

# Examining the Validity of Input Lung CT Images Submitted to the AI-Based Computerized Diagnosis

Aleksandra A. Kosareva<sup>1,2\*</sup>, Dzmitry A. Paulenka<sup>2</sup>, Eduard V. Snezhko<sup>2</sup>, Ivan A. Bratchenko<sup>3</sup>, and Vassili A. Kovalev<sup>2</sup>

<sup>1</sup> Belarusian State University of Informatics and Radioelectronics, 6 P. Brovki str., Minsk 220013, Belarus

<sup>2</sup> The United Institute of Informatics Problems of NAS of Belarus (UIIP NASB), Biomedical Image Analysis Department, 6 Surhanava str., Minsk 220012, Belarus

<sup>3</sup> Samara University, Department of Laser and Biotechnical Systems, 34 Moskovskoe shosse, Samara 443086, Russia

\* e-mail: [kosareva@bsuir.by](mailto:kosareva@bsuir.by)

**Abstract.** A well-designed CAD tool should respond to input requests, user actions, and perform input checks. Thus, an important element of such a tool is the pre-processing of incoming data and screening out those data that cannot be processed by the application. In this paper, we consider non-trivial methods of chest computed tomography (CT) images verifications: modality and human chest checks. We review sources to develop training datasets, describe architectures of convolution neural networks (CNN), clarify pre-processing and augmentation processes of chest CT scans and show results of training. The developed application showed good results: 100% classification accuracy on the test dataset for modality check and 89% classification accuracy on the test dataset for checking of lungs presence. Analysis of wrong predictions showed that the model performs poorly on biopsy of lungs. In general, the developed input data validation model shows good results on the designed datasets for CT image modality check and for checking of lungs presence. © 2022 Journal of Biomedical Photonics & Engineering.

**Keywords:** image classification; medical imaging; convolutional neural network; deep learning; computer-aided diagnosis; computed tomography; input validation.

Paper #3500 received 23 Jun 2022; revised manuscript received 05 Sep 2022; accepted for publication 16 Sep 2022; published online 30 Sep 2022. doi: [10.18287/JBPE22.08.030307](https://doi.org/10.18287/JBPE22.08.030307).

## 1 Introduction

Thanks to the advent of Artificial Intelligence and Deep Learning methods a large number of CAD (Computer-Aided Diagnosis) tools are becoming available. One of the challenges of using such tools in real practice is incorrect input data, including incorrect modalities and other anatomical organs that are mistakenly fed into the system but not intended to be processed by the specific software. In this case, the diagnostic results will lead to incorrect results from the CAD system. For example, in one trial among 7830 Chest XRay images received from various hospitals in 10 countries, there were 564 cases, containing flawed data of such kind [1]. The input data may contain errors, for example, upside down images, different file format, different modality, wrong organs, etc. [2–4].

In this paper, we consider the validity check of input chest CT images. This important phase of any software for medical image processing and analysis is rarely considered in detail during scientific research. If some data is not valid for a specific application (for example, the shape, modality, content or orientation of the image does not match the business logic of the application), it should be dropped out and given a clear descriptive message. For computed tomography (CT) scans of human lungs, there are three main verification processes:

1. Trivial checks to verify that the input image is indeed a 3D scan with the required parameters in its header (check the actual format of an image file, check image volume rank, check that the image contains non-zero data, etc.);

2. Modality check to verify that the input image is CT scan and not ultrasound, PET, MRI or binary mask image;

3. Human chest check to test that input CT scan contains lungs in necessary proportion to be processed by the application.

In this article we review different datasets and describe CNN architectures to perform non-trivial methods of chest CT images verification. Currently these non-trivial checks can only be solved using neural networks. The developed methods show good results on the designed datasets. We obtained 100% classification accuracy on the test dataset for modality check and 89% classification accuracy on the test dataset for checking lung presence. Analysis of wrong predictions showed that the model for checking lung presence performs poorly on biopsy of lungs.

## 2 Materials and methods

### 2.1 Materials

**Dataset used for CT image modality check.** The training dataset (acronym DS1) is divided into two classes (CT and non-CT) with 113 images in each class (total 226 3D images). The 3D image data classes of CT and “Non-CT” (ultrasound, PET, MRI, binary masks) used in this study originated from fifteen datasets [5]:

1. COVID-19 CT segmentation (14 CT scans, 7 3D masks [6]).
2. Lung CT Segmentation Challenge 2017 (14 CT scans [7]).
3. Lung Nodule Analysis 2016 (14 CT scans [8]).
4. Private CT dataset from UIIP NASB (15 CT scans).
5. Pancreas-CT (TCIA) (14 CT scans [9]).
6. DeepLesion (14 CT scans [10]).
7. Head-Neck-PET-CT (14 CT scans [11]).
8. ACRIN-NSCLC-FDG-PET (ACRIN 6668) (14 CT scans, 12 PET images [12]).
9. Anti-PD-1 Immunotherapy Lung (12 PET images [13]).
10. Ultrasound data of a variety of liver masses (B-mode-and-CEUS-Liver) (15 ultrasound images [14]).
11. Prostate MRI and Ultrasound with Pathology and Coordinates of Tracked Biopsy (Prostate-MRI-US-Biopsy) (10 ultrasound images, 7 MRI Images and 5 3D masks [15]).
12. OASIS (10 MRI images [16]).
13. BRAINS Imagebank (10 MRI images [17]).
14. Breast-MRI-NACT-Pilot (10 MRI images [18]).
15. MRI Dataset for Hippocampus Segmentation (10 MRI images, 5 3D masks [19]).

