2018 IEEE 31st International Symposium on Computer-Based Medical Systems

Lung Nodule Classification via Deep Transfer Learning in CT Lung Images

Raul Victor Medeiros da Nóbrega*, Solon Alves Peixoto†, Suane Pires P. da Silva‡, and Pedro Pedrosa Rebouças Filho§
Programa de Pós-Graduação em Ciência da Computação (PPGCC),
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Brazil
Email: *raulmedeiros@lapisco.ifce.edu.br, † solonalves@lapisco.ifce.edu.br,
‡ suanepires@lapisco.ifce.edu.br, §pedrosarf@ifce.edu.br

***Abstract*—Lung cancer corresponds to 26% of all deaths due to cancer in 2017, accounting more than 1.5 million deaths globally. Considering this challenging situation, several computer-aided diagnosis systems have been developed to detect lung cancer at early stages, which increases the patients' survival rate. Motivated by the success of deep learning in natural and medical image classification tasks, the proposed approach aims to explore the performance of deep transfer learning for lung nodules malignancy classification. For this, convolutional neural networks (CNN), such as VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile and NASNetLarge, were used as features extractors to process the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI). Next, the deep features returned were classified using Naive Bayes, MultiLayer Perceptron (MLP), Support Vector Machine (SVM), Near Neighbors (KNN) and Random Forest (RF) classifiers. Additionally, to compare the classifiers performance with themselves and with other ones in literature, the evaluation metrics Accuracy (ACC), Area Under the Curve (AUC), True Positive Rate (TPR), Precision (PPV), and F1-Score were computed. Finally, the best combination of deep extractor and classifier was CNN-ResNet50 with SVM-RBF, which achieved ACC of 88.41% and AUC of 93.19%. These results are equivalent to related works, even just using a CNN pre-trained on non-medical images. For this reason, deep transfer learning proved to be a relevant strategy to extract representative imaging biomarkers for lung nodule malignancy classification in chest CT images.**

***Index Terms*—Lung Nodule Classification, Convolutional Neural Networks, Transfer Learning, Computer-Aided Diagnoses, Computed Tomography.**

I. Introduction

Despite the fact that lung cancer mortality rate reduced 17%, from 2002 to 2014, among females and 43%, from 1990 to 2014, among males [1] lung cancer still accounts for 1.59 million deaths worldwide [2], corresponding to 26% of all deaths due to cancer in 2017 [1]. In addition, since more than 50% of lung cancer patients are diagnosed at a distant stage, for which the five-year survival rate is only 4% [2], lung cancer is considered an aggressive disease [3].

However, to address this issue, pulmonologist and radiologists started to use chest Computed Tomography (CT) images to diagnose lung cancers at an earlier stage [1]. These images have proved to be a valuable source of information, reducing lung cancer mortality by up to 20% in a high-risk group [1], [4].

Moreover, using only chest CT images and performing an automatic assessment of pulmonary nodules modern lung Computer-Aided Diagnosis (CAD) systems can support medical decision-making. The ultimate goal of such systems is to discriminate between malignant and non-malignant nodules [5]. In order to achieve this, several methods have been proposed to find representative imaging biomarkers at a macroscopic level in chest CT images. Overall, these extractors can be divided into two categories: hand-crafted, and data-driven.

Several studies investigating hand-crafted features extractors for lung nodule classification have been carried out. Among them, Histogram of Oriented Gradients (HOG) [3], [5], [6], Gabor [7], [8], Local Binary Pattern (LBP) [3], [5], [7], [8], Gray-Level Co-occurrence Matrix (GLCM) [6]–[10].

Furthermore, the rising quantity and quality of public-available datasets made it possible for data-driven methods, e.g. Convolutional Neural Network (CNN), learn how to extract representative features from raw data [11]. However, to perform a proper training of such deep models a large amount of data it is necessary and this remains a challenge in the medical domain [12], [13]. To overcome this, a strategy called transfer learning suggests that features learnt to solve a specific task can be useful for tasks from other domains [14]–[16].

