Automatic Out-of-Distribution Detection Methods for Improving the Deep Learning Classification of Pulmonary X-ray Images

Andrey A. Dovganich, Alexander V. Khvostikov, Yakov A. Pchelintsev, and Andrey A. Krylov
Department of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia
Email: a.dovganich@yandex.ru, khvostikov@cs.msu.ru, yakov.pchelintsev@gmail.com, kryl@cs.msu.ru

Yong Ding
Zhejiang University, Hangzhou, China
Email: dingy@vlsi.zju.edu.cn

Mylene C. Q. Farias
University of Brasilia, Brasilia, Brazil
Email: mylene@unb.br

***Abstract*—In this paper, we investigated the possibility of using medical differential criteria to determine the level of radiation in X-ray images of the lungs. We developed a new method for automatic determination and calculation the number of visible vertebrae in the pulmonary X-ray images and proposed a system of automatic out-of-distribution detection that can be used together with deep learning-based systems of pulmonary X-ray image analysis, in particular with the task of tuberculosis detection. The proposed method and system were evaluated using three X-ray lung datasets (Montgomery County chest X-ray dataset, Shenzhen chest X-ray dataset and Tuberculosis X-ray TBX11K dataset). We demonstrated that using the proposed system of out-of-distribution detection allows to enhance the tuberculosis classification results up to 1.3% using the same classification model. We also showed that the proposed system allows to automatically train a composite model which considers X-ray radiation level of the image, which is more effective compared to the traditional one-part model.**

***Index Terms*—X-ray images, deep learning, quality assessment, out-of-distribution detection, detection of vertebrae**

I. Introduction

Algorithms for automatic processing of medical data that utilize deep neural networks are growing in popularity not only among researchers but also among their end users in medical organizations. As for image data, these deep neural networks are represented mainly by convolution neural networks and are used to solve various image analysis tasks including classification, segmentation, object detection and anomaly detection. With the increasing availability of medical datasets, hardware development and the appearance of simple tools for implementing deep learning-based pipelines, these algorithms become easier to deploy and customize for any specific task.

However, being extremely dependent on the training dataset, neural network-based algorithms demand the data to be representative and include examples close to all possible inputs at the same time having as few outliers ("bad", "garbage" examples) as possible. Besides, complex, and therefore powerful, neural networks contain a lot of tunable parameters, which makes them prone to overfitting even to very rare abnormal examples in case of small training datasets. Thus, the task of image quality assessment remains open in various areas [1]-[4].

This paper is devoted to solving the task of automatic radiation level evaluation for the pulmonary chest X-ray images. These images are used to make a diagnosis in case of many diseases, for instance, Tuberculosis (TB) [5], [6] and estimation of lung damage for COVID-19 patients [7]-[9].

The chest X-ray images in existing available training datasets, as well as the chest X-ray images obtained by medical experts in practice, may vary widely in contrast and number of details due to the difference of the X-ray hardness (radiation) level used during the image acquisition [10]. Moreover, public datasets often don't provide enough information to determine the image acquisition conditions of a certain image. This and also presence of outliers of various nature (inappropriate image obtained due to some exporting issues, an incorrect perspective angle, incorrect focus, too harsh or insufficient exposure etc.) in training datasets and real medical data complicates the correct inference for end-use applications of machine learning algorithms. Therefore, various algorithms for automatic elimination of outliers and distinction of hard and soft X-ray images are in much demand now since they can help to obtain more consistent

Manuscript received November 25, 2021; revised April 10, 2022.

doi: 10.18178/joig.10.2.56-63

datasets, expand the scope of deep learning-based systems of image analysis and make them more robust.

Such approaches of dataset cleaning have already been considered, for instance, in tasks of human face processing [11] and sensor fault detection [12].

In this paper we propose an algorithm for automatic out-of-distribution detection for improving the deep learning classification of pulmonary X-ray images. The main idea is to thin the training dataset by removing outliers, which is then can be followed by the division of the dataset into two parts depending on the hardness of images, that are used to train two distinct neural networks to solve the end task for their own image domain, which may increase the inference quality.

II. Used Datasets

In this work for training and evaluating all constructed X-ray classification models we used three different tuberculosis datasets.

The first dataset is Montgomery County chest X-ray set (MC) [13]-[15]. It is collected through the cooperation with Department of Health and Human Services, Montgomery County, Maryland, USA. MC dataset consists of 138 grayscale X-ray images, 80 of which correspond to healthy cases and 58 correspond to cases with manifestations of TB. Images resolution is ~4000×4900 pixels, the color depth is 8 bits per pixel.