We also used an additional dataset to test the trained model, which contains 50 non CT series. Test dataset includes cases from the aforementioned datasets, as well as additional datasets of digital mammography and radiography:

1. The VICTRE Trial: Open-Source, In-Silico Clinical Trial for Evaluating Digital Breast Tomosynthesis (35 mammography images [20]).

2. CPTAC-PDA (15 XRay images [21]).

**Dataset used for checking of lungs presence.** The training dataset (acronym DS2) is divided into two classes (lungs and non-lungs) with 293 images in each class for training and 41 images for testing part (Fig. 1), total 668 3D images.

### Dataset (size is 31.6 GB)

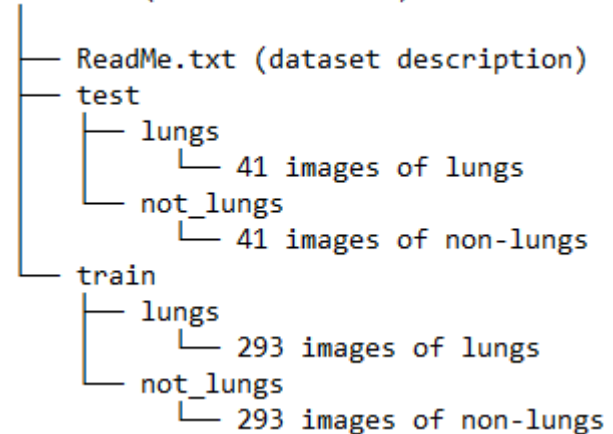


Fig. 1 Dataset DS2 directory structure which is used for checking of lungs presence.

The CT image data used in this study originated from the following twelve sources.

1. STOIC2021-COVID-19 AI Challenge (57 CT scans of lungs [22]).
2. Private dataset of Tuberculosis Portals (81 CT scans of lungs [1]).
3. ACRIN-NSCLC-FDG-PET (ACRIN 6668) (256 CT scans of different body parts [12]).
4. CT images with lung caverns from National Center of Tuberculosis and Lung Diseases, Georgia (13 CT scans of lungs and 11 3D lung masks).
5. MosMedData: Chest CT Scans with COVID-19 Related Findings COVID19\_1110\_1.0 (24 CT scans of lungs and 11 3D lung masks [23]).
6. RibFrac Dataset: A Benchmark for Rib Fracture Detection, Segmentation and Classification (23 CT scans of lungs [24]).
7. Coronacases Initiative and Radiopaedia (10 CT scans of lungs).
8. IH Pancreas-CT Dataset (43 CT scans of abdomen [9]).
9. Head-Neck-PET-CT (68 CT scans of the head and upper body [11]).
10. Private CT dataset from UIIP NASB (21 CT lung masks).
11. CQ500 Dataset (23 CT scans of the head and upper body [25]).
12. NODE21 dataset (27 CT images of box-shaped regions containing lung nodules [26]).

Some smaller part of dataset images was artificially created and added to the train and test datasets:

1. Artificial biopsy images and multiple lungs to classify them as non-lungs;
2. Doubled lungs to classify them as non-lungs;
3. Slightly cropped images at the top and bottom of lungs to classify them as lungs;
4. Significantly cropped images of lungs to classify them as non-lungs;
5. Images of chest with head to classify them as lungs.

## 2.2 The Method of CT Image Modality Check

After trivial checks which are not considered in this paper, the next step is to check if the modality of an input data is CT scan or not. Medical input data are often provided in DICOM format, when a modality attribute (0008,0060) is available, so the modality check is trivial in this case. However, for a general purpose only image data are available, thus the more sophisticated task of modality estimation by image content becomes the only way. In order to solve this problem, we use a Deep Learning technique based on convolutional neural networks. The CNN was trained on the DS1 image dataset (see above), which contains images of various modalities including CT, ultrasound, PET, and MRI. Examples are presented in Fig. 2.

CNN architecture is shown in Fig. 3 and consists of the following layers: Conv3D, MaxPool3D, Dropout, GlobalAveragePooling3D.

Neural network parameters were selected in practice during numerous training sessions. The study showed that the input layer with resolution  $64 \times 64 \times 64$  and convolutional layer with 25 output filters turned out to be acceptable for the modality check method. Training data were normalized and rotated by 90, 180, and 270 degrees, which eliminates inaccuracies associated with image rotation.