Motivated by the success of deep transfer learning in computer-aided diagnosis systems [2], [12], [13], [15]–[20], we address the challenge of lung nodules malignancy classification in chest CT images using deep features. Specifically, this work intends to:

1) Extract deep features from Lung Image Database Consortium image collection (LIDC-IDRI) [21] dataset, using 11 different CNNs models pre-trained on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [22] such as: VGG16 [23], VGG19 [23], MobileNet [24], Xception [25], InceptionV3 [26], ResNet50 [27], Inception-ResNet-V2 [28], DenseNet169, DenseNet201 [29] and NASNetMobile and NASNetLarge [30].
2) Apply 5 classification methods, such as: Naive Bayes [31], Multilayer Perceptron (MLP) [32], Support Vector Machine (SVM) [33], Near Neighbours (KNN) [34] and Random Forest (RF) [35], for each extracted features set.
3) Compare the results among themselves and with related works through 5 evaluation metrics: Accuracy (ACC),

2372-9198/18/$31.00 ©2018 IEEE
DOI 10.1109/CBMS.2018.00050

Area Under the Curve (AUC), True Positive Rate (TPR), Precision (PPV) and F1-Score.

The remainder of this paper is organized as follows. In Section II, an overview of deep models as feature extractors and machine learning methods are presented. In Section III, the evaluation experiments are formulated, and their results are reported. In Section IV, the experiments outcomes are explored and a discussion is elaborated. Finally, Section V concludes the paper.

II. DEEP FEATURE EXTRACTION WITH CNNs

Although developed in the late 80's by LeCun [11], CNNs have only become popular in 2012, with the rise of large publicly available databases such as the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [22]. With 1.2 million natural images and more than 1000 classes, ImageNet has become an important benchmark in computer vision techniques [22].

Moreover, since [22] won the ImageNet 2012 challenge, several attempts have been made to surpass their results, among them: VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile and NASNetLarge.

However, the application of these deep learning models is limited in the medical field, as there are not large enough datasets to train such models from scratch [12], [13]. To overcome this issue, an approach called Transfer Learning has been widely investigated for Computer-Aided Diagnosis methods [2], [12], [13], [15]–[20].

The Transfer Learning approach suggests that the transformation of space performed to increase the discriminatory power of a generic dataset, e.g., ImageNet, can be useful to solve other problems, such as the LIDC-IDRI dataset [14]–[16].

In the specific case of CNNs, to use a pre-trained model, two steps must be taken: remove the output layer from the model (last layer) and resize the output of the remaining model to a one-dimensional vector [36], [37]. Thus, the new model only performs a transformation in the input data (extraction of characteristics) and no more a classification.

III. OVERVIEW MACHINE LEARNING TECHNIQUES

Naive-Bayes Classifier consists of a probabilistic method based on Bayes Decision Theory and is used for the categorization of objects according to the individual probability of belonging to a certain class [31]. This technique divides the samples in accord with the value of the posterior probability, estimated on the basis of the values of the conditional densities and the prior probabilities [31]. Naive-Bayes Classifier is characterized by assuming that the attributes (variables) of the pattern are statistically independent [31].

Multi-Layer Perceptron (MLP) represents a set of perceptrons intended for solving non-linearly separable problems, being a generalization of the simple perceptron [32]. Its structure consists of layers of neurons, in which each layer is attached to the next layer. This method starts at the input layer, where the attribute vector is presented to the network [32]. Then pulses are produced which will be propagated by the following layers, called hidden layers. Through this process, a final solution is returned in the output layer [32].

Support Vector Machines are based on the Statistical Learning Theory created by [33]. This technique has as main objective to classify through surfaces that increase the distance between the classes [33]. In essence, SVM was designed to solve binary problems, so this method becomes complicated when considering multiclass problems. Thus approaches such as one-versus-one [38] and one-versus-all [38] are examples of variations of this method for application in multiclass issues.

The k-Nearest Neighbors (kNN) corresponds to a classification algorithm of the supervised type. Basically, the operation of this technique consists of determining the classification label of a pattern based on the k neighboring patterns coming from a training set. The number of neighbors is the parameter in kNN [39].

Random Forest is a classification algorithm based on the decision tree method [40]. This technique has unsupervised learning and aims to create several decision trees employing a subset of randomly chosen attributes from the original set [40]. With the construction of the tree sets, it is feasible to perform the classification of which presents the best knowledge gain for the decision making of a given problem [40].

