An Ensemble of Triplet Neural Networks for Differential Diagnostics of Lung Cancer

Lev Utkin, Anna Meldo, Maxim Kovalev, Ernest Kasimov
Peter the Great St.Petersburg Polytechnic University (SPbPU)
St.Petersburg, Russia
lev.utkin@gmail.com, anna.meldo@yandex.ru
maxkovalev03@gmail.com, kasimov.ernest@gmail.com

***Abstract*—A new classification subsystem of a lung cancer computer-aided-diagnosis systems is proposed in the paper. Its implementation is based on two main approaches. First, the computed tomography images of segmented suspicious lung nodules are represented by means of five histograms characterizing the shape, inner and outer structures of nodules. This representation significantly reduces the dimensionality of data. Second, an ensemble of triplet neural networks is used to take into account atypical cases of lung cancer and to improve accuracy of the classification subsystem usage. An architecture of the developed triplet network and peculiarities of the triplet network ensemble training process are considered in detail. The corresponding results of numerical experiments with using public dataset LUNA16 show outperforming properties of the proposed classification subsystem.**

I. Introduction

One of the common malignancy tumors in many countries is lung cancer [1]. As a result, many computer-aided-diagnosis (CAD) systems, which automatically detect and classify the suspicious pulmonary nodules or lung lesions, have been developed in last time to assist radiologists in the screening process and in confirming a certain diagnosis [2]. It should be noted that one of the most important modalities for lung cancer diagnosis is the computed tomography (CT) which has a lot of advantages in comparison with other modalities [3]. The main symptom of lung cancer at early stage is the pulmonary nodule which is detected by means of CT. Therefore, CAD systems aim to detect all nodules and to classify them in order to correctly diagnose a disease.

Nowadays, a lot of CAD systems solving the lung cancer diagnosis task consist of two subsystems. The first subsystem called as the CADe system aims to segment and detect lung nodules. The segmentation stage (the CADe system) delineates the lung parenchyma and then detects possible nodules on the lung parenchyma. The second subsystem called as the CADx system is concentrated on a special feature representation of the detected suspicious lung nodules and on their classification as malignant or benign in order to reduce false positive cases. Most efficient lung cancer CAD systems in the recent years are based on using machine learning algorithms, in particular, deep learning ones, due to the fact that the deep learning models achieve better results in many areas such as segmentation and classification tasks compared with other traditional computing techniques. Analyses and comparison of the lung cancer CAD system architectures are available in [4], [5], [6]. To the best of our knowledge, the most efficient lung cancer CAD system nowadays is proposed by Ardilla et al. [7]. Approaches using deep learning methods to classification of lung nodules dominate at present. However, a comprehensive review of the lung nodule classification methods in the CADx systems [8] has illustrated that an important remaining question in the classification implementation is how to represent a segmented nodule image by new features in order to reduce the number of false positive cases and to use simple machine learning models for classification.

In spite of the permanent accumulating of training data, one of the problems of CAD systems, especially CADx systems, is a lack of sufficient data concerning with many atypical cases of the lung disease. For example, one may meet tuberculosis looking like a cancer on CT images, which does not belong to cancer disease. Moreover, some diseases do not have the nodular structure, for example, lymphoma, hamartomas, lepidic adenocarcinoma [9]. As a result, it is very difficult to correctly classify the lung cancer in these cases because the number of examples with these atypical cases are usually small. It should be noted that exactly the CADx system plays a crucial role in classifying atypical cases.

Two ways for simplifying the above problem can be pointed out. The first way is to reduce the dimensionality of lung nodule CT image description after segmentation (after the CADe system) and to apply some special low-dimensional feature representation. One of the adequate feature representation of lung nodules is a set of the following five histograms [10] which characterize the surface, the inner structure and surroundings of every lung nodule. These are two inner and outer chord length histograms based on the chord method [11], two inner and outer radiodensity histograms, and an inner histogram of the radiodensity changes.

It should be noted that the above dimensionality reduction can be conditionally viewed as one of the radiomics approaches which are based on radiological quantitative image features. Radiomics approaches [12], [13] try to build the prediction model based on the extracted two dimensional (2D) or three dimensional (3D) radiological quantitative image features of lung nodules based on prior knowledge of what fea-

ISSN 2305-7254

tures and characteristics are significant. However, in contrast to the radiomics approaches, we do not use prior knowledge about significance of features because this information is very often absent. We try to propose an adequate and simple feature representation such that the significance of features is implicitly determined by a classifier.