## 2.3 Method of Checking of Lungs Presence

To determine if a CT scan contains images of a person's lungs and not another part of the body, we used CNN to classify lungs and non-lungs, i.e. a binary classification task was implemented. The CNN was trained on the DS2 image dataset, which included chest CT scans with lungs and other body parts (non-lungs).

Obvious CT scans of non-lungs type are masks, head, abdomen, limbs, scout preliminary scans, empty scans, etc. However, the difference between lungs and non-lungs is not as obvious as it seems. We agreed that the following categories are "non-lungs" class:

- biopsy scans of lungs;
- double, triple and multiple lungs in one image;
- whole body scans including lungs, abdomen, head, limbs, etc.;

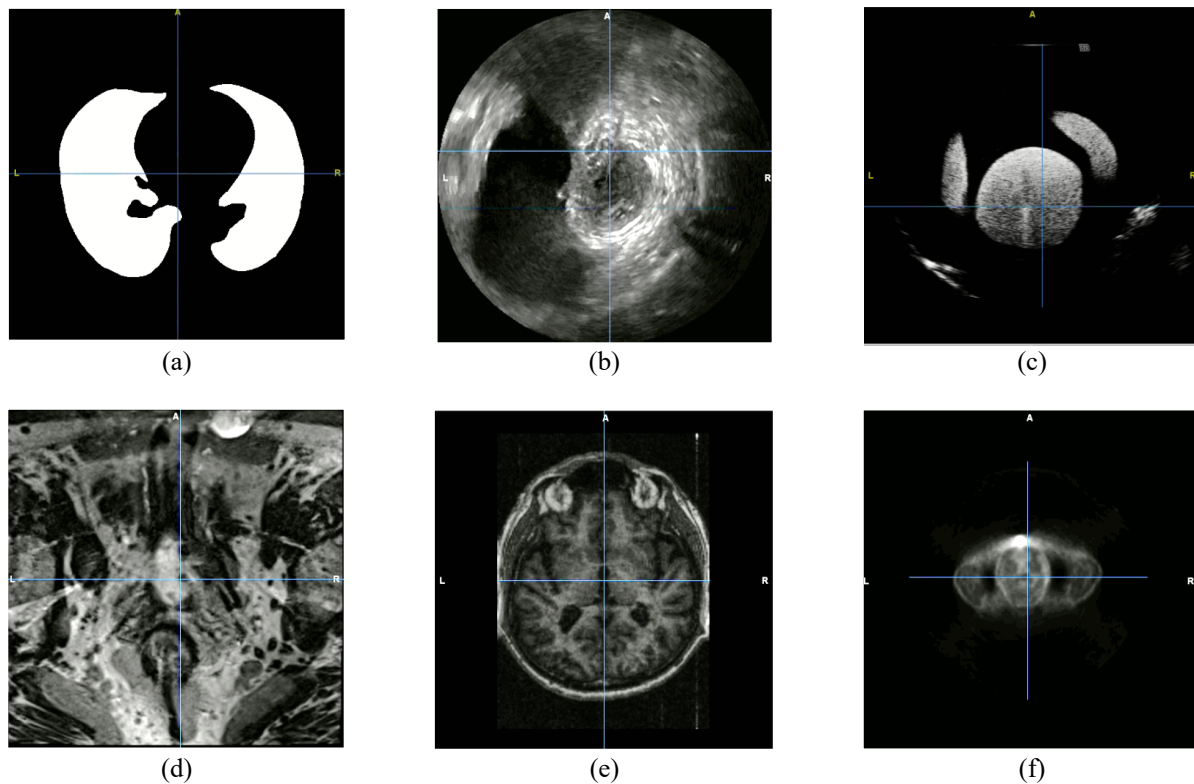


Fig. 2 Examples of 2D layers of 3D images of the "non-CT" class in the dataset DS1 for CT images modality check: (a) binary mask of lungs ( $512 \times 512$  pixels); (b) ultrasound images of prostate ( $452 \times 452$  pixels); (c) ultrasound images of baby phantom ( $532 \times 416$  pixels); (d) MRI images of prostate ( $256 \times 256$  pixels); (e) MRI images of brain ( $256 \times 256$  pixels); (f) PET image of body ( $128 \times 128$  pixels).

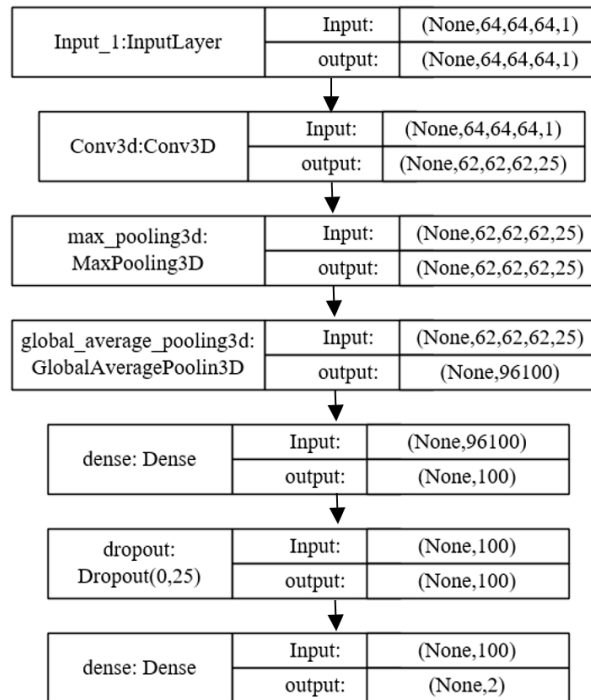


Fig. 3 Convolutional neural network architecture for CT images modality check.

