sextants are below the right inferior pulmonary vein.

For CT scans:

– upper sextants are above keel level;

– middle sextants are between the keel and right inferior pulmonary vein;

– lower sextants are below the right inferior pulmonary vein.

In this formulation, dividing the lungs into lobes based on anatomical image maps *is a challenging task in itself*, as it cannot be adequately captured by a simple mathematical algorithm due to its reliance on complex anatomical structures. In addition, there is always the possibility of human error in annotation, since some lesions in the lungs are on the border of the sextants selected by the radiologist.

According to the CXR images processing pipeline proposed in the article [8], lung masks were obtained automatically using the application program interface (API) of the *LungExpert* [9, 10]. This software is available online for testing at URL: https://lungs.org.by.

When cropping the lungs according to their automatically generated mask, several difficulties arose. The lung mask is incorrectly positioned under the following conditions:

– lung collapse is present (fig. 2, *a*);

– poor quality of the original image (fig. 3, *a*);

– too large black background on the image (fig. 3, *b*);

– presence of foreign objects on the image.

To identify images with incorrectly generated lung masks, they were viewed in two ways: automatically, using algorithmic processing of lung mask regions, and by manual review. This semi-automatic approach allowed minimizing the number of incorrectly generated masks.

For manual review of the images, previews were created that included a lung mask overlaid on the CXR and a rectangular cropping frame where the image should be cropped (fig. 4).

To organize semi-automatic search and correction of images with *incorrectly generated masks*, the ratio of two lung mask areas was used. To calculate the ratio of two lung mask areas, the number of lung mask contours is counted. Ideally, there should be two contours for the left and right lung. The first two contours with the largest areas $S1$ and $S2$ are then found. If there are third, fourth, etc. contours, *they are ignored*. The ratio of lung areas $S2/S1$ is in the range [0, 1], where $S1$ is the area of the largest contour and $S2$ is the area of the second largest contour.

To determine the $S2/S1$ ratio value for the correctly generated mask, this ratio was calculated for each of the images in the original dataset. A total of 10,475 masks and corresponding ratios were obtained (tabl. 1).

Table 1
Number of images with lung area ratio in the intervals

Column No.	1	2	3	4	5	6	7	8	9	10	11
Lung area ratio interval	0.0	(0.0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1.0]
Number of images	514	***35***	***54***	***132***	***179***	***363***	***671***	1,378	2,519	3,223	1,407

Of this tabl. 1, 514 images of column 1 had only one lung mask contour ($S2/S1=0$). Single-lung masks were considered "correct" and processed according to a separate procedure described below.

The images with different ratios were then partially manually reviewed. After manual review, a ratio of 0.6 was *experimentally* chosen as the boundary for "correct" and "incorrect" lung masks.

If the ratio is strictly greater than 0.6 ($S2/S1>0.6$), we consider the mask to be "correct" (see columns 8–11 in tabl. 1). In total, 230 images from columns 8–11 were manually reviewed and subsequently excluded from the final dataset due to errors in lung mask definition.

If the ratio is less than or equal to 0.6 ($S2/S1\leq0.6$) and not equal to zero ($S2/S1\neq0$), we consider the mask to be "incorrect" (see columns 2–7 in tabl. 1). We excluded the "incorrect" masks from further study and did not cut them into sextants. This is a simplified approach that could probably be improved.

Before cropping the 'correct' masks, a padding of 3% is added to both the width and height of the bounding frame, but not exceeding the actual image dimensions.

Having selected a rectangular frame to cut, the next step is to divide that frame into six parts (sextants). After a meeting with the radiologist, it was decided to divide the lungs into sextants *evenly by area of mask*. Thus, the cropping frame is first divided vertically into three equal parts according to the mask area (two horizontal lines in fig. 5), so the area of each part should be one-third of the area of the whole lung mask. The cropping frame is then divided exactly in half horizontally (one vertical line in fig. 5). An approach involving overlapping sextants could potentially provide better performance. The investigation of overlapping sextant configurations is intended for future research.

To correctly process masks with one lung (514 images of column 1 in tabl. 1), a second missing lung must be drawn on the image. Fig. 6 shows how the cropping frame is formed around the mask for the case where only one lung (contour) was detected on the CXR image.

In fig. 6 value d is the estimated distance between existing and missing lungs and W_1 is the width of one lung mask. To find the estimated value of d, we obtained all masks with (at least) two contours and found the ratio k of the distance between the lungs (the two largest contours) to the sum of the widths of the two lungs using eq. (1):

$$k = \frac{d}{W_1+W_2}, \tag{1}$$

where d is the distance between the two lungs; W_1 is the width of the right lung; W_2 is the width of the left lung. Fig. 7 explains the variables in eq. (1).

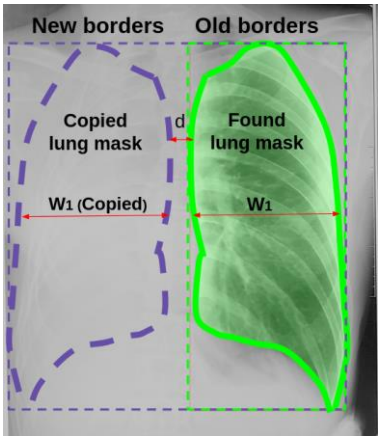


Fig. 6. An image explaining the procedure for selecting the cropping frame for a single lung

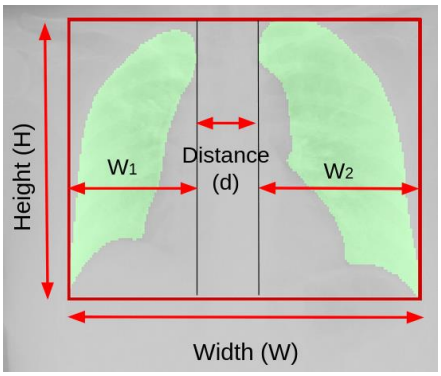


Fig. 7. Image explaining the notations in the eq. (1)

Tabl. 2 shows the results of the calculated k ratio in eq. (1) among the original CXR dataset.

