

Deep Feature Learning for Pulmonary Nodule Classification in a Lung CT

Bum-Chae Kim¹, Yu Sub Sung², and Heung-Il Suk^{1,*}

¹Department of Brain and Cognitive Engineering, Korea University, Republic of Korea

²Biomedical Imaging Infrastructure, Department of Radiology, Asan Medical Center, Republic of Korea

*Corresponding author: hisuk@korea.ac.kr

Abstract— In this paper, we propose a novel method of identifying pulmonary nodules in a lung CT. Specifically, we devise a deep neural network by which we extract abstract information inherent in raw hand-crafted imaging features. We then combine the deep learned representations with the original raw imaging features into a long feature vector. By taking the combined feature vectors, we train a classifier, preceded by a feature selection via *t*-test. To validate the effectiveness of the proposed method, we performed experiments on our in-house dataset of 20 subjects; 3,598 pulmonary nodules (malignant: 178, benign: 3,420), which were manually segmented by a radiologist. In our experiments, we achieved the maximal accuracy of 95.5%, sensitivity of 94.4%, and AUC of 0.987, outperforming the competing method.

Keywords- Pulmonary nodule classification; Lung cancer; Deep learning; Stacked denoising autoencoder

I. INTRODUCTION

The lung cancer mortality is reported as the most common cause of death worldwide [1]. A variety of ways are being tried to reduce the mortality ratio. It is known that once the cancer is diagnosed in an early stage, there exist therapeutic effects and contributes to overcome it. Furthermore, to mitigate the potential misdiagnosis due to mostly fatigues of doctors while reading a huge amount of CT scans, the computer-assisted screening or intervention has been of great interest.

From a clinical perspective, nodules of sized larger than 3mm is mostly referred to as pulmonary nodules [2] and the increasing-sized nodules are more likely to become cancerous. Therefore, diagnosis through the detection and observation of the nodule is important for screening. In this regard, the computer-assisted screening system has been proposed for the last decades, although those are not used in the clinic due to their low performance.

Recently, motivated by the great success of deep learning in the fields of computer vision and speech recognition, there have been efforts to apply the technique for medical diagnosis, especially for nodule detection in CT. For example, Roth *et al.*, used Convolution Neural Network (CNN) [3], one of most widely used deep learning models, for nodule detection [4]. Ciompi *et al.*, used CNN for feature extraction to identify pulmonary peri-fissural nodules [5]. To enhance the detection accuracy, they jointly exploited the information from axial, coronal, sagittal slices. In the meantime, Fakoor *et al.* and Kumar *et al.*, studied, independently, in cancer genomic data

classification [6] and pulmonary nodule classification [7] by using Stacked AutoEncoder (SAE) [6]. Although the previous deep learning-based methods have presented the effectiveness in their own experiments, they mostly ignored morphological information such as perimeter, circularity diameter, integrated density, median, skewness, and kurtosis of a nodule, which cannot be extracted by conventional deep models. In this paper, we propose to use a deep model to find latent information inherent in morphological features and to combine the deep-learned representations with the original morphological features. As for deep feature learning, we use a Stacked Denoising AutoEncoder (SDAE) [8]. Our work is inspired by Suk *et al.*'s work [11], where they combined original neuroimaging features with the deep learned features for Alzheimer's disease diagnosis.

II. PROPOSED METHOD

A. Dataset and Morphological Features

We collected CT scans from 20 patients (male/female: 7/13, age: 63.5 ± 7.7). Pulmonary nodules were manually segmented by a well-trained radiologist. In total, we have 178 malignant and 3,420 benign nodules (Table 1). Figure 1 shows examples of pulmonary samples with high intra- and inter-class variations, which make the nodule classification very challenge. From each nodule, we extracted 96 morphological features, namely, area, mean Hounsfield Units (HU)¹, standard deviation, mode, min, max, perimeter, circularity diameter, integrated density, median, skewness, kurtosis, raw integrated density, ferret angle, min ferret, aspect, ratio, roundness, solidity, entropy, run length matrix (44 values) [9], and gray-level co-occurrence matrix (32 values) [10].