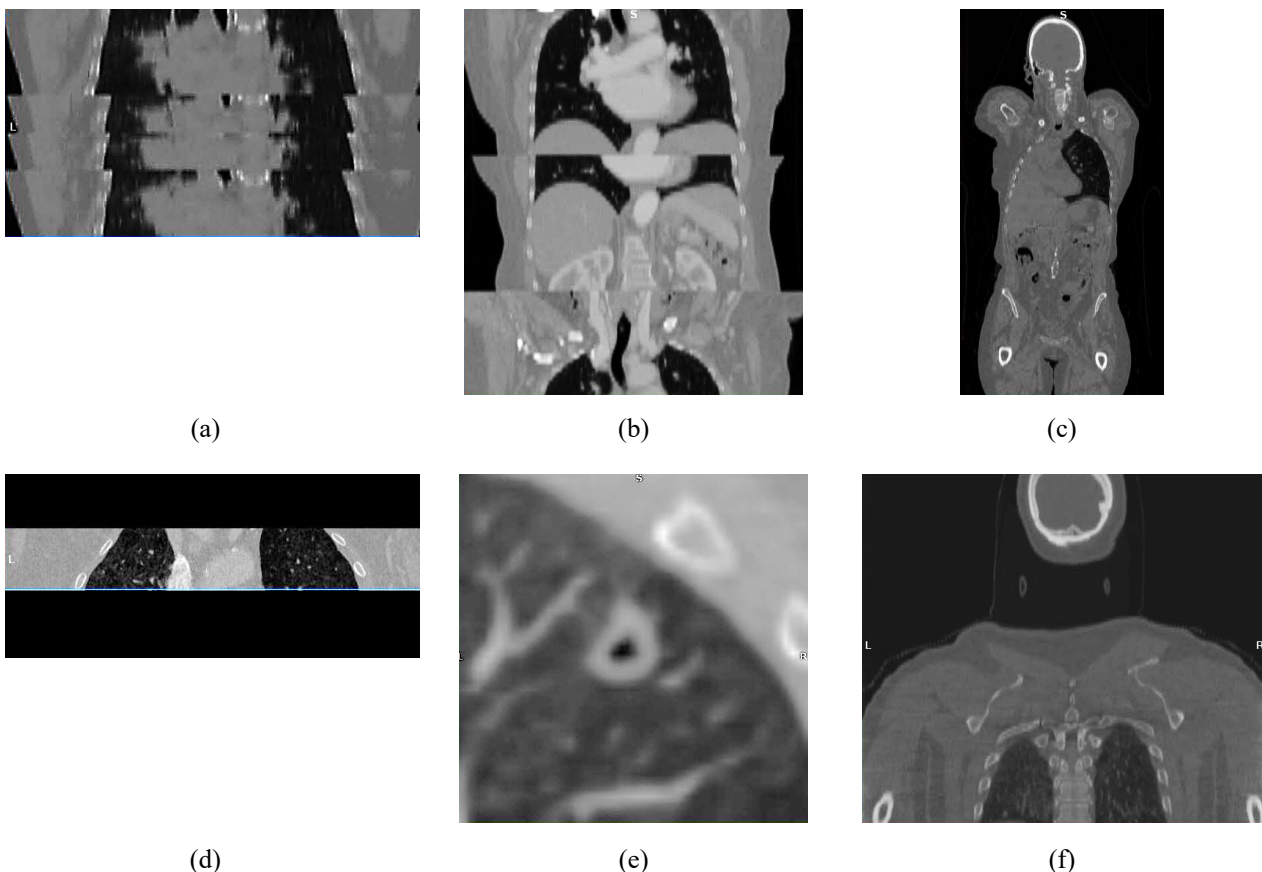


Fig. 4 Examples of 2D layers of 3D images of the non-obvious cases of “non-lungs” class in the DS2 dataset for checking of lungs presence: (a) biopsy scan of lungs (512×43 pixels); (b) triple lungs in one image (512×87 pixels); (c) whole body scan including lungs, abdomen, head, limbs, etc. (512×287 pixels); (d) pancake-like (significantly cropped) scan of lungs (512×51 pixels); (e) 3D box-shaped region containing lung nodules (50×50 pixels); (f) lungs occupy less than 50% of the coronal axis (512×135 pixels).



- pancake-like (significantly cropped) scans of lungs;
- 3D box-shaped regions containing lung nodules;
- if the lungs occupy less than 50 % of the coronal axis.
- Examples of non-obvious types are shown in Fig. 4.