The aforementioned histograms allow us to significantly reduce the nodule CT image representation dimensionality. However, the problem of atypical cases cannot be solved only by applying the first way. Therefore, the second way is to consider the few-short learning [14] which is a framework for dealing with the small number of training examples in some classes. There are several methods for solving the few-short learning problem. One of them is to use the so-called Siamese neural network [15], [16]. The Siamese neural network is composed of two identical neural networks with shared parameters. It aims to compare pairs of examples and to make decision about semantic similarity or dissimilarity of examples in every pair. However, our experiments have shown that the Siamese neural network does not provide outperforming results due to the small amount of training data.

Another architecture which can be regarded as an extension of the Siamese neural network is the Triplet neural network (TrNN) [17]. Sometimes, it is viewed as a neural network with the triplet loss (see, for example, [18]). The TrNN is comprised of three identical neural networks with shared parameters. If we denote three input examples for the networks as X^-, X^+, X, then the TrNN encodes the pair of distances between each of X^- and X^+ against the reference X. In other words, the vector X represents the anchor example, X^+ represents a positive example and X^- represents a negative example. The dissimilarity between X and X^+ should be less than the dissimilarity between X and X^-. It is shown in the paper that the best choice of a loss function is the so-called lossless triplet loss [19].

The TrNN improved the results of the lung cancer classification, but this improvement was negligible. Therefore, we propose to apply an ensemble of the TrNNs. Ensemble-based methods can be viewed as techniques that develop a group of classifiers and then combine them to produce improved results. They often produce more accurate solutions in comparison with single models used separately [20], [21], [22]. One of the groups of ensemble-based methods is bagging [23]. The well-known bagging model is the random forest [24], which uses a large number of randomly built individual decision trees in order to combine their predictions. Generally, bagging methods generate an ensemble of independent models in which each model is trained using a sample of examples taken from the original dataset as a replacement [21]. Majority voting of predictions provided by every model is performed to compute the final prediction of a new example.

The proposed ensemble of TrNNs as a part of the lung cancer CADx system surprisingly have demonstrated very promising results. Therefore, the paper aims to propose an architecture of the developed TrNN and to provide the corresponding results of numerical experiments which show nice properties of the TrNN ensemble. In order to train and to test the TrNN, we use the well-known dataset LUNA16 (https://luna16.grand-challenge.org/) which contains respective CT images in DICOM (Digital Imaging and Communication in Medicine) format.

The CADe system is not studied in the paper because it may have various implementations. In particular, it can be implemented based on the well-known conventional image processing methods, including shape based methods, edge based methods, morphological methods. These methods cannot be called intelligent. In contrast to the conventional image processing methods, deep learning approaches are successfully used for segmentation recently. Since we assume that the segmented nodules are represented in the form of the histograms, the CADx system does not depend on the CADe system implementation. Various image segmentation approaches to detect lung nodules can be found, for example, in survey papers [5], [6], [25].

A general architecture of the considered CAD system is shown in Fig. 1. The CADe system produces CT images of suspicious lung nodules. The CADx system transforms the images into histograms which are classified by the ensemble of TrNNs. The final decision about a disease is based on the separate TrNN results, and it is computed by using the standard majority voting scheme.

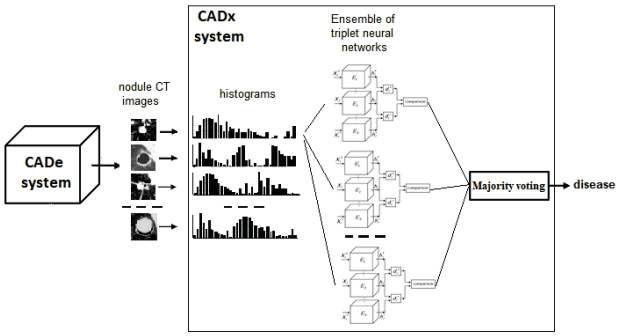

Fig. 1. A general architecture of the CAD system

The rest of the paper is organized as follows. In Section 2, a brief description of the chord method and the histogram representation of the lung nodule CT images is given. TrNNs are explained in Section 3. Peculiarities of an ensemble of TrNNs and an architecture of neural networks which form the TrNN for the histogram classification are considered in Section 4. Numerical examples illustrating the proposed ensemble of TrNNs are provided in Section 5. Section 6 contains some concluding remarks.