The second dataset is Shenzhen chest X-ray set (Shenzhen) [13]-[15]. It is collected through the cooperation with Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China. Shenzhen dataset is composed of 326 norm cases and 336 cases with manifestations of TB, leading to 662 grayscale X-ray images in total. Images resolution is ~3000×2900, color depth is 8 bits per pixel.

These two datasets have been made available by the US National Library of Medicine.

The third dataset is the Tuberculosis X-ray (TBX11K) dataset [16]. TBX11K dataset contains 11200 X-ray grayscale images, including 5000 healthy cases, 5000 sick but non-TB cases, and 1200 cases with manifestations of TB with corresponding bounding box annotations for tuberculosis areas. TBX11K dataset is collected through the long-term cooperation with major hospitals in Nankai. All images are with a size of 512×512 and have color depth of 8 bits.

III. Proposed Method

In this paper, we proposed an automatic system for detection out-of-distribution pulmonary X-ray images based on radiologist's experience of data annotation for improving application possibilities of deep learning models used for pulmonary X-ray image analysis.

In order to cluster pulmonary X-ray images into groups corresponding to the radiation level radiologists usually count the number of visible vertebrae in the spinal curve. Fixing a threshold for the number of visible vertebrae they can split the images into "soft" and "hard" X-ray images. Here it should me mentioned that there is no one medical gold-standard rule or threshold to split the X-ray images into "soft" and "hard" classes and this terminology is more qualitative than quantitative emphasis. In this paper we use the splitting threshold which fits well to the obtained TB classification task according to the recommendations of medical experts.

It is obvious that to create an effective learning-based model for pulmonary X-ray image analysis, the images that are used for training and evaluation of this model have to be from the same distribution. And one of the most accurate way of describing and clustering the distribution of encountered in practice X-ray images from the medical point of view is based on the calculation of the quantity of visible vertebrae in the image. Thus, in this paper we propose a new algorithm for automatic estimation of the number of visible vertebrae in X-ray images and demonstrate the possibility of its application for detecting out-of-distribution cases which has a positive effect on the use of neural network models for pulmonary X-ray image classification.

In this paper we propose a new method for automatic determination of the X-ray radiation level based on the number of detected vertebrae, demonstrate the possibility of its' application for detecting out-of-distribution X-ray images and describe several ways in which this method can be used to build more effective deep learning-based systems for TB classification task. The flowchart of the proposed automatic out-of-distribution detection system for pulmonary X-ray images is shown in Fig. 1.

The proposed method of automatic determination of the X-ray radiation level based on the number of detected vertebrae includes four stages: image preprocessing, isolation of the spine region, detection of central and boundary spinal curve lines, vertebrae detection and counting.

A. Preprocessing

Firstly, all input images are resized to the 512×512 resolution. Then we try to resolve the problem of X-ray images illumination spectrum distortion. There are three main possible methods of contrast enhancement: contrast stretching, classic histogram equalization and adaptive histogram equalization.

Contrast stretching stretches the histogram of the image across the entire intensity scale. It works well when some part of the spectrum is completely absent. The standard histogram equalization method works well for images with very low overall contrast. However, histograms of pulmonary X-ray images cover the entire spectrum of intensity values, but low contrast is observed in local areas. The main advantage of Adaptive Histogram Equalization (AHE) is that it can provide better contrast in local areas compared to the traditional histogram equalization methods. Whereas traditional methods consider the entire image, AHE utilizes a local contextual region. The comparison of applying all three algorithms on a sample pulmonary X-ray image is shown in Fig. 2. In this work we use AHE as the most suitable for the current task of X-ray image processing.