Binary classification of lungs from non-lungs is a relatively simple task, so a simple 4-block CNN was used (Fig. 5).

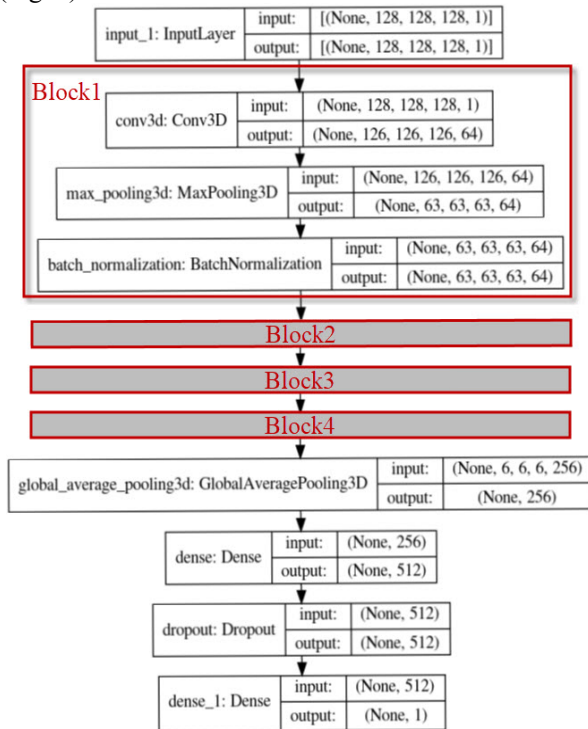


Fig. 5 Convolutional neural network architecture for checking of lungs presence.

Each block of the 4-block neural network consists of the same three layers: Conv3D, MaxPool3D, and BatchNormalization. After four blocks, the GlobalAveragePooling3D layer is added. Convolution layers and the penultimate dense layer have ReLU activation and the final dense layer has Sigmoid activation to perform binary classification.

Files are provided in NifTI format with the “.nii.gz” extension. To read the scans, we use the “nibabel” package. CT scans store raw voxel intensities in Hounsfield units (HU), which vary widely and may differ from dataset to dataset. Above 500 HU are bones with different radio intensity, so this is used as a higher bound. Air and subtle tissues usually associated with  $[-1500, 0]$  HU range. A threshold between  $-1500$  and  $500$  is commonly used to normalize CT scans. To preprocess the data, we do the following [27]:

1. Apply threshold  $[-1500, 500]$  HU;
2. Scale the HU values down to the  $[-1, 1]$  range;
3. Resize width, height and depth to  $128 \times 128 \times 128$ , the volume has to be cubic for the further augmentation;

4. Rotate the volumes by 90 degrees, so the orientation is fixed.

The last step with rotation by 90 degrees is not necessary, because of subsequent augmentation with random rotations. To increase preprocessing speed, we used a “multiprocessing” Python module so that all or part of the CPU cores can be used in a parallel mode.

Augmentation is used both for the training and validation sets. The reason to use augmentation for the validation set is that we have many augmented CT images in our testing set. Also note that a random transposition followed by a flip is equivalent to random rotation. Augmentation function:

1. Permutes randomly (transposes) three axes of a 3D CT image;
2. Flips randomly a volume out of eight different choices, including identity flip (or None), which does nothing (flips =  $[None, 0, 1, 2, (0,1), (0,2), (1,2), (0,1,2)]$ );
3. Finally makes sure that the volume stays between the  $[-1, 1]$  range.

Neural network architecture, augmentation and hyperparameters were modified several times. After each modification and training of the model it was used to the datasets mentioned above to find wrong predictions between lungs and non-lungs (false-positive and false-negative predictions). All images found with incorrect predictions were visually reviewed and added to our dataset.

## 3 Results and Discussions

### 3.1 CT Image Modality Check

The trained model for method of modality check achieved 100% classification accuracy on the test dataset, which was not involved in model training. All of the images were attributed to the desired class. Model accuracy and loss for train and validation sets are shown in Fig. 6.

Results of binary classification of CT from non-CT images are shown in Table 1.

Table 1 Results of binary classification of CT from non-CT images.

Dataset	Accuracy, %	Number of images	Number of wrong predictions
train	100	214	0
validation	100	12	0
test	100	280	0

The model was trained five times and each time it showed one hundred percent result on a test dataset. The train and validation datasets changed when learning every time. Therefore, the results of the model predictions do not depend on the data composition in the training set.