II. The chord method and histograms in CADx

In order to reduce the dimensionality of the segmented lung nodule images and to simplify their classification, a specific feature representation of the images is proposed. The feature representation uses the idea of the chord method [11] which is based on constructing many line segments that are called by

chords, and they link pairs of points on the shape boundary. The set of normalized chord lengths can be viewed as a chord length histogram which is invariant to size, translation and rotation of the shape [11]. Therefore, the first histogram representing the lung nodule CT image is the chord length histogram, and it characterizes the nodule shape or surface. It is important to note that the chord length histograms of malignant nodules and benign nodules or vessels differ from each other. This fact allows us to classify histograms which are of a low dimensionality avoiding the high dimensionality of images.

However, in order to improve the classification accuracy, four additional histograms are proposed. In particular, the second histogram characterizes the inner structure of every lung nodule, and it is constructed by randomly selecting points on every chord and by determining the radiodensity (Hounsfield units) at these points. As a result, we can get the histogram of the radiodensity values inside the considered nodule (the inner histogram).

The third histogram again characterizes the nodule shape. To get this histogram, it is proposed to put the lung nodule into a virtual cube which can be viewed as the environment around the nodule. Then the histogram is constructed by drawing chords connecting points at the virtual cube around the nodule and at the nodule surface. Chords are constructed as perpendiculars to every face of the cube. This histogram can be viewed as the outer histogram.

In order to take into account the environment around a nodule, values of the radiodensity are measured at points on chords constructed between the above cube and the nodule surface. This histogram is also outer, and it characterizes the nodule surrounding.

In order to take into account changes of the radiodensity inside the nodule, the fourth histogram is constructed by computing differences between the radiodensity values determined at neighboring points (at two neighboring spheres of different radius inside the nodule). Spheres of different increasing radii are constructed from the approximate center of the nodule, so that the last sphere covers the most distant points of the nodule surface.

Finally, the concatenated five histograms form a histogram as a new feature representation of every lung nodule CT image. The summed number of bins can be regarded as the number of new features.

The next problem is to classify these histograms. However, the main difficulty in training a classifier in the CADx system is that the training set contains only a few examples of many atypical lung cancer cases. In order to cope with this difficulty, we propose to apply an ensemble of the TrNN to classify the suspicious lung nodules.

III. Triplet neural networks

Let $\{(X_i, y_i), i = 1, ..., N\}$ be a dataset of feature vectors $X_i \in \mathbb{R}^m$ (histograms of suspicious nodules) of size m with labels y_i such that every label corresponds to a certain diagnosis. For simplicity, we consider a special case when

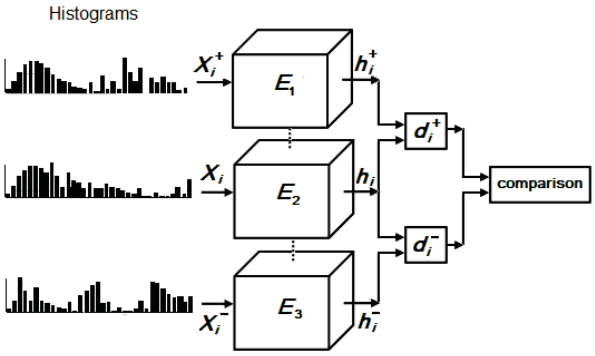

Fig. 2. A general structure of the TrNN, where E_1, E_2, E_3 are three identical neural networks with shared weights

y_i corresponds to two outcomes: malignant and benign, i.e., $y_i \in \{0, 1\}$. The value m is the total number of bins in five histograms.

It has been mentioned that the TrNN consists of three identical neural networks with shared parameters. It is trained from triplets $\langle X_i, X_i^+, X_i^- \rangle$ such that every triplet is constructed from the pairwise similarity data as follows: for each *anchor* example X_i, a *positive* example X_i^+ is selected (X_i and X_i^+ are semantically similar or from the same class), and a *negative* example X_i^- is selected (X_i and X_i^- are semantically dissimilar or from different classes). In fact, X_i and X_i^+ are different viewpoints of the same disease, and X_i^- comes from a different disease.

A general structure of the TrNN is shown in Fig. 2, where three neural networks are denoted as E_1, E_2, E_3. The TrNN maps the triplet $\langle X_i, X_i^+, X_i^- \rangle$ into a learned feature space with a new triplet (a new feature representation) $\langle h_i, h_i^+, h_i^- \rangle$. Since the anchor example X_i is semantically closer to all positive examples X_i^+ than to all negative examples X_i^-, then the relative similarity and dissimilarity between examples in the triplet $\langle X_i, X_i^+, X_i^- \rangle$ are measured by the Euclidean distances between the corresponding vectors in the triplet $\langle h_i, h_i^+, h_i^- \rangle$. In other words, training the TrNN brings X_i and X_i^+ close in the feature space, and pushes X_i and X_i^- far apart.