Table 2
Number of images with value k in the intervals

Number of images	10,118	30	9	4	1
Interval of k values	(0, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]

According to tabl. 2, most of the images that have at least two contours on the mask have a k value between 0 and 0.3 (over 99.5% of all images). The average value of the k over the whole dataset is 0.079. Thus, it was decided to use a k value equal to 0.3 to construct a cropping frame with a missing lung. This choice is due to the intention to preserve more information in the image when the cropping frame is automatically selected.

If $k = 0.3$ and according to eq. (1) the estimated value of d for one-lung mask is equal to:

$$d = k \cdot (W_1 + W_2) = k \cdot 2 \cdot W_1 = 0.6 \cdot W_1. \quad (2)$$

Therefore, if the number of lung contours is equal to one, a new cropping frame width is calculated automatically by mirroring one lung to the other half of the thorax with the missing lung. The eq. (3) for calculating the width W of the cropping frame is:

$$W = 2 \cdot W_1 + d = 2.6 \cdot W_1. \quad (3)$$

Quantitative analysis of the obtained data. An important step of this study was to collect statistical information about the existing markup of lung images based on radiologists' annotations (feature labels). All feature labels are divided into five groups:

1. general features for describing the lungs as a whole (5 features);
2. features for describing cavities in the lungs (5 features);
3. features for describing nodules in the lungs (11 features);
4. features for describing infiltrate (density anomalies) in the lungs (3 features);
5. feature to describe the presence and magnitude of pulmonary collapse (1 feature).

In total 25 features were analyzed. Unfortunately, there is no possibility to show all tables and charts in this paper. Therefore, here is an example of one labeling feature "Small Nodules" *in the original CXR dataset*.

The "Small Nodules" feature is the integer number from 0 to 100%. The frequency and distribution of non-empty values of the "Small Nodules" feature are shown in tabl. 3, 4, and fig. 8, 9.

Table 3
Number of lobes with non-empty values for the "Small Nodules" feature

	Upper Right lobe	Upper Left lobe
> 0	4,120	3,616
= 0	3,418	2,965
	Middle Right lobe	Middle Left lobe
> 0	2,819	2,828
= 0	2,402	2,424
	Lower Right lobe	Lower Left lobe
> 0	1,529	1,325
= 0	1,551	1,563

Table 4
Number of lobes with non-zero values in the intervals for the "Small Nodules" feature

Value range, %	(0, 10]	(10, 20]	(20, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 70]	(70, 80]	(80, 90]	(90, 100]
Number of lobes	8,272	4,110	1,325	1,123	651	276	189	168	50	73

Fig. 8. Percentage distribution of lobes with non-zero values for the "Small Nodules" feature

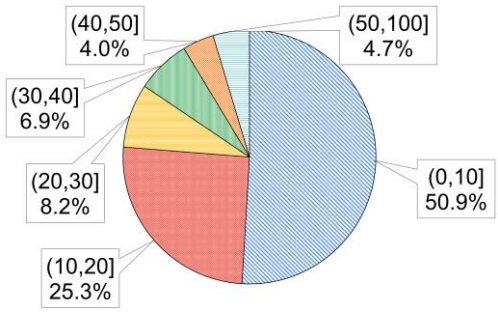


Fig. 9. Frequency of occurrence of lobes with non-zero values for the "Small Nodules" feature

Based on the quantitative analysis of the data obtained, the following conclusions can be drawn.

A strong imbalance in the annotated data was found when working with different groups of feature labels. This strong imbalance exists for both the distribution of binary responses ("Yes" or "No") and for integer numerical values from 0 to 100%.

For binary responses ("Yes" or "No"), most data have a value of "No".

For integer numerical values from 0 to 100%, most non-zero data is in the range [1, 10]% and zero values are the most common.

Upper sextants are affected more frequently than lower sextants. The right lung is affected generally more frequently than the left lung. The exception is only the "Collapse" feature, for which lower sextants are affected more frequently than upper sextants, and the right-left distribution is approximately the same.

The lack of labeled data can be partially addressed by *merging data for different sextants into one group*. Thus, circumventing the lack of data for training. To increase the training dataset size, we merged data with the same labels for the upper right/left, middle right/left, and lower right/left sextants into one group.

As a result of data merging, the task of recognizing lesions in lung lobes is reduced from regression (predict a number) to binary (a class exists or does not exist in the lung). Thus, in tabl. 5, the labeling features are combined from the 20 tasks into three binary tasks.