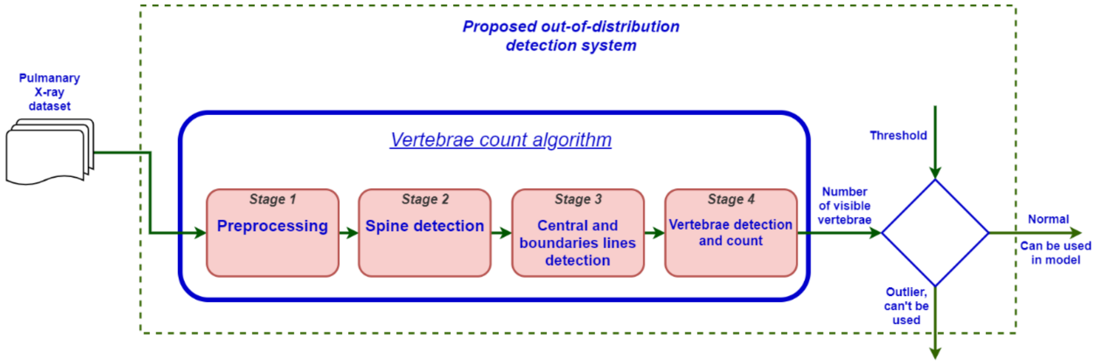


Figure 1. Flowchart of the proposed system for automatic out-of-distribution detection.

B. *Spine Detection*

To isolate the spine region inside the X-ray image, we apply pretrained CNN ResNet-34 [17]. For each image we locate a bounding box region on spine area. We used MC and Shenzhen datasets described above for this task, since they contain the spine area annotation. The data was divided into two subsets for training and validation (test). The ratio of the number of TB X-rays for training and test is 4:1. We trained ResNet-34 for 10 epochs with the initial learning rate of 10^{-3} and L1 loss function. In addition, Intersection over Union (IoU) value is used for quality assessments after every epoch. Results of the metrics during training are shown in Table I.

TABLE I. L1-LOSS AND IOU VALUES WHILE TRAINING RESNET-34 MODEL FOR SPINE AREA DETECTION

Epoch	L1-loss on training set	L1-loss on test set	IoU on training set	IoU on test set
2	0.534	0.257	0.001	0.012
3	0.136	0.064	0.470	0.627
5	0.053	0.053	0.741	0.676
10	0.049	0.050	0.755	0.698

C. *Detection the Central and Boundary Lines of Spine*

After extracting the spine region, we further detect the locations of vertebrae. In general, the spine usually appears with a higher intensity in the cropped spine Region of Interest (ROI). Therefore, we can detect the edges of the spine by using the sums of the intensity and gradient. In order to correctly detect the vertebrae, we first detect the central and boundary lines of spine using two-steps algorithm:

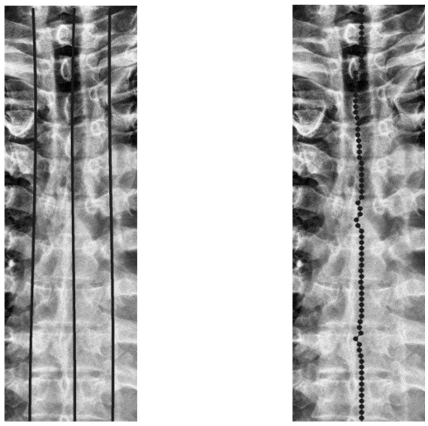
1) detection of the Central Line Segment (CLS),
2) detection of the spine boundary.

The first step is aimed to detect the central line segment (CLS) of the vertebrae. In this step, a number of rectangle windows with a size of $H \times W$ pixels are placed with one-pixel increment along the top of the spine ROI from left to right with overlapping. The sums of intensity inside each rectangle window are calculated. If one rectangle window has the largest sum of intensity, the top middle point of this window is used as the first reference point for CLS:

$$s(x,y) = \sum_{i=0}^{H} \sum_{j=-W/2}^{W/2} I(x+j, y+i) \qquad (1)$$

$$(x_{max}, y_{max}) = argmax_{(x,y)}(s) \qquad (2)$$

Next, the current rectangle window with a maximum sum of intensity is moved down by p pixels, and then a search for the next reference point is initiated in the range of q pixels on both of its sides. This search is shifted by 1 pixel at once and then records the intensity sum of the corresponding window. The window with the maximum sum of intensity value is then assigned to the current window and its top middle point is defined as the second reference point for CLS. This procedure is repeated until n reference points are detected, which are afterwards transformed into a CLS applying a polynomial fitting method (Fig. 2).



(a) (b)

Figure 2. Detection the central and boundary lines of spine with the proposed algorithm: (a) detection of the points for Central Line Segment (CLS), (b) detection of the CLS and spine boundary.