Let us denote the Euclidean distance between vectors h_i and h_i^- as $d_i^- = \left\| h_i - h_i^- \right\|_2^2$, and the Euclidean distance between vectors h_i and h_i^+ as $d_i^+ = \left\| h_i - h_i^- \right\|_2^2$. Then one of the most popular triplet loss functions for training the TrNN called Margin Ranking Loss is of the form:

$$L = \sum_{i=1}^{M} \max\left(0, \delta - d_i^- + d_i^+\right). \quad (1)$$

Here δ is a margin (a tuning parameter) that is enforced between positive and negative pairs, M is the number of all possible triplets in the training set. By using the above triplet loss function, we enforce the TrNN to adjust its weights to achieve the condition $d_i^- > \delta + d_i^+$. In fact, the block "comparison" in Fig. 2 performs comparison of distances d_i^-

and d_i^+ in order to determine whether they fulfil the above condition for distances or not.

Another loss function called Ratio Loss was proposed by Hoffer and Ailon [17]. The Ration Loss uses the SoftMax function as follows:

$$L = \sum_{i=1}^{M} \left\| q_i^+, q_i^- - 1 \right\|_2^2, \tag{2}$$

where

$$q_i^+ = \frac{\exp(d_i^+)}{\exp(d_i^+) + \exp(d_i^-)}, \; q_i^- = 1 - q_i^+. \tag{3}$$

Several different triplet loss functions have been proposed in literature, for example, a triplet-center loss [26], a batch-hard based triplet loss [27]. Arsenault in his blog proposes to use the so-called lossless triplet loss function [19], which is of the form:

$$L = \sum_{i=1}^{M} \left(-\ln \; -\frac{d_i^+}{\beta} + 1 + \varepsilon \right) - \ln \; -\frac{M - d_i^-}{\beta} + 1 + \varepsilon \Big) \Big). \tag{4}$$

Here β is a scaling factor, ε is a parameter avoiding the case $\ln(0)$. It turns out that this loss function provides the best accuracy results for the lung cancer classification when the histogram representation is used for suspicious lung nodule images.

IV. An ensemble of triplet neural networks

In order to classify histograms representing the suspicious lung nodule CT images, we propose an architecture of every neural network being one of three components of the TrNN. It is shown in Table I. One can see from Table I that the number of units in the input layer corresponds to the number of bins in every histogram consisting of five small histograms ($m = 108$). The output layer contains only two units. Activation functions for all layers are PReLU (a parametric ReLU [28] of the form $f(x) = \max(0, x) + a \min(0, y)$) except for the output layer where the sigmoid activation function is used. A decision about a class of the analyzed lung nodule is carried out by looking for k nearest neighbors in the feature space, i.e., by measuring the Euclidean distances between the corresponding vectors in the triplet $\langle h, h^+, h^- \rangle$. The preliminary comparison analysis of different loss functions has shown that the TrNN with the lossless triplet loss provides the best accuracy results. Therefore, we use this loss function for training the proposed TrNN.

It turns out that a single TrNN provides unsatisfactory results. Therefore, an ensemble of TrNNs is proposed. In order to train the ensemble, a procedure for a random selection of triplets having the structure $\langle X_i^-, X_i^+, X_i \rangle$ for every TrNN should be implemented. The following procedure for preparing a set of triplets $\mathcal{T}_t$ for training the t-th TrNN is proposed.

1) n random examples $\{c_1, ..., c_n\}$ are selected with replacement from the class "malignant".

TABLE I. The architecture of neural networks from the TrNN

Layer	Dimension	Activation
Input	108	PReLU
Hidden 1	36	PReLU
Hidden 2	12	PReLU
Hidden 3	6	PReLU
Hidden 4	3	PReLU
Output	2	Sigmoid

2) n random examples $\{b_1, ..., b_n\}$ are selected without replacement from the class "benign".
3) For every value c_i, $i = 1, ..., n$, a set of pairs $\langle c_i, c_j \rangle$, $j = 1, ..., n$, is constructed.
4) q examples are randomly selected from the set of selected examples $\{b_1, ..., b_n\}$ and form a set $\{b_1^*, ..., b_q^*\}$.
5) The set of triplets $\mathcal{T}_t$ consists of examples $\langle c_i, c_j, b_l^* \rangle$, $l = 1, ..., q$. The total number of obtained triplets for training a single TrNN is $S = 2qn(n-1)$.