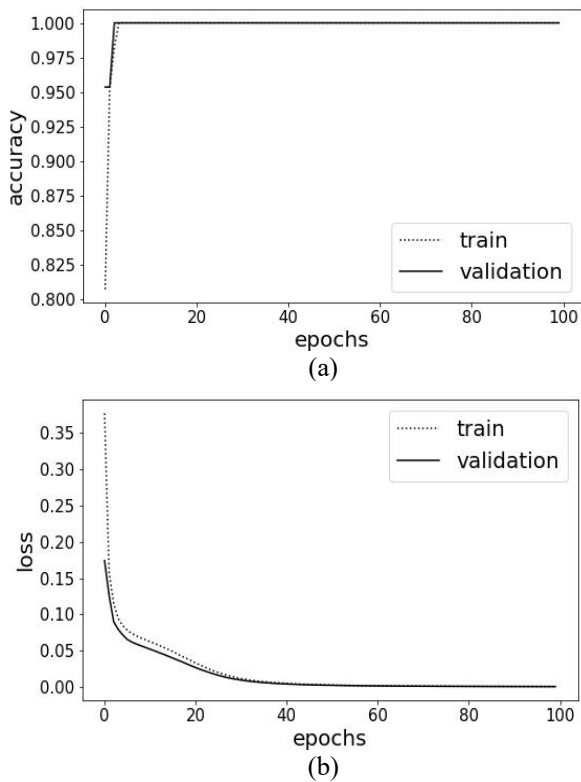


Fig. 6 Training diagram for model accuracy (a) and loss (b) for classification of CT from non-CT images in 100 epochs for dataset DS1.

Regarding 100% accuracy in modality verification, we did a similar study to verify the modality of X-ray images (similar to CT), and we are preparing a new article on this topic. During the experiments, it turned out that the same method showed less than 70% accuracy for X-ray images. So, we had to significantly increase the training samples up to 1000 images for each class and use more advanced architecture of EfficientNet CNN to achieve 99.8% accuracy. For this reason, we believe that CT images have significant differences from MRI, ultrasound, binary masks, histological images, etc.

Finally, based on our results we can conclude that the method can be used for CT images modality check. Despite the fact that the model showed a good result on the CT images, this approach needs to be explored to solve problems of determining other modalities. This is necessary for the development of medical (non-CT) imaging services.

### 3.2 Checking of Lungs Presence

Results for binary classification of lungs from non-lungs are not so unambiguous. Model training used a large patience for callback “keras.callbacks.EarlyStopping” equal to 50 epochs and stopped training automatically after 147 epochs. Model accuracy and loss for train and validation sets are shown in Fig. 7.

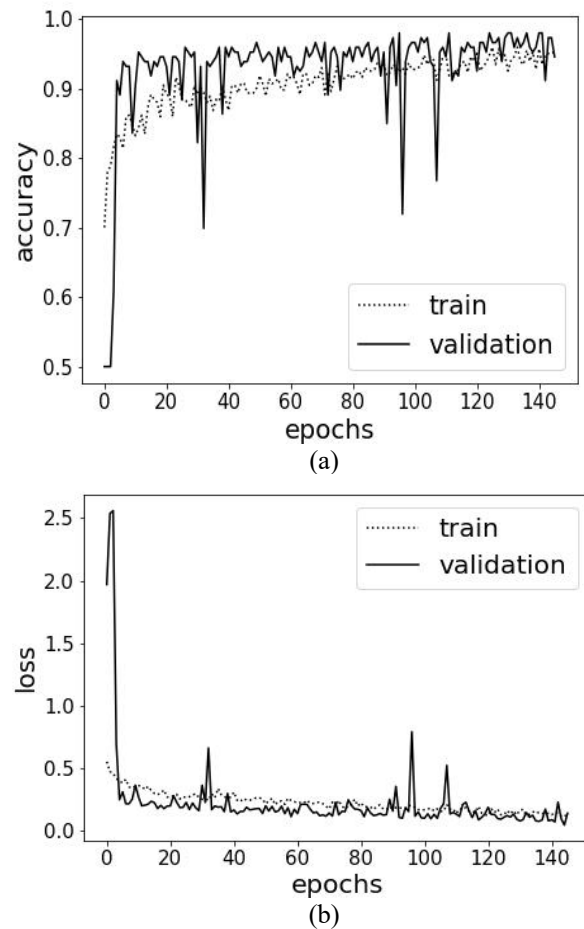


Fig. 7 Training diagram for model accuracy (a) and loss (b) for classification of lungs from non-lungs in 147 epochs for dataset DS2.

Results of checking of lungs presence are shown in Table 2.

Table 2 Results for classification of lungs from non-lungs.

Dataset	Accuracy, %	Number of images	Number of wrong predictions
train	98.64	440	6
validation	98.63	146	2
test	89.02	82	9

From the total 17 wrong predictions only one CT scan is from the STOIC2021 dataset. Other 16 CT images are from the ACRIN-NSCLC-FDG-PET (ACRIN 6668) dataset [12]: eleven are artificially created; three are biopsy scans; one real CT scan with chest and part of the head; one real pancake-like (significantly cropped) scan of lungs.