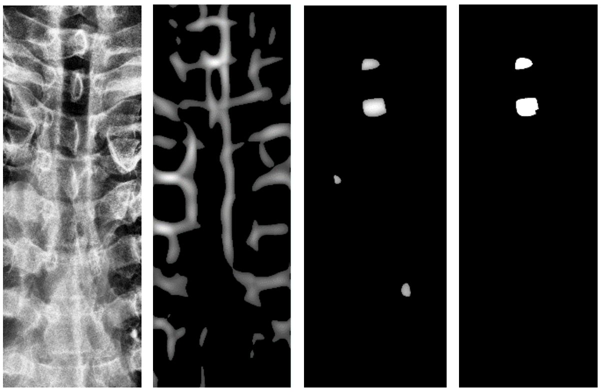
Next, in the second step, the boundary points of the spine along the normal direction of the detected central line segment are determined. This step utilizes two small sliding windows, each of 11×5 pixels. The pair of sliding windows moves at most by r pixels along both sides in normal directions of the corresponding CLS point. The top middle of the pair is chosen as the boundary point of the spine when their intensity difference is maximal. The boundary detection procedure continues until all points of the CLS are explored. The corresponding window of the final point for this CLS is reconstructed for sequential detection of the CLS until all boundaries of the spine are found. Finally, all spine boundary points in each side are dependently fitted by polynomial fitting with three degrees into the spine boundary. In the experiments, we set the following parameters: $H = \frac{H_I}{8}$, $W = \frac{W_I}{4}$, $p = \frac{H}{4}$, $q = \frac{W}{2}$, $r = \frac{W_I}{2}$, where H_I is the image height and W_I is the image width.

D. Vertebrae Detection and Counting

After detecting the central and boundary lines of spine we detect vertebrae inside the detected boundaries using the modified ridge map. The map is built through the following steps:

1) Finding the second derivatives of the image.
2) Creating matrix $Q = \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix}$, where L is the pixel intensity of the source image.
3) Finding eigenvalues λ_1 and λ_2 of matrix Q where $\lambda_1 > \lambda_2$.
4) Building ridge maps I_1 and I_2 using λ_1 and λ_2 values, where $I_i(x,y) = \frac{\lambda_i(x,y)}{\max(\lambda_i)} * 255,\ i = 1,2$ (see Fig. 3b, Fig. 3c).
5) Maps binarization using thresholds θ_1 и θ_2.
6) Intersection I_R of I_{1_θ} and I_{2_θ} (see Fig. 3d).

Then the number of independent connected areas along CLS is counted. The result is the number of vertebrae detected in the image.



(a) (b) (c) (d)

Figure 3. The steps of the proposed vertebrae detection algorithm for the spine region of a sample pulmonary X-ray image: (a) source image, (b) ridge map I_1, (c) ridge map I_2, (d) maps intersection I_R.

IV. APPLICATIONS OF THE PROPOSED METHOD

This section is devoted to several possible applications of the proposed vertebrae detection method for automatic thinning of TB X-ray datasets, which can noticeably improve the TB classification accuracy. In particular, we consider several thresholds of vertebrae limit k for outlier detection and two strategies of splitting data into subsets. For each case we train the same TB classification model and compare the obtained results.

In this paper we have chosen DenseNet architecture [18] for the considered TB classification task since it is still one of the state-of-the-art architectures for classification with limited amount of training data. We take a DenseNet model pretrained on the ImageNet dataset [19], replace the classification head with fully-connected layer with two outputs (since we consider binary classification: TB or healthy) and fine-tune the model on the pulmonary X-ray data for 20 epochs. In these experiments we use Stochastic Gradient Descent (SGD) optimizer with initial learning rate of 10^{-3} and momentum 0.9, the batch size was 8. We decay the learning rate by gamma 10^{-1} every 7 epochs.

We performed 4 types of experiments according to the number of detected vertebrae used as a threshold for thinning the dataset:

1) $k \geq 0$. No dataset thinning is performed, all images are used for training and validation.
2) $k > 0$. The number of detected vertebrae is more than 0. Using this scheme, we expect to remove only "bad" images from datasets. For example, X-ray images with poor focus or composition.
3) $3 \leq k \leq 8$. Number of detected vertebrae is between 3 and 8 inclusive. In this experiment we expect to leave only the most typical X-ray images that are received with correct hardware settings.
4) $5 \leq k \leq 7$. Number of detected vertebrae is between 5 and 7 inclusive. Here we want to further narrow the range of appropriate X-ray images and see whether it can enhance the classification results compared to the second type of threshold.