The above procedure is repeated T times to get T sets of triplets $\mathcal{T}_1, ..., \mathcal{T}_T$ for all TrNNs in the ensemble. The values n and q can be viewed as tuning parameters.

Suppose that the t-th TrNN in the ensemble in the testing phase returns an output $y_t^* \in \{0, 1\}$ computed in accordance with k nearest neighbors such that its values correspond to "benign" and "malignant", respectively. Then the ensemble output can be computed by averaging the TrNN outputs, i.e., it is computed as

$$y^* = T^{-1} \sum_{t=1}^{T} y_t^*. \tag{5}$$

V. Numerical experiments

In order to illustrate the proposed ensemble of TrNNs for classifying histograms representing the suspicious lung nodules, the following four accuracy measures are analyzed: Accuracy (the ratio of correctly predicted observation to the total observations), Precision (the ratio of correctly predicted positive observations to the total predicted positive observations), Recall (the ratio of correctly predicted positive observations to all observations in actual class), F1-score (the weighted average of Precision and Recall). Dataset LUNA16 is used for training the total CAD system. The dataset consists of a total of 888 chest CT scans. Results of segmentation are 25 098 histograms such that 24670 and 428 histograms correspond to benign and malignant cases, respectively. All histograms are characterized by the numbers a_i, $i = 1, ..., 5$, of bins (each bin represents a predefined interval of lengths (a_1,a_3), radiodensity values (a_2,a_4), radiodensity value differences (a_5). Numbers of bins in the histograms are tuning parameters. However, we take for following parameters of histograms: $a_1 = a_4 = 24$, $a_2 = 40$, $a_3 = a_5 = 10$. As a result, the total number of features for classification is $m = 108$. In order to perform numerical experiments with balanced datasets for testing, the number of testing examples is 500: 250 benign cases and 250 malignant cases.

TABLE II. ACCURACY MEASURES AS FUNCTIONS OF THE NUMBER k OF NEAREST NEIGHBORS

k	Accuracy	Precision	Recall	F1-score
1	0.916	0.926	0.904	0.915
3	0.916	0.922	0.908	0.915
5	0.918	0.926	0.908	0.917
7	0.918	0.926	0.908	0.917
9	0.918	0.926	0.908	0.917

Parameters of the TrNN training are $n = 600$ (the number of selected examples from the class "benign"), $q = 10$ (the number of selected examples from the restricted class "benign"), $T = 80$ (the largest number of TrNNs in the ensemble). The learning rate is 10^{-3}, the number of epochs is 1000, the batch size is 5000.

To evaluate the average accuracy measures, we perform experiments with 100 repetitions. We also study how numbers of TrNNs in the ensemble impact on the accuracy measures and show results corresponding to the "optimal" number of TrNNs from the accuracy point of view. It has been mentioned that the lossless triplet loss gives the largest accuracy. Therefore, this loss function is used in all numerical experiments with parameters $\beta = 2$ and $\varepsilon = 0.001$.