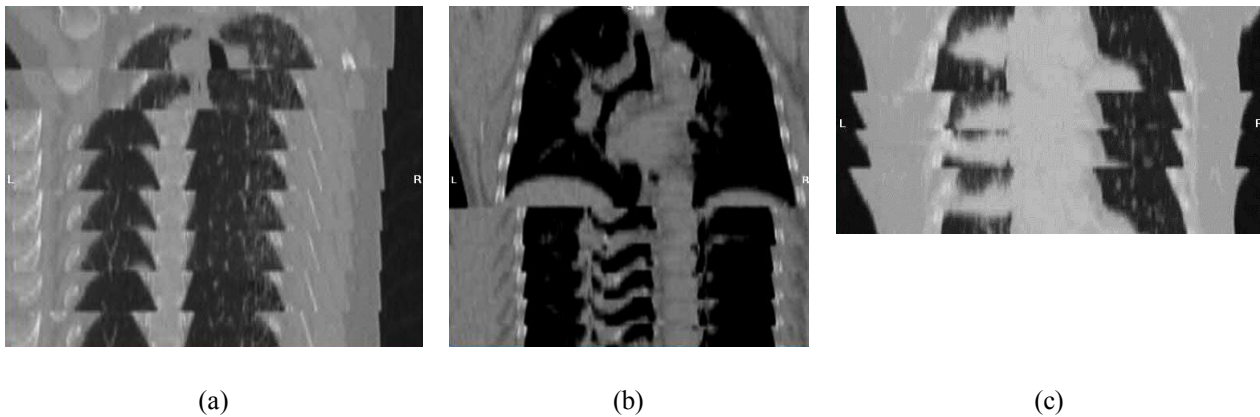


Fig. 8 Examples of biopsy CT images, coronal view (2D layer image sizes are  $512 \times 69$  (a),  $512 \times 65$  (b), and  $512 \times 43$  pixels).

Eleven artificially created scans with wrong predictions are: four out of total six artificially made biopsy scans; three out of total eight artificially made doubled lungs; two out of total ten artificially made pancake-like (significantly cropped) scans of lungs; and two out of total fifteen artificially made CT scans of lungs with part of the head. A possible reason for the low classification accuracy for artificially created images is their small number, only about a dozen for each non-standard case in DS2. Perhaps we need to add more artificial images to improve the classification accuracy.

Analysis of wrong predictions showed that the model performs poorly on biopsy of lungs. Seven out of total nine (real and artificially created) biopsy CT scans were incorrectly classified as lungs. A possible reason why CT biopsies are incorrectly classified as “lungs” is that their axial view does not differ from normal lungs. However, coronal and sagittal views differ a lot from the lungs class (Fig. 8).

To improve the classification accuracy of biopsy CT scans, we have to add more real and artificial biopsy scans and try more complex standard neural network architectures.

#### 4 Conclusions

The developed validation model shows good results on datasets DS1 and DS2: 100% for CT image modality

check and 89.02% for checking of lungs presence. For example, we found nine chest CT scans of lungs which were mistakenly placed in the dataset of head CT scans “CQ500 Dataset” [25]. These scans of lungs were visually checked and added into the “lungs” folder to improve model accuracy. The task of verifying medical data requires further study. We are going to check 2D images for XRay of human lungs. It is necessary to develop a more general method capable of solving the evaluation problem simultaneously at different levels. In general, our trained models performed well and are already being used in our application called “AI-based software for computer-assisted diagnosis of lung diseases using chest X-Ray and CT images” (LungExpert) [1] for screening of incoming data.

#### Disclosures

The authors declare that they have no conflict of interest.

#### Acknowledgments

This work was carried out with the financial support of NIAID, NIH, USA, in the framework of the CRDF G-DAA9-20-67103-1 Project.

#### References

1. A. Rosenthal, A. Gabrielian, E. Engle, et al., “The TB Portals: an Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis,” *Journal of Clinical Microbiology* 55(11), 3267–3282 (2017).
2. S. Kaplan, D. Handelman, and A. Handelman, “Sensitivity of neural networks to corruption of image classification,” *AI Ethics* 1, 425–434 (2021).
3. D. Guan, W. Yuan, Y.-K. Lee, and S. Lee, “Identifying mislabeled training data with the aid of unlabeled data,” *Applied Intelligence* 35(3), 345–358 (2011).
4. A. P. Brady, “Error and discrepancy in radiology: inevitable or avoidable?” *Insights Imaging* 8(1), 171–182 (2017).
5. K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, “The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository,” *Journal of Digital Imaging* 26(6), 1045–1057 (2013).
6. J. P. Cohen, P. Morrison, and L. Dao, “COVID-19 Image Data Collection,” arXiv:2003.11597v1 (2020).