Besides that, for the first three types of thresholds mentioned above we consider 2 different dataset splitting strategies:

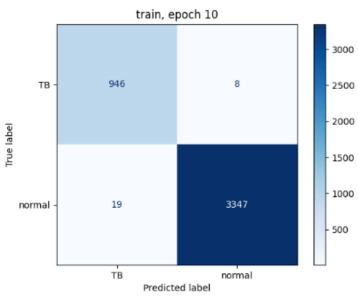
1) "One-part model". We use the whole thinned dataset as it is.
2) "Two-part model". We further split the thinned dataset into two subsets, corresponding to the "soft" and "hard" radiation level and train a separate DenseNet model for each of the subsets and then treat them as a single composite model, which analyzes the level of radiation and based on it performs prediction using one of the internal models.

The training scheme for all the experiments was organized as follows:

- MC, Shenzhen and TBX11K datasets were united into one single dataset.
- The united dataset was divided into train and test sets with proportion of 4:1.
- Train and test set were filtered using vertebrae counting method with selected threshold.
- Pretrained model was fine-tuned on the obtained dataset for 20 epochs.

A. $k \geq 0$. One-Part Model

In this experiment we do not perform any thinning for the used dataset and used all 5386 images.



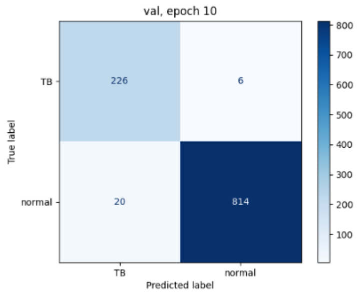


Figure 4. Confusion matrices for experiment with $k \geq 0$ (one-part model).

The obtained confusion matrices for the best epoch by validation are shown in Fig. 4. Balanced accuracy of the trained model on the validation set was 0.975.

B. $k>0$. One-Part Model

In this experiment we only remove "bad" X-Ray images from the dataset. This includes partially broken images, images with incorrect focus, non-pulmonary images, that can accidentally be included in the dataset while downloading from the database and so on. In the source united dataset there were 5386 images, after thinning 5040 images are left (94% of the source dataset). The obtained confusion matrices for best epoch by validation are shown in Fig. 5. Balanced accuracy of the trained model on the validation set was 0.984.

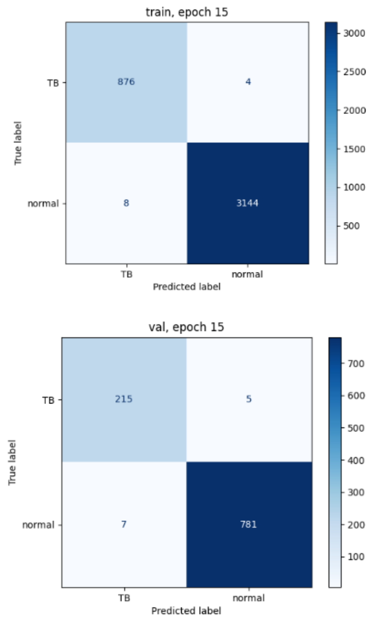


Figure 5. Confusion matrices for experiment with $k>0$ (one-part model). 94% of the source dataset is used.

C. $3 \leq k \leq 8$. One-Part Model

In this experiment we leave only those images that contain 3 to 8 vertebrae, which tends to represent the most typical X-ray images that are received with correct hardware settings. In this case after thinning the united dataset there were 3925 images left (73% of the source dataset). The obtained confusion matrices for best epoch by validation are shown in Fig. 6. Balanced accuracy of the trained model on the validation set was 0.985.

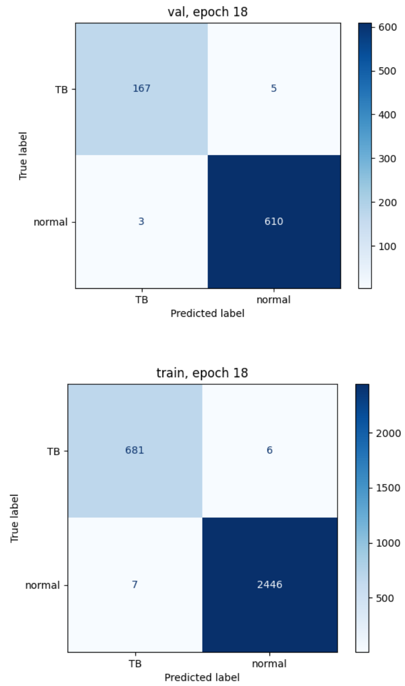


Figure 6. Confusion matrices for experiment with $3 \leq k \leq 8$ (one-part model). 73% of the source dataset is used.