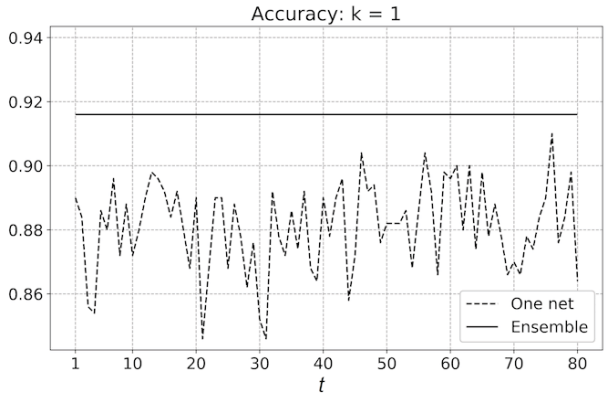

Fig. 3. Accuracy as a function of the TrNN number t and of the ensemble

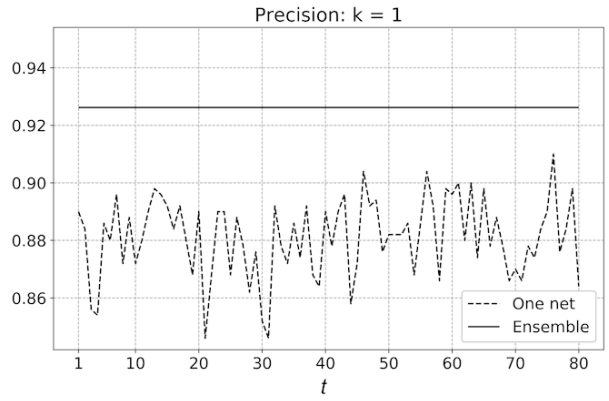

Fig. 4. Precision as a function of the TrNN number t and of the ensemble

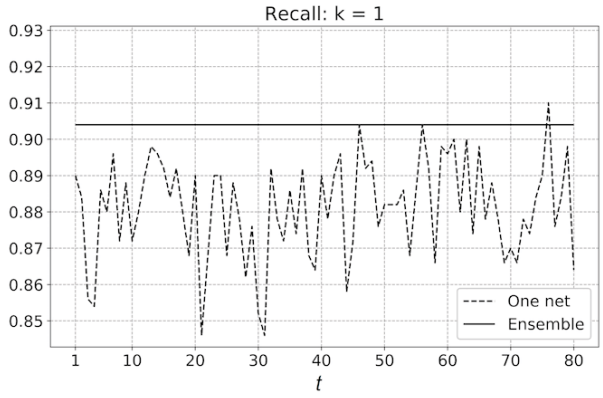

Fig. 5. Recall as a function of the TrNN number t and of the ensemble

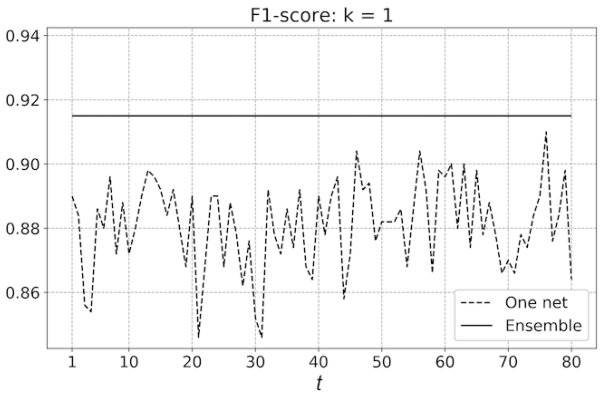

Fig. 6. F1-score as a function of the TrNN number t and of the ensemble

In order to illustrate the difference between accuracy measures of separate TrNNs and the ensemble of TrNNs, we show in Figs. 3-6 dependencies of accuracy measures (Accuracy, Precision, Recall, F1-score) on every TrNN and on the ensemble. The accuracy measures of the ensemble are depicted by solid lines whereas the same measures of separate TrNNs are depicted by dashed lines. The dashed lines have a jumping behavior because every single TrNN is trained by using randomly selected training data and it has its inherent classification ability. It can be seen from Figs. 3-6 that separate TrNNs are mainly inferior to the ensemble. This is the main reason for implementing the idea to use the ensemble of TrNNs to improve classification accuracy measures.

In order to study how the number T of TrNNs in the ensemble impacts on the classification performance, we consider dependencies of accuracy measures (Accuracy, Precision, Recall, F1-score) on T. These dependencies are shown in Figs. 7-10, respectively. It can be seen from Figs. 7-10 that all accuracy measures are increased with increase of T. At the same time, it is interesting to observe some interval of T where every accuracy measure is not strongly changed. For example, this interval for the Accuracy is from $T = 15$ till $T = 65$.

The next question is how the parameter k of nearest neigh-

bors for decision making impact on the accuracy measures. The corresponding results are shown in Table II. It can be seen from Table II that the increase of accuracy measures is observed only for $k \leq 5$. Values of k larger than 5 do not change the accuracy measures. This implies that there exists some optimal smallest k which provide the largest values of accuracy measures.

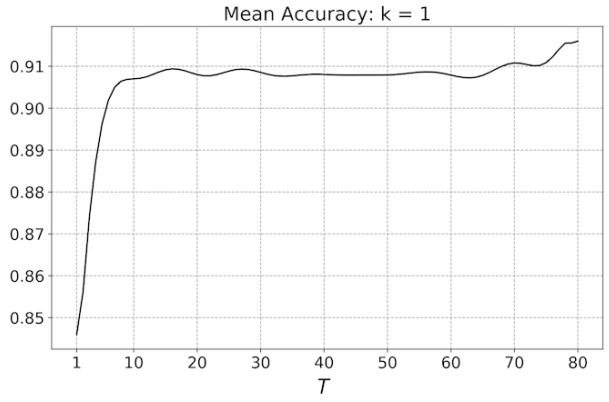

Fig. 7. Accuracy as a function of the number T of TrNNs in the ensemble

Finally, we compare the proposed ensemble of TrNNs with the random forest which is also based on the bagging method and is one of the best classifiers. Moreover, we use an ensemble of random forests consisting of two standard random forests and two completely-random tree forests [29]. The corresponding results are given in Table III. One can clearly see from Table III that the ensemble of TrNNs significantly outperforms the ensemble of random forests. Only the precision measure achieves the highest value for the ensemble of random forests.