IV. METHODOLOGY

An overview of the approach explored in this work is presented in Figure 1. First, LIDC/IDRI dataset annotations are processed and pulmonary nodules images are extracted. Later, several CNNs are transformed into features extractors and them applied to the aforementioned images, producing a new dataset. Lastly, several classification methods are applied to this new dataset and evaluation metrics are computed. All of these stages are described in detail in the next subsessions.

A. LIDC/IDRI dataset

Composed of 1018 cases, LIDC/IDRI dataset [21] contains, for each exam, a set of chest CT images and XML file with medical annotations from 4 thoracic radiologists. More specifically, there are 7371 injuries segmented and labeled as nodules by at least one specialist and 2669 of them are also marked as lager than 3 mm.

To extract nodules images from a given LIDC/IDRI exam, the centroids from all sementations annotations are computed. Next, a Euclidean distance matrix is also calculated and the subsets of centroids, that have a distance other nodule smaller than 0.8 times its radius, are merged. This combination process consists of computing the median malignancy and the mean centroid among all close centroid. Then, for each remaining centroids, 3 images of 16x16 voxels were extracted (centered on the nodules centroids), one at each exam plane (Sagittal, Coronal, Axial). Furthermore, images mean intensity value was subtracted from them, the result was normalized between 0 and 1 and resized to fit a target deep model input size.

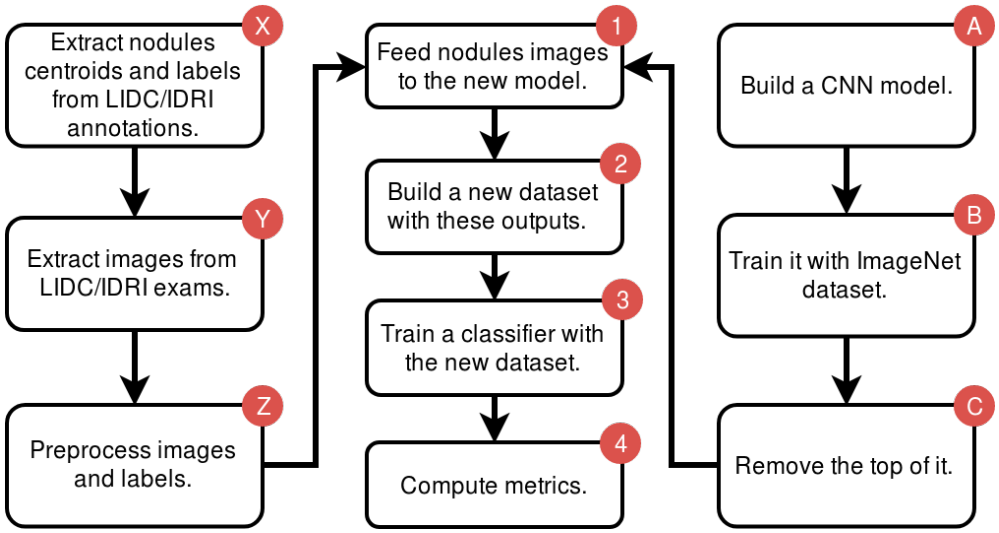

Figure 1. The block sequence X,Y and Z represent the steps to preprocess LIDC/IDRI information. The block sequence A,B and C represent the steps to build a deep feature extractor. The block sequence 1 and 2 represent the steps to build a new dataset and the blocks 3 and 4 the evaluation of the extracted features discriminative power.

Finally, these images are concatenated depthwise, assembling a 3 channels image.

It is noteworthy that, the annotation XML file provides the malignancy level for each one of the 2669 nodules. These levels are integer values ranging from 1 to 5 and indicate the malignancy degree of each nodule. However, this work aims to solve a two-class problem, building a system able to discriminate between malignant and non-malignant nodules images. For this reason, classes '1' and '2' were defined as non-malignant nodules, class '3' was disregarded and classes '4' and '5' defined as malignant nodules. Finally, at the end of this subsession, from the initial 7371 nodules only 1536 remain, more specifically, 1146 are nonmalignant and 390 are malignant.