Table 5
Three binary classification tasks for cavities, nodules, and infiltrates

Feature name [6]	Value range	Task
1. Small Cavities (less than 3 cm)	[0.0, 100.0]%	Reduce to the binary classification task.
2. Medium Cavities (3–5 cm)	[0.0, 100.0]%	
3. Large Cavities (more than 5 cm)	[0.0, 100.0]%	**Cavities** exist or don't exist in a sextant
4. Does any Large cavity belong to a multi-sextant cavity?	Yes / No	
5. Can Multiple cavities be seen?	Yes / No	
6. Small Nodules (less than 3 mm)	[0.0, 100.0]%	Reduce to the binary classification task.
7. Medium Nodules (5–15 mm)	[0.0, 100.0]%	
8. Large Nodules (15–30 mm)	[0.0, 100.0]%	**Nodules** exist or don't exist in a sextant
9. Huge Nodules (more than 30 mm, tuberculoma)	[0.0, 100.0]%	
10. Any calcified or partially-calcified Nodules?	Yes / No	
11. Any non-calcified Nodules?	Yes / No	
12. Any clustered Nodules (nodules 2–5 mm apart)?	Yes / No	
13. Can Multiple Nodules be seen?	Yes / No	
14. Low/ground glass Density (active fresh nodules)	[0.0, 100.0]%	
15. Medium Density (stabilized fibrotic nodules)	[0.0, 100.0]%	
16. High Density (calcified nodules, typically sequella)	[0.0, 100.0]%	
17. Low/ground glass Density	[0.0, 100.0]%	Reduce to the binary classification task.
18. Medium Density	[0.0, 100.0]%	
19. High Density	[0.0, 100.0]%	**Infiltrates** exist or don't exist in a sextant
20. Collapse	[0.0, 100.0]%	**Not considered**

Collapse is not considered, because it is easily detected visually even by a non-specialist.

Five general features for describing the lungs as a whole are not considered in this research, because we only work with lobes here.

Creating training, validation, and testing datasets. If different lung lobes from the same patient are both in the training and test datasets, this will lead to overtraining of the neural network and overestimation of classification accuracy. To exclude different fragments of the same patient in the training and test datasets all CXR images were divided into three groups according to patient ID: training (70% of all patient IDs), validation (15%), and testing (15%). That is, each ID should belong only to the training, validation, or testing dataset.

Dataset directories for training, validation, and testing were created using the algorithm in fig. 10.

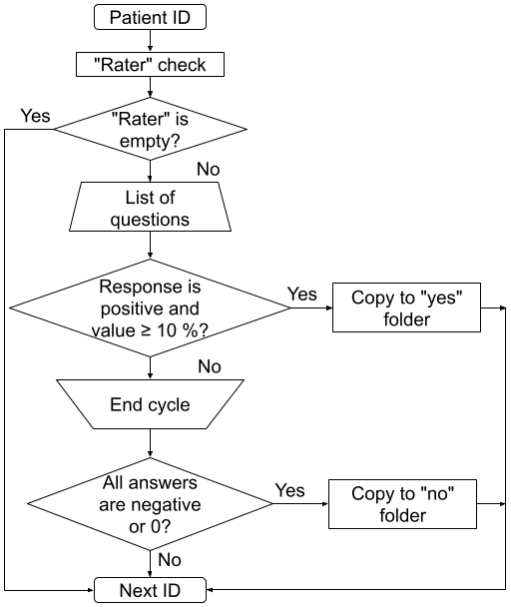


Fig. 10. Algorithm for creating datasets

The algorithm presented in fig. 10 includes two main checks.

1. Verification that an expert has annotated the lungs. By checking the "Rater" field in "Overall Characteristics" is not empty. If the "Rater" field is empty, the CXR image is not annotated and is excluded from datasets.

2. Division of lung sextants into classes according to various conditions. For example, separation into a "cavity" class in the "yes" directory and a "not cavity" class in the "no" directory.

Conducting experiments. During the initial experiments, the following neural network architectures have been trained and tested: Xception, InceptionResNetV2, EfficientNet (B0, B1, B2, B3, B5), DenseNet (121, 169, 201), MobileNet (V2, V3Small). Various hypotheses were tested for the data and hyperparameters of the neural networks.

In this study, the following hyperparameters were used during training: the models were trained for a total of 150 epochs with a batch size of 5, optimization was performed using the Adam optimizer with an initial learning rate of 0.00001, and the learning rate was scheduled to decay exponentially with a decay rate of 0.96 every 10,000 steps. Additionally, early stopping with a patience of 15 epochs was applied to prevent overfitting and ensure efficient convergence.

As an example, tabl. 6 shows the classification accuracy for the nodule lesions with the final terms for the data described below.

Table 6
Classification accuracy for the nodule lesions on the test dataset

Neural network architecture	Classification accuracy for the *nodule* lesion on the test dataset, %	Number of trainable parameters, mln	Number of epochs to train
Xception	72.7 / 72.8 / 72.4 / 71.8 / 72.1 / 72.7	20.8	17 / 16 / 20 / 19 / 17 / 19
InceptionResNetV2	72.4 / 72.3	54.3	20 / 28
DenseNet201	72.8 / 72.9	18.1	50 / 39
DenseNet169	73.2 / 73.6 / 73.4	12.5	54 / 41 / 55
EfficientNetB5	62.7 / 65.9 / 66.2	28.3	43 / 48 / 65
EfficientNetB3	66.2	10.7	49
EfficientNetB0	69.5	4.0	121
MobileNetV2	71.1 / 70.1	2.2	121 / 131
MobileNetV3Small	63.4	1.5	16

Some architectures have multiple training sessions separated by a "/" sign.

For tabl. 6 the datasets for training, validation, and testing are the same, as well as all the hyperparameters of the neural network. The difference is in the architecture, in the preliminary shuffle of the data, and the random augmentation.

Both lightweight and large architectures to test the accuracy and heat maps were used. The accuracy difference between *training sessions of the same architecture* is approximately ±0.5%, i.e., for Xception 72.8–71.8=1.0 and 1.0/2 = ±0.5.