VI. Conclusion

A new approach for implementing the lung cancer CADx system based on applying the ensemble of TrNNs has been considered in the paper. The main aim of the approach was to take into account atypical cases of the suspicious lung nodules.

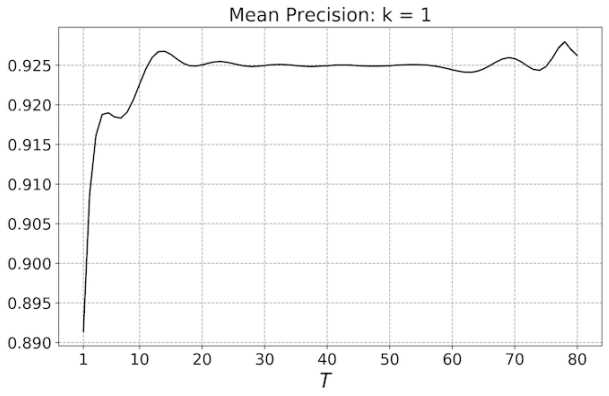

Fig. 8. Precision as a function of the number T of TrNNs in the ensemble

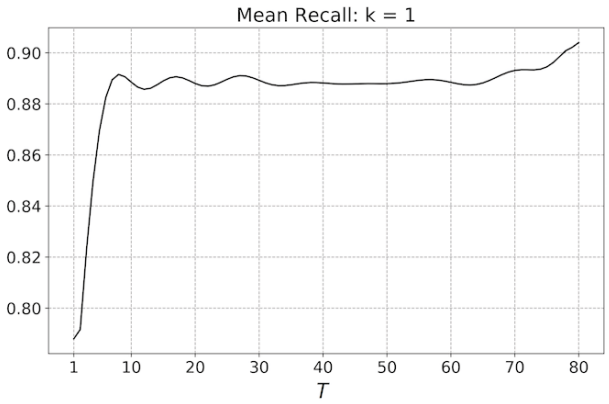

Fig. 9. Recall as a function of the number T of TrNNs in the ensemble

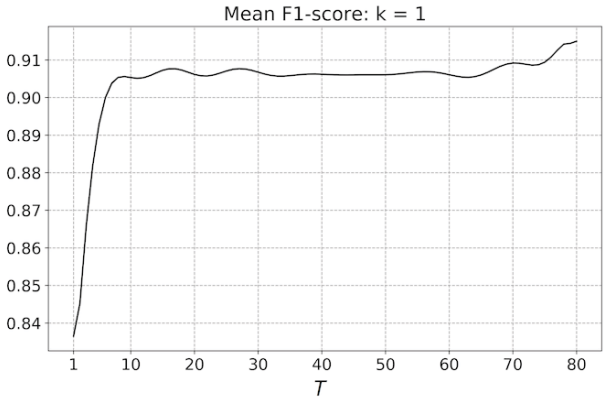

Fig. 10. F1-score as a function of the number T of TrNNs in the ensemble

The numerical results show that the proposed ensemble of TrNNs jointly with the histogram feature representation provide outperforming accuracy results in comparison with use of the single TrNN.

Three basic ideas were used for implementing the CADx system, which distinguish the system from similar CAD systems. First, every lung nodule is represented by means of five histograms characterizing its shape, inner and outer structures. Second, the TrNN is used for taking into account atypical cases of lung cancer. Third, the ensemble of TrNNs is used to improve the accuracy measures of the CADx system.

Main numerical results are obtained for the histogram representation of lung nodules. Moreover, the architecture of networks in every TrNN is adapted to this representation. However, the approach is general and can be applied to various feature representations. In particular, it is interesting to study

TABLE III
Comparison of the random forest and the ensemble of TrNNs

	Accuracy	Precision	Recall	F1-score
Ensemble of random forests	0.638	1.000	0.276	0.437
Ensemble of TrNNs	0.918	0.926	0.908	0.917

the ensemble of TrNNs consisting of 3D convolutional neural networks for processing parts of CT images in their initial 3D form. This is a direction for further research.