B. Feature Extraction

Initially models VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile and NASNetLarge were built. Next, trained on 1.2 million natural images (ImageNet dataset). Then, model top layers, fully connected layers close to the output, were removed. Lastly, all model were applied to the 1536 remaining nodules images, generating a new datasets of each model.

C. Malignancy Classification

In order to analyze the extracted features representativeness, 5 classifiers were applied: Naive Bayes, KNN, SVM, MLP and RF. The Naive Bayes classifier used has a gaussian kernel. The hyperparameter "number of neighbors" from KNN classifier was chosen through a grid search, testing the odd values from 1 to 11.

SVM, MLP and RF hyperparameters were chosen through a random search with 20 iterations over each fold of a 5-fold cross-validation process. The random search for SVM explored the hyperparameters C and Y with 2^{-5} , 2^{-4} , 2^{-3} , ..., 2^{15} and 2^{-15} , 2^{-14} , ..., 2^{3} values, respectively [41].

The MLP was trained using the Adam algorithm [42] with an initial learning rate of 5e-04, learning rate decay of 0.01, maximum number of 1000 iterations, tolerance for the optimization of 1e-4 and epsilon of 1e-8. Moreover, a random search was performed to find the number neurons used in the hidden layer. This search ranged from 2 to 1000 neurons.

The random search for RF was used to find: which number of features to consider when looking for the best split (1,2,3,...,or 10); which tree maximum depth should be used (6 or none); which minimum number of samples required to split an internal node (1,2,3,...,or 10); which minimum samples per leaf should be used (1,2,3,...,or 10); whether bootstrap samples should be used when building trees or not, and which function should be used to measure the quality of a split (gini or entropy). Furthermore, RF train process used 3000 estimators.

D. Evaluation metrics

To evaluate the classifiers average generalization performance in a 10-fold cross-validation, accuracy (ACC), sensitivity (True Positive Rate, TPR), positive predictive value (PPV), and F-Score, area under curve (AUC) metrics were computed at each fold. More specifically, ACC is the number of samples correctly classified over all samples; TPR is the rate of positive samples correctly classified; PPV is the number of positive samples correctly classified over the number of samples classified as positive; F-Score is the harmonic mean between TPR and PPV; and AUC is the area under the plot of TPR against the rate of negative samples correctly classified at various threshold configurations and indicates the diagnostic capacity of a binary classifier. Finally, at the end of this process, each metric has its 10 values average.

E. Materials

All experiments were conducted on Ubuntu 17.10 operating system with 16GB RAM, Intel Core i5 processor. The deep extractors were implemented using the keras 2.1.4 library and classifiers were implemented using the scikit-learn 0.19.1 library.

V. Results And Discussion

This section presents the experimental results achieved by applying the Naive Bayes, KNN, SVM, MLP and RF classifiers on 11 datasets generated by the deep extractors: VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile and NASNetLarge.

The results are presented in two parts: the first one makes a comparison among deep extractors through the evaluation metrics ACC, AUC, F-Score, TPR, PPV and aims to obtain the best combination between deep extractor and classifier; the second one, makes a comparison among the best combination achieved with other approaches found on literature.

In Table 1, the higher value of TPR is from the combination between Resnet50 deep extractor and Linear SVM classifier, achieving an AUC of 92.67%, an ACC of 86.98%, a F-score of 77.08%, a TPR of 85.64% and a PPV of 70.51%.

Moreover, the second higher value of TPR and the higher value of AUC are from the combination between Resnet50 deep extractor and SVM RBF classifier, achieving an AUC of 93.19%, an ACC of 88.41%, a F-score of 78.83%, a TPR of 85.38% and a PPV of 73.48%.

Furthermore, the combination between deep extractor and classifier with higher values of ACC and F-Score is from Xception with MLP, achieving an ACC of 89.91%, an AUC of 92.90%, a F-score of 79.11% , a TPR of 75.13 % and a PPV of 84.03%.

However, as missing a malignant cancer (false negative) would be worse than a warning about the suspiciousness of a non-malignant nodule (false positive), TPR is considered more important than PPV when selecting the better classifier for lung nodules malignancy. It is noteworthy that, ResNet50 deep extractor achieved the two higher TPR values, leading us to select it for our approach.