According to tabl. 6, there is an accuracy difference between architectures from 62.7 to 73.6% or around 10.9% which is a large percentage for binary classification. Generally, large architectures with more trainable parameters give higher accuracy, but this is not always the case. For example, MobileNetV2 either very fast and has an accuracy that is comparable to architectures 10 times larger. MobileNetV2 is not the best in terms of accuracy, but when speed is also important, this neural network is indispensable.

Oddly enough, EfficientNetB5 with 28.3 million trainable parameters performs much worse (66.2% for the third training session) than its counterpart EfficientNetB0 with 4.0 million parameters (69.5%). *No signs of overfitting were observed for EfficientNetB5*. Training took 65 epochs and *validation* accuracy did not exceed 66.6%, which corresponds to minimal validation loss in the early stopping callback checkpoint with the patience of 15 epochs.

As a result of numerous workouts with different architectures and hyperparameters of these architectures, we decided to stop at Xception architecture [11]. Although DenseNet169 performed slightly better. *In this study, we focus not on architecture but on the data itself and the reasons how to raise the accuracy above 75% in future research.*

The quality of the resulting heat maps is important because visualization of the lesions is also the goal of the study. Not all neural networks produce consistent results that can be interpreted by medical professionals. For example, EfficientNet models showed useless heat maps because of the nature of their architecture. DenseNet models also showed poorly interpretable heat maps. Later, the following sections present radiologist-verified heat maps for each identified type of pathology.

We would like to note that similar results were obtained by another team when solving the problem of COVID-19 search on X-ray images [12]. According to their results, Xception architecture is also best for classification tasks.

In general, the following conversions and improvements to make datasets were made:

– randomly divide patient IDs into three categories: training (70% of all patient IDs), validation (15%), and testing (15%), so different sextants of the same patient cannot be in both the training and test datasets;

– CXR image evaluator (rater) must be non-empty;

– remove from the dataset all annotations whose value is less than or equal to 10% (≤10%);

– apply various augmentations to the images;

– normalize images to the range [–1, 1];

– equalize the datasets so that the number of "yes" and "no" cases are equal;

– convert from float32 to float16 to make dataset smaller;

– apply checkpoints with early stopping to minimize loss value;

– etc.

Two main improvements significantly enhanced classification accuracy: data augmentation and filtering out *data values less than or equal to 10% (≤10%).* Other modifications had minimal impact on performance.

The decision to remove values ≤10% was based on observed improvements in classification accuracy. In our dataset, the values represent numeric annotations assigned by the radiologist to different regions of the image. However, many of these values fell within the low range (0, 10]% and appeared to contribute little meaningful information.

For example, as shown in fig. 9, the "Small Nodules" label included 50.9% of non-zero values within the (0, 10]% range. Despite their high frequency, removing these low values improved accuracy by reducing noise and highlighting more relevant areas.

To evaluate this effect, we conducted several experiments using the Xception architecture:

1. accuracy on test data with values >10% = 72.8%;
2. accuracy on full-range test data (0–100]% = 70.9%;
3. accuracy on test data with values >20% = 70.4%.

The difference between the first and second results is 1.9%, indicating that excluding low values has a measurable positive impact. While the exact reason for this improvement is not fully understood, we suspect that values ≤10% may introduce noise or inconsistencies into the model's predictions. Further investigation – ideally involving domain experts such as radiologists – would help clarify this.

Further removal of values less than or equal to 20% reduces classification accuracy by about 2%. The test dataset in the second check also contains data >10%. In future research, it will be interesting to train on >10% data but evaluate on (0, 100]% data.

Properly done augmentation almost always enhances results. In this study, augmentation improved accuracy by about 3%. Augmentation consists of:

– random horizontal and vertical flips;

– three random rotations on 90, 180, and 270 degrees;

– random brightness change within ±5%;

– small random contrast change.

In addition, nodule lesions were divided into "small-medium" and "large-huge" subclasses (see rows 6–9 in tabl. 5) to test the hypothesis that on divided subgroups the neural network binary classification will be better (fig. 11).

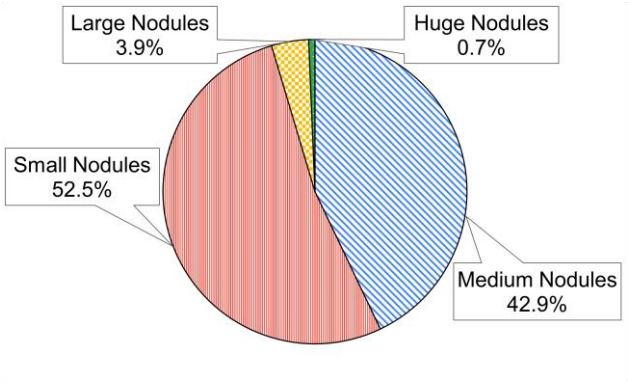


Fig. 11. Percentage of nodules sizes

However, this did not give a significant improvement:

– for "small-medium" nodules size (5,955 images in total) Xception accuracy is 76.59%;

– "large-huge" nodules size (287 images in total) Xception accuracy is 64.37%.

In comparison for all nodule sizes in one class, the accuracy is 72.83%. Here "large-huge" has lower accuracy than "small-medium". *This is counterintuitive*, because "large-huge" nodules should be more visible. We assume there is not enough data for "large-huge" nodules to train a neural network.