D. $5 \leq k \leq 7$. One-Part Model

In this experiment we further narrow the boundaries of the used images and consider only those that have 5 to 7 vertebrae. Here the amount of image left in the united dataset after thinning was 1934 (36% from the source dataset). The obtained confusion matrices for best epoch by validation are shown in Fig. 7. Balanced accuracy of the trained model on the validation set was 0.967.

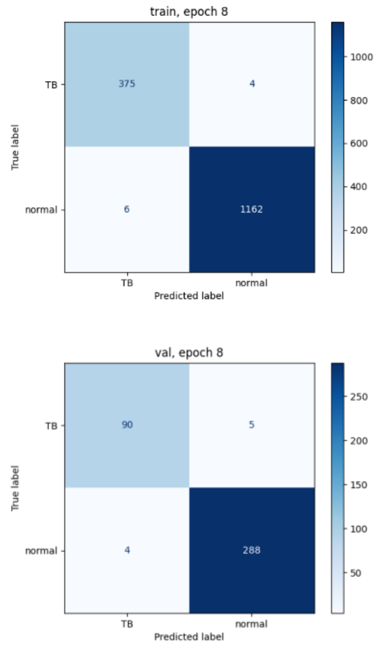


Figure 7. Confusion matrices for experiment with $5 \leq k \leq 7$ (one-part model). 36% of the source dataset is used.

E. $k \geq 0$. Two-Part Model

In this experiment we do not perform any thinning on the used dataset and used all 5386 images. The dataset is split into two sets according to the level of radiation. The first set contains images with $k \leq 4$ vertebrae, the second set contains images with $k \geq 5$ vertebrae. Two independent models are trained on these datasets. The balanced accuracy was calculated as for the "ordinary" single model and reached the value of 0.979.

F. $k>0$. Two-Part Model

In this experiment besides deleting "bad" images from the dataset we also split the dataset into two parts according to the level of radiation and process these two parts with separate models. In particular, the first model is trained only on images that have 1-4 vertebrae ($0<k\leq 4$) and the second model is trained on images that have more than 4 vertebrae ($k \geq 5$). The obtained two models are treated as one composite model. The balanced accuracy was calculated as for the "ordinary" single model and reached the value of 0.987.

G. $3 \leq k \leq 8$. Two-Part Model

In this experiment we select only those images that have the most typical level of radiation, that correspond to 3 to 8 visible vertebrae ($3 \leq k \leq 8$) and further split them into classes (with "soft" and "hard" level of radiation). First model is trained only on images with up to 4 vertebrae ($3 \leq k \leq 4$), the second model is trained on images with 5 to 8 vertebrae ($5 \leq k \leq 8$). The obtained two models are treated as one composite model. The balanced accuracy was calculated as for the "ordinary" single model and reached the value of 0.988.

H. Comparison of Experimental Results

The comparative results of all mentioned above experiments are shown below in Table II.

TABLE II. COMPARATIVE RESULTS OF THE CARIED OUT EXPERIMENTS ON DATASET THINNING USING THE PROPOSED AUTOMATIC METHOD OF VERTEBRAE DETECTION. THE PROBLEM OF BINARY CLASSIFICATION (TB, HEALTHY) FOR PULMONARY X-RAY IMAGES WAS CONSIDERED

Model Type	Dataset thinning	Dataset size (%)	Balanced accuracy
One-part	no ($k \geq 0$)	5386 (100%)	0.975
Two-part	no ($k \geq 0$)	5386 (100%)	0.979
One-part	$k > 0$	5040 (94%)	0.984
Two-part	$k > 0$	5040 (94%)	0.987
One-part	$3 \leq k \leq 8$	3925 (73%)	0.985
Two-part	$3 \leq k \leq 8$	3925 (73%)	0.988
One-part	$5 \leq k \leq 7$	1934 (36%)	0.966

As it can be seen from the table, dataset thinning using the proposed automatic out-of-distribution detection method can noticeably enhance the results of the CNN-based classification in the TB diagnostics using pulmonary X-ray images. The best results are obtained when only images with $3 \leq k \leq 8$ range of visible vertebrae are left in dataset and two models are used, each trained for its own level of radiation.