Moreover, despite the fact that ResNet50 with linear SVM has achieved the higher TPR, his value is only 0.26% higher than ResNet50 with SVM RBF. In addition, SVM RBF outstands linear SVM 1.43 in ACC, 0.52 in AUC, 1.75 in F-score and PPV 2.97. Thus, this comparative analyze among metrics lead us to select SVM RBF classifier to be combined with ResNet50 deep extractor.

In Table 2, [45] results achieved superior metric AUC and values if compared to any other approach analyzed. However, the statistical validation performed in our approach has almost twice more samples as theirs and still achieved an AUC of 93.19%, differing only 5.81% from theirs. A similar situation can be seen at the [8] and [6] approaches.

Moreover, If compared with works with more than 1300 nodules images ResNet50 with SVM RBF approach achieved, not only the higher value of AUC, but also higher value of ACC and the second higher value of TPR, differing only 1.62% from the [43] approach. However, it is noteworthy that, this TPR difference is mitigated even further by the fact that our work used 11% more images in its statistical validation.

Table I
Accuracy (ACC), Area Under Curve (AUC), F-Score, Sensitivity (TPR), Positive Predictive Value (PPV) for the evaluated combinations between deep feature extractors and classifiers.

Model	Classifier	ACC(%)	AUC(%)	F-Score(%)	TRP(%)	PPV(%)
Xception	Bayes	83.66	83.87	62.49	54.87	73.66
	MLP	89.91	92.90	79.11	75.13	84.03
	Nearest Neighbors	87.89	89.99	76.36	77.44	75.82
	Random Forest	88.61	90.98	75.48	70.00	82.64
	SVM Linear	86.85	92.55	76.68	84.87	70.44
	SVM RBF	88.28	92.93	78.61	84.87	73.40
ResNet50	Bayes	83.20	82.65	63.08	56.41	72.19
	MLP	89.13	92.25	76.24	69.23	85.37
	Nearest Neighbors	86.85	89.38	73.21	70.77	76.23
	Random Forest	88.80	92.15	77.75	76.92	78.97
	SVM Linear	86.98	92.67	77.08	85.64	70.51
	SVM RBF	88.41	93.19	78.83	85.38	73.48
MobileNet	Bayes	83.21	89.57	69.05	73.59	65.38
	MLP	87.76	91.08	74.81	72.05	78.09
	Nearest Neighbors	87.57	91.04	72.94	66.41	81.20
	Random Forest	87.18	91.64	73.13	68.46	78.68
	SVM Linear	84.25	89.64	70.70	75.13	67.07
	SVM RBF	88.67	91.79	75.73	70.51	82.21
VGG16	Bayes	77.87	85.34	24.81	14.62	90.83
	MLP	83.53	88.55	67.07	66.41	68.07
	Nearest Neighbors	84.05	85.23	64.55	57.18	74.76
	Random Forest	84.83	89.82	60.09	45.90	88.70
	SVM Linear	80.27	87.90	67.61	80.77	58.28
	SVM RBF	86.53	89.40	72.21	69.49	75.86
VGG19	Bayes	76.43	83.60	15.26	8.46	86.67
	MLP	87.69	90.75	74.10	70.00	79.33
	Nearest Neighbors	86.00	88.31	70.25	65.13	77.19
	Random Forest	87.11	89.55	73.27	70.26	76.90
	SVM Linear	82.03	89.61	69.90	82.31	60.84
	SVM RBF	84.90	89.82	72.40	78.21	67.74
InceptionV3	Bayes	80.08	84.50	63.30	67.69	59.59
	MLP	87.83	91.79	74.45	70.77	81.08
	Nearest Neighbors	87.04	89.46	73.61	71.03	76.84
	Random Forest	85.81	89.91	69.21	63.08	77.14
	SVM Linear	87.50	91.41	77.31	83.85	71.88
	SVM RBF	88.87	92.50	77.34	74.62	80.44
InceptionResNetV2	Bayes	83.52	83.91	62.63	54.87	73.46
	MLP	89.00	92.63	77.10	73.33	82.70
	Nearest Neighbors	87.50	88.91	75.51	75.90	75.56
	Random Forest	86.72	90.22	75.05	78.97	71.80
	SVM Linear	87.24	92.53	77.34	85.38	70.90
	SVM RBF	87.04	92.48	76.84	84.62	70.61
NASNetMobile	Bayes	82.81	85.23	54.58	41.28	82.13
	MLP	87.37	89.98	72.24	65.64	81.03
	Nearest Neighbors	85.10	86.98	69.55	67.18	72.63
	Random Forest	84.51	87.66	70.33	73.08	68.09
	SVM Linear	83.73	88.92	72.22	83.33	64.00
	SVM RBF	85.22	89.84	73.67	81.28	67.81
NASNetLarge	Bayes	83.21	85.42	66.44	65.64	67.41
	MLP	88.08	90.79	74.26	69.23	82.41
	Nearest Neighbors	87.63	87.97	74.53	71.54	78.07
	Random Forest	87.56	90.17	72.06	63.33	84.02
	SVM Linear	87.57	91.00	76.75	81.03	73.42
	SVM RBF	88.28	90.09	75.32	70.77	80.98
DenseNet169	Bayes	81.32	82.25	45.64	31.28	86.80
	MLP	86.52	89.67	68.94	61.03	82.26
	Nearest Neighbors	86.46	89.10	72.72	71.54	74.57
	Random Forest	86.53	91.16	74.41	77.44	72.07
	SVM Linear	86.13	91.35	75.15	82.82	68.95
	SVM RBF	86.66	90.24	72.91	70.77	75.70
DenseNet201	Bayes	81.76	86.68	47.09	32.56	87.56
	MLP	87.18	91.83	71.01	62.31	83.84
	Nearest Neighbors	87.11	89.57	73.12	69.49	77.76
	Random Forest	87.50	91.70	71.80	63.08	83.83
	SVM Linear	85.81	92.47	75.26	84.62	68.14
	SVM RBF	87.82	90.66	74.95	72.05	78.78