The classification accuracy achieved on the test datasets for cavities, infiltrates, and nodules is presented in the respective sections below. Importantly, none of the test samples were used during the training of the neural network. A detailed performance evaluation across individual classes is provided in the confusion matrices shown in tabl. 8, 10, 12.

Cavities detection. The dataset size for cavities detection to train neural networks is shown in tabl. 7.

Table 7
Dataset size for cavities detection to train neural networks

Class name	Training	Validation	Test
"Yes" or cavities are present	2,476	387	553
"No" or cavities are absent	2,476	387	553
Total	4,952	774	1,106

Five feature labels were used to create the dataset (see rows 1–5 in tabl. 5). At least one non-zero value or affirmative ("Yes") feature label is sufficient to assign the lung lobe image to the "Yes" class with some cavities in the image (see the algorithm in fig. 10). For the second "No" class or without cavities in the image all five features are required to have zero or negative answers. The same approach is applied to infiltrates and nodules but with different feature labels.

Classification accuracy for the test dataset is 73.1%.

Table 8
Confusion matrix for cavities binary classification

		Actual	
		Positive (with cavities)	Negative (without cavities)
Predicted	Positive	TP=408	FP=153
	Negative	FN=145	TN=400

In comparison, random guessing has an accuracy of 50%. Thus, the obtained accuracy of 73.1% is good. Incorrect neural network predictions: false negative (FN=145) and false positive (FP=153) are roughly comparable. In medicine, FN is more significant than FP and should be smaller (FN < FP), because missing an illness can cost a life. However, we have not explored this issue in depth.

Fig. 12 shows heat maps for the true positive (TP) responses automatically obtained by the neural network on the test dataset.

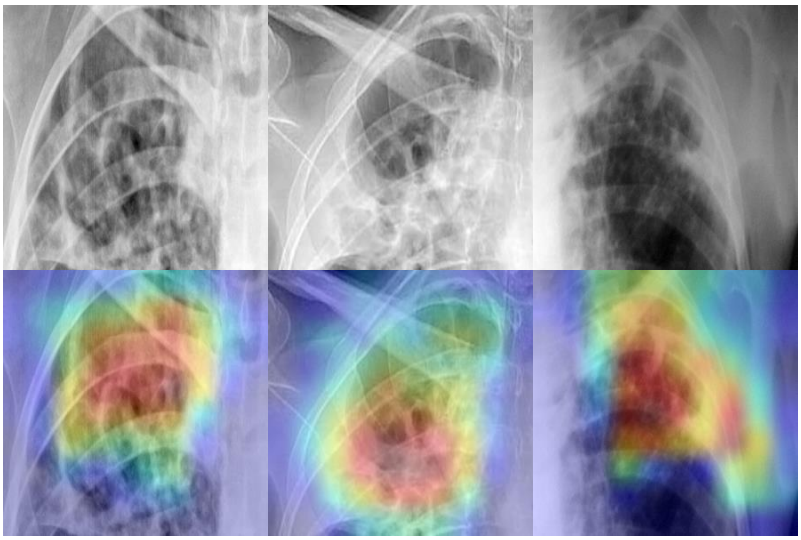


Fig. 12. Examples of TP outcomes for cavities and corresponding heat maps

The correctness of the heat maps in fig. 12 *was agreed with the radiologist*.

Infiltrates (anomalous density) detection. The dataset size for infiltrates detection to train neural networks is shown in tabl. 9.

Table 9
Dataset size for infiltrates (anomalous density) detection to train neural networks

Class name	Training	Validation	Test
"Yes" or infiltrates are present	4,935	1,120	977
"No" or infiltrates are absent	4,935	1,120	977
Total	9,870	2,240	1,954

Three feature labels were used to create the dataset (see rows 17–19 in tabl. 5).

Classification accuracy for the test dataset is 71.9%.

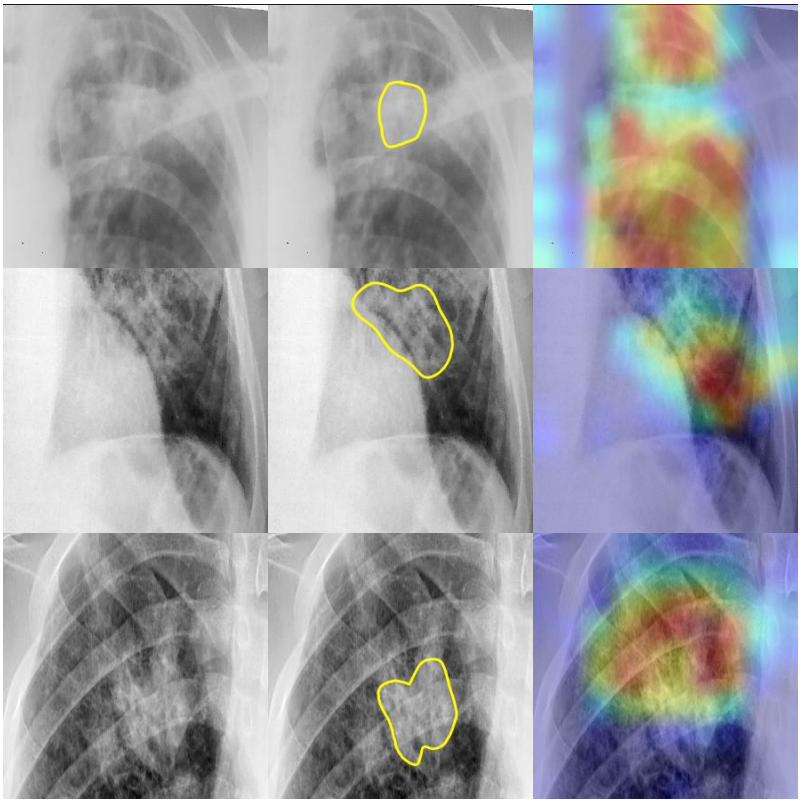
Table 10
Confusion matrix for infiltrates binary classification

		Actual	
		Positive (with infiltrates)	Negative (without infiltrates)
Predicted	Positive	TP=730	FP=302
	Negative	FN=247	TN=675

Fig. 13 shows heat maps for the TP responses automatically obtained by the neural network on the test dataset.