It also should be mentioned that the experiment with two-part model on $5 \leq k \leq 7$ vertebrae range was not performed due to the too small amount of data which would be used for training each of the models. The choice of splitting data in the experiment with $3 \leq k \leq 8$ and two-part model into sets with $k=\{3, 4\}$ and $k=\{5, 6, 7, 8\}$ rather than $k=\{3, 4, 5\}$ and $k=\{6, 7, 8\}$ is also explained by the distribution of the used X-ray images. In the second way of splitting the ratio of images used to train two models would be too unbalanced.

V. IMPLEMENTATION DETAILS

The experiments were performed on a PC with Intel Core i7-9750H 2.60 GHz CPU, 16 GB memory, and NVIDIA GeForce RTX 2060 6 GB GPU. The networks for classification and spine area detection were implemented based on pytorch 1.8.1 framework in Python 3. Pillow 8.2.0 was used for image processing tasks. We also used matplotlib and scipy 1.6.3. To create data annotation for spine area detection Amazon SageMaker Ground Truth service was used.

The time of processing of one typical pulmonary X-Ray image for vertebrae detection and counting is 2-3 seconds. Despite the fact that the proposed algorithm will take a lot of time to process a large dataset, the operation on the whole dataset should be performed only once and it still takes much less time than training the model. For all newly received images, the achieved processing time of 2-3 seconds per image is more than satisfactory for medical examinations and screening.

VI. CONCLUSION

In this paper, we demonstrated the importance of using medical differential criteria to determine the level of radiation in X-ray images of the lungs and showed its applicability for the task of tuberculosis detection with deep neural networks. We developed a new method for automatic determination and calculation the number of visible vertebrae in the pulmonary X-ray images and proposed a system of automatic out-of-distribution detection for pulmonary X-ray image analysis.

Using the proposed system of out-of-distribution detection allows to enhance the tuberculosis classification results up to 1.3% within the same classification model. Here we calculate the growth of the obtained balanced accuracy with the proposed dataset thinning technique and using two-part model compared to the balanced accuracy obtained for classification with one-part model without dataset thinning. The proposed system allows to reduce the share of incorrect classifications by 2 times: from 2.5% to 1.2%. It also allows to automatically train a composite model considering X-ray radiation level, which is more effective compared to the traditional CNN models.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Andrey A. Dovganich and Yakov V. Pchelitsev conducted the research; Alexander V. Khvostikov, Andrey S. Krylov, Yong Ding, and Mylene C. Q. Farias revised and edited the paper together; all authors had approved the final version.

AKNOWLADGMENT

The work was funded by RFBR, CNPq and MOST according to the research project 19-57-80014 (BRICS2019-394).

REFERENCES

[1] I. Fantini, C. Yasuda, M. Bento, L. Rittner, F. Cendes, and R. Lotufo, "Automatic MR image quality evaluation using a deep CNN: A reference-free method to rate motion artifacts in neuroimaging," *Computerized Medical Imaging and Graphics*, vol. 90, p. 101897, 2021.

[2] Q. Zhang, E. Hann, K. Werys, C. Wu, I. Popescu, E. Lukaschuk, and S. K. Piechnik, "Deep learning with attention supervision for automated motion artefact detection in quality control of cardiac T1-mapping," *Artificial Intelligence in Medicine*, vol. 110, p. 101955, 2020.

[3] K. Nousiainen, T. Mäkelä, A. Piilonen, and J. I. Peltonen, "Automating chest radiograph imaging quality control," *Physica Medica*, vol. 83, pp. 138-145, 2021.

[4] Y. Ding, *Visual Quality Assessment for Natural and Medical Image*, Zhejiang University Press and Springer-Verlag GmbH Germany, 2018.

[5] R. O. Panicker, B. Soman, G. Saini, and J. Rajan, "A review of automatic methods based on image processing techniques for tuberculosis detection from microscopic sputum smear images," *Journal of Medical Systems*, vol. 40, no. 1, pp. 1-13, 2016.

[6] R. D. J. Samuel and B. R. Kanna, "Tuberculosis (TB) detection system using deep neural networks," *Neural Computing and Applications*, vol. 31, no. 5, pp. 1533-1545, 2019.

[7] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, pp. 1-14, 2021.

[8] M. E. Chowdhury, *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665-132676, 2020.

[9] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635-640, 2020.