Table II
THE COMBINATION BETWEEN THE RESNET50 DEEP EXTRACTOR AND SVM (RBF) CLASSIFIER IS COMPARED WITH RELATED WORKS THROUGH THE METRICS ACCURACY (ACC), AREA UNDER CURVE (AUC), F-SCORE, TRUE POSITIVE RATE (TPR), POSITIVE PREDICTIVE VALUE (PPV).

Approach	Number of nodules	ACC(%)	AUC(%)	F-Score(%)	TRP(%)	PPV(%)
Our Approach	1536	88.41	93.19	78.83	85.38	73.48
Shen *et al.* [5]	1375	86.80	-	-	-	-
Shen *et al.* [3]	1375	87.14	93.00	-	77.00	-
Aerts *et al.* [43]	1375	83.21	89.00	-	87.00	-
Han *et al.* [7]	1356	-	92.70	-	-	-
Hussein *et al.* [2]	1340	91.26	-	-	-	-
Zhu *et al.* [44]	1004	90.44	-	-	-	-
Dhara *et al.* [6]	891	-	95.05	-	89.73	-
Kang *et al.* [45]	776	95.41	99.00	-	95.68	-
Wei *et al.* [9]	746	85.20	-	85.80	85.80	85.95
Wei *et al.* [8]	366	91.00	98.40	91.80	92.10	92.30
Sergeeva *et al.* [10]	321	81.30	89.60	76.50	70.90	83.00
Ma *et al.* [46]	157	82.70	-	-	80.00	-

VI. CONCLUSION

In this paper, we explore deep learning models as feature extractors to tackle the challenging problem of lung nodule malignancy classification, which is represented here by the LIDC/IDRI dataset.

Moreover, it was demonstrated that deep transfer learning from non-medical images is a relevant strategy to extract representative imaging biomarkers at the macroscopic level in chest CT images.

Among models and classifiers explored, the combination between ResNet50 and SVM RBF was selected as the best one, achieving an AUC of 93.19%, an ACC of 88.41%, a F-score of 78.83%, a TPR of 85.38% and a PPV of 73.48%. Such results outstand among related works, specially AUC and ACC metrics.

However, this approach still has room for improvements, e.g., apply a training strategy called finetune on ResNet50 with LIDC/IDRI dataset, combine features extracted from different deep models and/or combine deep features with hand-crafted features.