The radiologist's expertise revealed that the neural network *correctly identified* the presence of the infiltrates in fig. 13, but their location on the heat map is not where the red color is located.

The *approximate* location of the infiltrate is highlighted by the *non-radiologist* in fig. 13, *b*.



a) *b)* *c)*

Fig. 13. Examples of TP outcomes for infiltrates: *a)* unlabeled input images; *b)* approximate ground truth; *c)* corresponding heat map

If compare the heat map and the approximate location of the infiltrate, we can conclude that the *neural network pays attention not to the infiltrate itself (the light spot), but to the gradient*, i.e., the transition from less dense to more dense areas at the infiltrate boundary. The areas from lower to higher density (gradients) are highlighted by the neural network, but not the infiltrate itself.

Nodules detection. The dataset size for nodules detection to train neural networks is shown in tabl. 11.

Table 11
Dataset size for nodules detection to train neural networks

Class name	Training	Validation	Test
"Yes" or nodules are present	8,834	2,006	1,796
"No" or nodules are absent	8,834	2,006	1,796
Total	17,668	2,012	3,592

Eleven feature labels were used to create the dataset (see rows 6–16 in tabl. 5).

Classification accuracy for the test dataset is 72.8%.

Table 12
Confusion matrix for nodules binary classification

		Actual	
		Positive (with nodules)	Negative (without nodules)
Predicted	Positive	TP=1,218	FP=402
	Negative	FN=578	TN=1,394

Fig. 14 shows heat maps for the TP responses automatically obtained by the neural network on the test dataset.

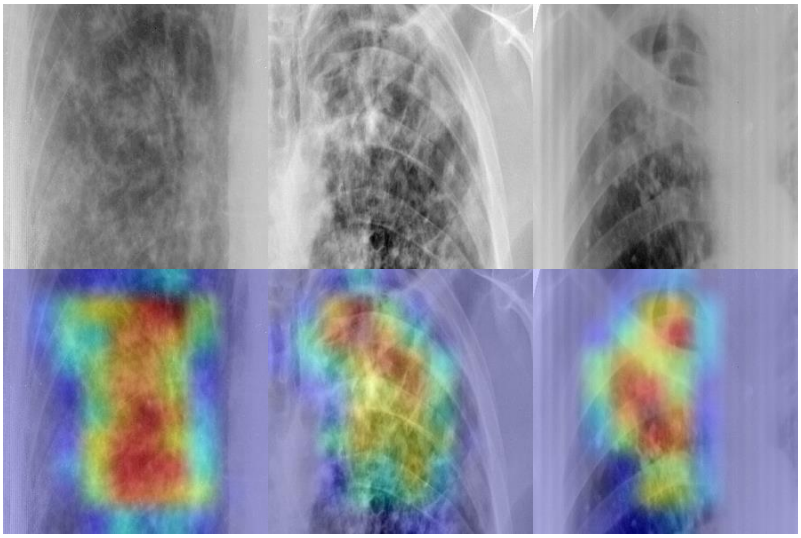


Fig. 14. Examples of TP outcomes for nodules and corresponding heat maps

The correctness of the heat maps in fig. 14 *was agreed with the radiologist*.

Discussion of results. The average classification accuracy of the Xception model for three lesion types: cavities (73.1%), infiltrates (71.9%), and nodules (72.8%), is equal to 72.6%.

Three improvements made the greatest contribution to accuracy:

– filtering of data by numerical values greater than 10% (>10%);

– excluding images with empty "Rater" field in "Overall Characteristics";

– image augmentation.

Overall, classification accuracy improved by an average of 5%.

It was found that *classification errors in all three lesion types are primarily related to the quality of the original dataset*. That's why it's hard to improve accuracy over 75%.

It was observed that some unlabeled sextants, in which the lung lobe should be free of lesions, were classified as pathological by the neural network. For example, *according to the radiologist*, sextants in fig. 15 and fig. 16 contain lesions, and the neural network shows they are, but they didn't contain any markup in the dataset and are considered as healthy.

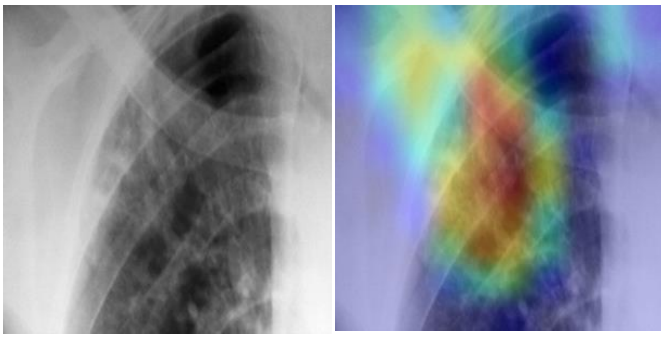


Fig. 15. There are nodules in the image, but they are not present in the annotated data