7. J. Yang, G. Sharp, H. Veeraraghavan, W. Van Elmpt, A. Dekker, T. Lustberg, and M. Gooding, “[Data from Lung CT Segmentation Challenge](#),” The Cancer Imaging Archive (2017).
8. S. G. Armato, G. McLennan, L. Bidaut, et al., “[The Lung Image Database Consortium \(LIDC\) and Image Database Resource Initiative \(IDRI\): A Completed Reference Database of Lung Nodules on CT Scans: The LIDC/IDRI thoracic CT database of lung nodules](#),” *Medical Physics* 38(2), 915–931 (2011).
9. H. Roth, A. Farag, E. B. Turkbey, L. Lu, J. Liu, and R. M. Summers, “[Data From Pancreas-CT](#),” The Cancer Imaging Archive (2016).
10. K. Yan, X. Wang, L. Lu, and R. M. Summers, “[DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning](#),” *Journal of Medical Imaging* 5(3), 036501 (2018).
11. M. Vallières, E. Kay-Rivest, L. Perrin, X. Liem, C. Furstoss, N. Khaouam, P. Nguyen-Tan, C.-S. Wang, and K. Sultanem, “[Data from Head-Neck-PET-CT](#),” The Cancer Imaging Archive (2017).
12. P. Kinahan, M. Muzi, B. Bialecki, B. Herman, and L. Coombs, “[Data from the ACRIN 6668 Trial NSCLC-FDG-PET](#),” The Cancer Imaging Archive (2019).
13. M. Patnana, S. Patel, and A. S. Tsao, “[Data from Anti-PD-1 Immunotherapy Lung](#),” The Cancer Imaging Archive (2019).
14. J. Eisenbrey, A. Lyshchik, and C. Wessner, “[Ultrasound data of a variety of liver masses](#),” The Cancer Imaging Archive (2021).
15. S. Natarajan, A. Priester, D. Margolis, J. Huang, and L. Marks, “[Prostate MRI and Ultrasound With Pathology and Coordinates of Tracked Biopsy \(Prostate-MRI-US-Biopsy\)](#),” The Cancer Imaging Archive (2020).
16. P. J. LaMontagne, T. LS. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko, M. E. Raichle, C. Cruchaga, and D. Marcus, “[OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease](#),” medRxiv 2019.12.13.19014902 (2019).
17. D. E. Job, D. A. Dickie, D. Rodriguez, A. Robson, S. Danso, C. Pernet, M. E. Bastin, J. P. Boardman, A. D. Murray, T. Ahearn, G. D. Waiter, R. T. Staff, I. J. Deary, S. D. Shenkin, and J. M. Wardlaw, “[A brain imaging repository of normal structural MRI across the life course: Brain Images of Normal Subjects \(BRAINS\)](#),” *NeuroImage* 144, 299–304 (2017).
18. D. Newitt, N. Hylton, “[Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy](#),” The Cancer Imaging Archive (2016).
19. K. Jafari-Khouzani, K. Elisevich, S. Patel, and H. Soltanian-Zadeh, “[Dataset of magnetic resonance images of nonepileptic subjects and temporal lobe epilepsy patients for validation of hippocampal segmentation techniques](#),” *Neuroinformatics* 9, 335–346 (2011).
20. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), “[The Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma Collection \(CPTAC-PDA\)](#),” The Cancer Imaging Archive (2018).
21. A. Badano, C. G. Graff, A. Badal, D. Sharma, R. Zeng, F. W. Samuelson, S. Glick, and K. J. Myers, “[Data from the VICTRE trial: open-source, in-silico clinical trial for evaluating digital breast tomosynthesis \(VICTRE\)](#),” The Cancer Imaging Archive (2019).
22. M.-P. Revel, S. Boussouar, C. de Margerie-Mellon, I. Saab, T. Lapotre, D. Mompoin, G. Chassagnon, A. Milon, M. Lederlin, S. Bennani, S. Molière, M.-P. Debray, F. Bompard, S. Dangeard, C. Hani, M. Ohana, S. Bommart, C. Jalaber, M. El Hajjam, I. Petit, L. Fournier, A. Khalil, P.-Y. Brillet, M.-F. Bellin, A. Redheuil, L. Rocher, V. Bousson, P. Rousset, J. Grégory, J.-F. Deux, E. Dion, D. Valeyre, R. Porcher, L. Jilet, and H. Abdoul, “[Study of Thoracic CT in COVID-19: The STOIC Project](#),” *Radiology* 301(1), E361–E370 (2021).
23. S. P. Morozov, A. E. Andreychenko, N. A. Pavlov, A. V. Vladzmyrskyy, N. V. Ledikhova, V. A. Gomboleviskiy, I. A. Blokhin, P. B. Gelezhe, A. V. Gonchar, and V. Yu. Chernina, “[MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset](#),” arXiv:2005.06465v1 (2020).
24. L. Jin, J. Yang, K. Kuang, B. Ni, Y. Gao, Y. Sun, P. Gao, W. Ma, M. Tan, H. Kang, J. Chen, and M. Li, “[Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet](#),” *eBioMedicine* 62, 103106 (2020).
25. S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, “[Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study](#),” *The Lancet* 392(10162), 2388–2396 (2018).
26. E. Sogancioglu, K. Murphy, and B. Van Ginneken, “[NODE21](#),” Zenodo (2021).
27. H. Zunair, A. Rahman, N. Mohammed, and J. P. Cohen, “[Uniformizing Techniques to Process CT Scans with 3D CNNs for Tuberculosis Prediction](#),” in *Predictive Intelligence in Medicine*, I. Rekik, E. Adeli, S. H. Park, and M. del C. Valdés Hernández (Eds. ), Springer International Publishing 12329, Cham, 156–168 (2020).