[10] G. A. Chuiko and V. M. Tsvetkov, "Effects of x-ray hardness on fluorogram informativeness," *Biomedical Engineering*, vol. 16, no. 4, pp. 117-119, 1982.

[11] H. W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *Proc. IEEE International Conference on Image Processing*, 2014, pp. 343-347.

[12] Y. Hu, *et al.*, "A statistical training data cleaning strategy for the PCA-based chiller sensor fault detection, diagnosis and data reconstruction method," *Energy and Buildings*, vol. 112, pp. 270-278, 2016.

[13] S. Jaeger, *et al.*, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, p. 475, 2014.

[14] S. Jaeger, *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, p. 233, 2014.

[15] S. Candemir, *et al.*, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, p. 577, 2014.

[16] Y. Liu, Y. H. Wu, Y. Ban, H. Wang, and M. M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2646-2655.

[17] K He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[18] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700-4708.

[19] J. Deng, *et al.*, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255.

Andrey A. Dovganich received the B.S. degree and M.S. degree in applied mathematics and informatics from Lomonosov Moscow State University, Russia, in 2015 and 2017. Since then he studies in the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University as a PhD student in the Laboratory of Mathematical Methods of Image Processing. His research interests are image processing and analysis, image quality assessment, machine learning, deep learning, convolutional neural networks.

Alexander V. Khvostikov received the B.S. degree and Ph.D. degree in applied mathematics from Lomonosov Moscow State University, Russia, in 2015 and 2019. Since then he works in the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University as a researcher in the Laboratory of Mathematical Methods of Image Processing. In 2020 he won the prize of the Lomonosov Moscow State University for young teachers and researchers who have achieved significant results in teaching and research activities. His research interests are image processing and analysis, computer vision, medical images, machine learning, deep learning, convolutional neural networks, hybrid methods.

Yakov A. Pchelintsev received the B.S. degree and M.S. degree in applied mathematics and informatics from Lomonosov Moscow State University, Russia, in 2017 and 2019. Since then he studies in the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University as a PhD student in the Laboratory of Mathematical Methods of Image Processing. His research interests are image processing and analysis, machine learning, deep learning, convolutional neural networks.

Andrey S. Krylov received the B.S. degree and Ph.D. degree in applied mathematics from Lomonosov Moscow State University, Russia, in 1978 and 1983, respectively. Since then he has worked in the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University. Now he is a full professor, head of the Laboratory of Mathematical Methods of Image Processing. His research interests include mathematical methods of image processing and computer vision. Editor-in-Chief of Springer journal "Computational Mathematics and Modeling" In 2016-2021 Member of Conference Programme Committees of more than 35 international conferences Science and Technics Leninskii komsomol Price 1989 Member IEEE from 2008. From 2020 he is Editor-in-Chief of Springer journal "Computational Mathematics and Modeling". In 2017-2021 Member of Conference Programme Committees of more than 35 international conferences. He is a member of IEEE and ACM.

Yong Ding received the B.S. degree and M.S. degree from School of Electronic Science & Applied Physics, Hefei University of Technology, Hefei, P.R. China in 1997 and 2000, respectively. Then, he received the Ph.D. degree from College of Electronic Science & Engineering, Nanjing, P.R. China in 2008. From 2000 to 2006, he was a Senior Engineer in R&D Center of Hisense. From 2006 to 2008, he was a Senior Project Leader of Architecture Design Department in Ominivision. Then he joined the faculty of Zhejiang University in 2009 as a full professor. His research interests concentrate on image objective quality assessment, digital video processing and associated SoC architectures. Up to now, he has authored more than 50 papers at Journals in these fields of research. And, he has made several plenary or invited talks on international conferences.

Mylene C. Q. Farias received her B.Sc. degree in electrical engineering from Federal University of Pernambuco, Brazil, in 1995 and her M.Sc. degree in electrical engineering from the State University of Campinas, Brazil, in 1998. She received her Ph.D. in electrical and computer engineering from the University of California Santa Barbara, USA, in 2004 for work in no-reference video quality metrics. Dr. Farias has worked as a research engineer at CPqD (Brazil), Philips Research Laboratories (The Netherlands) and Intel Corporation (Phoenix, USA) for video quality assessment tasks. Currently, she is an Associate professor in the Department of Electrical Engineering at the University of Brasilia. Her current interests include video quality metrics, video processing, multimedia signal processing, watermarking, and visual attention. Dr. Farias is a member of IEEE, the IEEE Signal Processing Society, ACM, and SPIE.