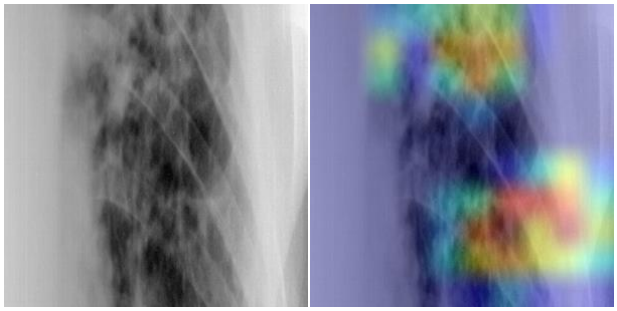


Fig. 16. There are cavities in the image, but they are not present in the annotated data

Another cause of classification errors is poor image quality in some cases. For example, in fig. 17, the misclassification result is affected by the low image contrast and the presence of artifacts in the background.

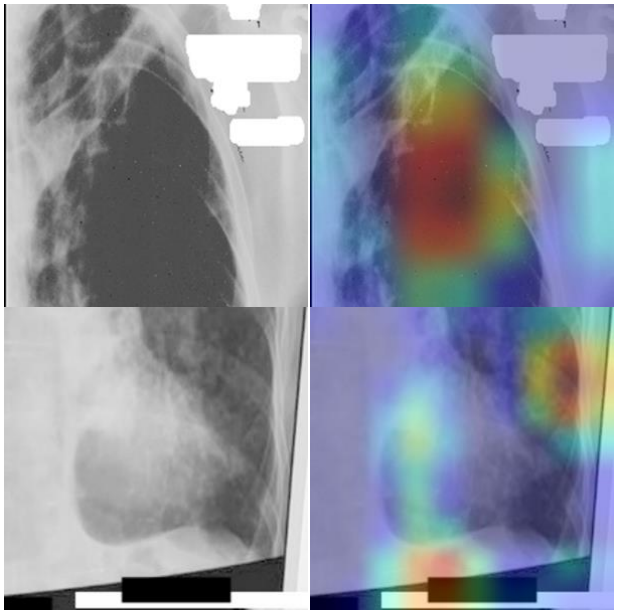


Fig. 17. Examples of low-quality images with FP outcome

Heat maps from the Xception model with TP outcomes for cavities and nodules are mostly understandable to radiologists.

Heat maps for infiltrates (see fig. 13) are not understandable to radiologists because the area of attention of the neural network is not directed to the area of pathology itself. Instead, the neural network pays attention to gradients or the area of transition from healthy tissue to pathology around or near the infiltrate. This may be because an infiltrate is an area of the lung with a higher density and appears as a lighter area on the CXR image. For this reason, infiltrate can be seen by analyzing changes in the brightness of the image (gradients), which is what the neural network does.

To get heat maps for infiltrate (anomalous density) that are understandable to experts, a separate study should be done. The most likely approach is to make preliminary gradient transformations of the CXR images to produce heat maps with the correct location of red regions.

From the perspective of radiologists, the classification of infiltrates represents a particularly challenging diagnostic task. We also had difficulties visualizing the infiltrate using heat maps. It was unexpected to get the result that the neural network pays attention during classification not to the lesion itself, but to the gradient area around it.

It is our contention that the following additional studies will serve to enhance the accuracy of the classification and the quality of the heat maps.

1. Using only digital images (cell phone photos must be excluded) should greatly improve the results.

2. Advanced data selection depending on various conditions would presumably improve the results.

3. Prepare a dataset based on descriptors labeled (annotated) by a single expert. And then fine-tune the trained model on this dataset. This method can improve the classification accuracy since the methodology of labeling images by one expert is the same.

4. Our team from Belarus annotated many three-dimensional CT images. CT is not CXR (different modality). Nevertheless, it is theoretically possible to obtain additional valuable descriptors that describe the position of the studied lesions (cavities, nodules, and infiltrate). For example, convert CT to CXR modality and apply the available CT annotations to increase the spatial accuracy predictions of the neural network for CXR scans.

Conclusions. For the Xception model, the binary classification accuracy on the test dataset is 73.1% for cavities, 71.9% for infiltrates, and 72.8% for nodules.

Heat maps associated with pathological processes in the lungs were constructed. Heat maps with true positive outcomes for cavities and nodules are mostly understandable to experts. To get heat maps for infiltrates that are understandable to experts, more research is needed.

Obtained results are good, but not excellent. Further investigation should be done to improve the classification accuracy and quality of the heat maps.

Authors' contributions. *D. A. Paulenka* carried out the quantitative analysis of the obtained data and conducted experimental studies, including training and testing models for nodule lesions. *A. A. Kosareva* proposed an algorithm for dividing the lung into sextants, created training, validation, and testing datasets of sextants from the original CXR database of the TB Portals Case Browser, and conducted experimental studies, including training and testing models for cavity and infiltrate lesions. *E. V. Snezhko* provided scientific supervision for the overall research process and conducted a literature review on the topic. *V. A. Kovalev* participated in discussions, summarization, and presentation of the obtained results.

References

1. Hansun S., Argha A., Liaw S.-T., Celler B. G., Marks G. B. Machine and deep learning for tuberculosis detection on chest X-rays: systematic literature review. *Journal of Medical Internet Research*, 3 July 2023, vol. 25, p. e43154. DOI: 10.2196/43154.