But the most interesting direction for further research and for improving the classification accuracy is a composition of ensemble-based TrNN classifiers which are trained by using different feature representations, for example, histograms and 3D patches of different sizes. This composition can be controlled by a meta-learner in order to enhance the accuracy. The obtained architecture can be viewed as an ensemble of ensembles.

We have considered the ensemble of TrNNs as a simple bagging model. However, it is interesting to investigate other ensemble-based models including the boosting and the stacking. Moreover, the proposed bagging model can be also improved by means of introducing some weights assigned to every TrNN such that the simple averaging of the TrNN outputs is replaced with the weighted averaging. This is another direction for further research.

Acknowledgement

The reported study was funded by RFBR, project number 19-29-01004.

The results of the work were obtained using computational resources of Peter the Great Saint-Petersburg Polytechnic University Supercomputing Center (www.spbstu.ru) which is registered as a center of collective usage (http://ckp-rf.ru/ckp/500675/).

References

[1] R. Siegel, K. Miller, and A. Jemal, "Cancer statistics, 2018," *CA Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, 2018.

[2] J. Zhang, Y. Xia, H. Cui, and Y. Zhang, "Pulmonary nodule detection in medical images: A survey," *Biomedical Signal Processing and Control*, vol. 43, pp. 138–147, 2018.

[3] G. Zhang, S. Jiang, Z. Yang, L. Gong, X. Ma, Z. Zhou, C. Bao, and Q. Liu, "Automatic nodule detection for lung cancer in ct images: A review," *Computers in Biology and Medicine*, vol. 103, pp. 287–300, 2018.

[4] P. Afshar, A. Mohammadi, K. Plataniotis, A. Oikonomou, and H. Benali, "From hand-crafted to deep learning-based cancer radiomics: Challenges and opportunities," Aug 2018, arXiv:1808.07954v1.

[5] V. Cheplygina, M. de Bruijne, and J. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," Apr 2018, arXiv:1804.06353.

[6] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[7] D. Ardila, A. Kiraly, S. Bharadwaj, B. Choi, J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019.

[8] D. Walawalkar, "A fully automated framework for lung tumour detection, segmentation and analysis," Jan 2018, arXiv:1801.01402.

[9] W. Brant and C. Helms, *Fundamentals of Diagnostic Radiology*, 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2012.

[10] A. Meldo and L. Utkin, "A new approach to differential lung diagnosis with ct scans based on the siamese neural network," in *IOP Conf. Series: Journal of Physics: Conference Series*, vol. 1236, 2019, pp. 012 058–5.

[11] S. Smith and A. Jain, "Chord distribution for shape matching," *Computer vision, graphics, and image processing*, vol. 20, no. 3, pp. 259–271, 1982.

[12] G. Lee, H. L. MD, H. Park, M.L.Schiebler, E. Beek, Y. Ohno, J. Seo, and A. Leung, "Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art," *European Journal of Radiology*, vol. 86, pp. 297–307, 2017.

[13] R. Thawani, M. McLane, N. Beig, S. Ghose, P. Prasanna, V. Velcheti, and A. Madabhushi, "Radiomics and radiogenomics in lung cancer: A review for the clinician," *Lung Cancer*, vol. 115, pp. 34–41, 2018.

[14] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[15] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 539–546.

[16] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, Lille, France, 2015, pp. 1–8.

[17] E. Hoffer and N. Ailon, "Deep metric learning using Triplet network," Dec 2018, arXiv:1412.6622v4.

[18] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Jun 2015, arXiv:1503.03832v3.

[19] M.-O. Arsenault. (2018, Feb) Lossless triplet loss. [Online]. Available: http://towardsdatascience.com/lossless-triplet-loss-7e932f990b24

[20] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New Jersey: Wiley-Interscience, 2004.

[21] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1–39, 2010.

[22] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton: CRC Press, 2012.

[23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[24] ——, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] M. Firmino, A. Morais, R. Mendoca, M. Dantas, H. Hekis, and R. Valentim, "Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects," *Biomedical engineering online*, vol. 13, no. 1, p. 41, 2014.

[26] X. He, Y. Zhou, Z. Zhou, S.Bai, and X. Bai, "Triplet-center loss for multi-view 3d object retrieval," Mar 2018, arXiv:1803.06189v1.

[27] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," Nov 2017, arXiv:1703.07737.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," Feb 2015, arXiv:1502.01852.

[29] F. Liu, K. Ting, Y. Yu, and Z.-H. Zhou, "Spectrum of variable-random trees," *Journal of Artificial Intelligence Research*, vol. 32, pp. 355–384, 2008.