In general, the deep extractor ResNet50 combined with SVM RBF classifier have achieved satisfactory results and can be integrated with a CAD system to give specialist doctors a second opinion about the malignancy suspiciousness of a lung nodule.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, jan 2017.

[2] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d cnn-based multi-task learning," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 249–260.

[3] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, pp. 663–673, 2017.

[4] P. M. Marcus, V. P. Doria-Rose, I. F. Gareen, B. Brewer, K. Clingan, K. Keating, J. Rosenbaum, H. M. Rozjabek, J. Rathmell, J. Sicks *et al.*, "Did death certificates and a death review process agree on lung cancer cause of death in the national lung screening trial?" *Clinical Trials*, vol. 13, no. 4, pp. 434–438, 2016.

[5] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.

[6] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung ct images," *Journal of digital imaging*, vol. 29, no. 4, pp. 466–475, 2016.

[7] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *Journal of digital imaging*, vol. 28, no. 1, pp. 99–115, 2015.

[8] G. Wei, H. Ma, W. Qian, F. Han, H. Jiang, S. Qi, and M. Qiu, "Lung nodule classification using local kernel regression models with out-of-sample extension," *Biomedical Signal Processing and Control*, vol. 40, pp. 1–9, 2018.

[9] G. Wei, H. Cao, H. Ma, S. Qi, W. Qian, and Z. Ma, "Content-based image retrieval for lung nodule classification using texture features and learned distance metric," *Journal of Medical Systems*, vol. 42, no. 1, p. 13, Nov 2017.

[10] M. Sergeeva, I. Ryabchikov, M. Glaznev, and N. Gusarova, "Classification of pulmonary nodules on computed tomography scans. evaluation of the effectiveness of application of textural features extracted using wavelet transform of image," in *2016 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT)*. IEEE, apr 2016.

[11] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.

[12] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, may 2016.

[13] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 76–84, 2017.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, oct 2010.

[15] C.-K. Shie, C.-H. Chuang, C.-N. Chou, M.-H. Wu, and E. Y. Chang, "Transfer representation learning for medical image analysis," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 711–714.

[16] J. J. Näppi, T. Hironaka, D. Regge, and H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography," in *Medical Imaging 2016: Computer-Aided Diagnosis*, G. D. Tourassi and S. G. Armato, Eds. SPIE, mar 2016.

[17] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, aug 2016.

[18] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," in *Medical Imaging 2016: Computer-Aided Diagnosis*, G. D. Tourassi and S. G. Armato, Eds. SPIE, mar 2016.

[19] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, jan 2017.

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

[21] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P.-Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke, "The lung image database consortium (LIDC) and image database resource

initiative (IDRI): A completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, jan 2011.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint arXiv:1610.02357*, 2016.
[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[28] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *AAAI*, 2017, pp. 4278–4284.
[29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jul 2017.
[30] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *arXiv preprint arXiv:1707.07012*, 2017.
[31] S. Theodoridis and K. Koutroumbas, *Pattern Recognition (Fourth Edition)*. USA: Academic Press, 2008.
[32] S. Haykin, *Neural Networks and Learning Machines*. McMaster University, Canada: Prentice Hall, 2008.
[33] V. N. Vapnik, *Statistical Learning Theory*. Nova Jersey, EUA: John Wiley & Sons, 1998.
[34] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
[35] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
[36] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
[37] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient transfer learning," *arXiv preprint arXiv:1611.06440*, 2016.
[38] K. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? an empirical study," in *Multiple Classifier Systems*, June 2005, pp. 278–285.
[39] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. C-24, no. 7, pp. 750–753, July 1975.
[40] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.
[41] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.
[43] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, 2014.
[44] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: 3d deep convolutional nets for automated pulmonary nodule detection and classification," *arXiv preprint arXiv:1709.05538*, 2017.
[45] G. Kang, K. Liu, B. Hou, and N. Zhang, "3d multi-view convolutional neural networks for lung nodule classification," *PloS one*, vol. 12, no. 11, p. e0188290, 2017.
[46] J. Ma, Q. Wang, Y. Ren, H. Hu, and J. Zhao, "Automatic lung nodule classification with radiomics approach," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 978 906–978 906.