2. Acharya V., Dhiman G., Prakasha K., Bahadur P., Choraria A., …, Kautish S. AI-assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model. *Computational Intelligence and Neuroscience*, 3 October 2022, vol. 2022, p. 2399428. DOI: 10.1155/2022/2399428.

3. Guo R., Passi K., Jain C. K. Tuberculosis diagnostics and localization in chest X-rays via deep learning models. *Frontiers in Artificial Intelligence*, 5 October 2020, vol. 3, p. 583427. DOI: 10.3389/frai.2020.583427.

4. Thomsen K., Christensen A. L., Iversen L., Lomholt H. B., Winther O. Deep learning for diagnostic binary classification of multiple-lesion skin diseases. *Frontiers in Medicine*, 2020, vol. 7, p. 574329. DOI: 10.3389/fmed.2020.574329.

5. Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, October 2019, vol. 128, no. 2, pp. 336–359. DOI: 10.1007/s11263-019-01228-7.

6. Rosenthal A., Gabrielian A., Engle E., Hurt D. E., Alexandru S., …, Tartakovsky M. The TB Portals: an open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. *Journal of Clinical Microbiology*, 2017, vol. 55, no. 11, pp. 3267–3282. DOI: 10.1128/jcm.01013-17.

7. Gabrielian A., Engle E., Harris M., Wollenberg K., Juarez-Espinosa O., ..., Tartakovsky M. TB DEPOT (Data Exploration Portal): A multi-domain tuberculosis data analysis resource. *PLOS ONE*, May 2019, vol. 14, no. 5, pp. 1–23. DOI: 10.1371/journal.pone.0217410.

8. Kosareva A., Paulenka D., Snezhko E. Chest X-ray image processing based on radiologists' textual annotations. *Open Semantic Technologies for Intelligent Systems (OSTIS): Conference Proceedings, Minsk,*

Belarus, 18–20 April 2024. Editorial board: V. V. Golenkov [et al.]. Belarusian State University of Informatics and Radioelectronics, vol. 8, pp. 293–302.

9. Snezhko E. V., Kovalev V. A., Kosareva A. A., Paulenka D. A. AI-based software for computer-assisted diagnosis of lung diseases using chest X-Ray and CT images. *The 1st Exhibition-forum of the IT-academgrad "Artificial Intelligence in Belarus": Conference Proceedings, Minsk, 13–14 October 2022*, pp. 80–87 (In Russ.).

10. Kovalev V. A., Tuzikov A. V., Snezhko E. V., Paulenka D. A. Software "AI-based software for computer-assisted diagnosis of lung diseases using chest X-Ray and CT images" (LungExpert). *Computer program registration certificate No. 1619 dated 02.08.2023* (In Russ.). Available at: http://search.ncip.by/depon/index.php?pref=1&lng=ru&page=3&target=1771 (accessed 15.09.2024)

11. Chollet F. Xception: deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*, pp. 1251–1258.

12. Nillmani, Sharma N., Saba L., Khanna N., Kalra M., …, Suri J. Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans. *Diagnostics*, September 2022, vol. 12, p. 2132. DOI: 10.3390/diagnostics12092132.

Information about the authors

Dzmitry A. Paulenka, Postgraduate Student, Lead Software Engineer, Laboratory of Biomedical Images Analysis, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: dmitri.pavlenko@gmail.com
https://orcid.org/0009-0007-9911-4356
https://www.researchgate.net/profile/Dzmitry-Paulenka
https://scholar.google.com/citations?user=2AX0it0AAAAJ

Aleksandra A. Kosareva, Junior Researcher, Laboratory of Biomedical Images Analysis, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: kosarevaaleksandra4317@gmail.com
https://www.researchgate.net/profile/Alexandra-Kosareva-3
https://www.scopus.com/authid/detail.uri?authorId=57934126700

Eduard V. Snezhko, Ph. D. (Eng.), Head of the Laboratory of Biomedical Images Analysis, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: eduard.snezhko@gmail.com
https://orcid.org/0000-0002-9843-0839
https://www.researchgate.net/profile/Eduard-Snezhko

Vassili A. Kovalev, Ph. D. (Eng.), Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: vassili.kovalev@gmail.com
https://orcid.org/0000-0002-8154-5875
https://www.researchgate.net/profile/Vassili-Kovalev-2
https://scholar.google.com/citations?user=-osN7dIAAAAJ

Информация об авторах

Павленко Дмитрий Анатольевич, аспирант (соискатель), ведущий инженер-программист, лаборатория анализа биомедицинских изображений, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: dmitri.pavlenko@gmail.com
https://orcid.org/0009-0007-9911-4356
https://www.researchgate.net/profile/Dzmitry-Paulenka
https://scholar.google.com/citations?user=2AX0it0AAAAJ

Косарева Александра Андреевна, младший научный сотрудник, лаборатория анализа биомедицинских изображений, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: kosarevaaleksandra4317@gmail.com
https://www.researchgate.net/profile/Alexandra-Kosareva-3
https://www.scopus.com/authid/detail.uri?authorId=57934126700

Снежко Эдуард Витальевич, кандидат технических наук, заведующий лабораторией анализа биомедицинских изображений, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: eduard.snezhko@gmail.com
https://orcid.org/0000-0002-9843-0839
https://www.researchgate.net/profile/Eduard-Snezhko

Ковалев Василий Алексеевич, кандидат технических наук, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: vassili.kovalev@gmail.com
https://orcid.org/0000-0002-8154-5875
https://www.researchgate.net/profile/Vassili-Kovalev-2
https://scholar.google.com/citations?user=-osN7dIAAAAJ