TABLE 1. DATA SUMMARY

Subjects		Nodule data	
Gender	Male: 7 Female: 13	Full dataset	Malignant: 178 Benign: 3,420
			Total : 3,598
Age [min/max]	63.5±7.7 [51/75]	Dataset used for performance evaluation	Malignant: 178 Benign: 200
			Total: 378

¹ HU: a quantitative scale for describing radiodensity.

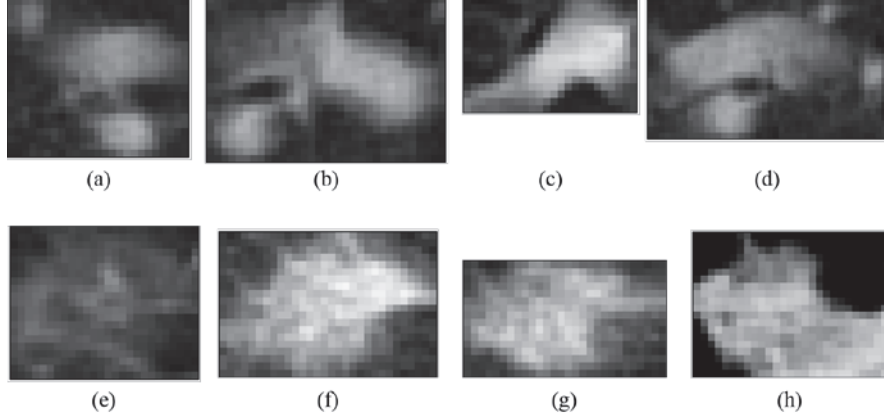


Figure 1. Examples of pulmonary nodules: (a)-(d) benign, (e)-(h) malignant.

B. Learning High-Level Relational Information

To better utilize the feature information, we use SDAE to discover latent non-linear relations among morphological features. SDAE is structured by stacking multiple auto-encoders in a hierarchical manner. An AE is a multi-layer neural network with an input layer, a hidden layer, and an output layer. The number of units in input and output layers is equal to the dimension of input features $\mathbf{x} \in \mathbb{R}^d$, i.e., d , while the number of hidden units can be arbitrary. In AE, the values of hidden (\mathbf{h}) and output (\mathbf{o}) units are obtained as follows:

$$\mathbf{h} = \phi(\mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)}) \quad (1)$$

$$\mathbf{o} = \mathbf{W}^{(2)} \mathbf{h} + \mathbf{b}^{(2)} \quad (2)$$

where $\phi(\cdot)$ is a non-linear sigmoid function. The parameters of $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$, $\mathbf{b}^{(1)}$, and $\mathbf{b}^{(2)}$ are learned such that the values of hidden units can recover the input feature values, i.e., $\mathbf{x} \approx \mathbf{o}$. However, to make the AE robust to unfavorable noises, we can change the training protocol slightly. Specifically, during training we explicitly contaminate the original input values by adding random noises but trained the model to have the values of the output layers to be as close as to the original uncontaminated values. This model is known as ‘*Denoising AutoEncoder*’ (DAE) [8].

Note that the values of hidden units can be used as alternative representation of an input features in a new space, where different dimensions denote different relations among the original features. By stacking multiple DAEs hierarchically such that the values of hidden units become the input to the next upper AE, we construct a deep architecture and we call it ‘*Stacked Denoising AutoEncoder*’ (SDAE).

One remarkable advantage of DAE is that its parameters can be learned in an unsupervised manner. So we can exploit as many training instances as possible regardless of the validity of their label information. This favorable characteristic can be further utilized in finding the “good” initial values of the parameters in SDAE by means of pre-training [11]. In a nutshell, an SDAE is first pre-trained in an unsupervised way

and the pre-trained parameter values are used as the initial value to train a deep neural network, structured SDAE with an additional label layer at the top. Then we fine-tune all the parameters in a supervised manner. After training a SDAE, we take the outputs of the top hidden layer, which are the input to the label layer in our SDAE, as high-level relational information inherent in the original morphological features. We finally concatenate the original features and the SDAE-learned features into a long vector and regard it as our new augmented feature vector.

C. Feature Selection and Classifier Training

From the previous work in the field of pattern recognition, it is well studied that feature selection before classifier learning is very helpful to improve a classifier’s performance [11]. Motivated by their work, we apply a feature selection method via statistical test between features and class labels. Specifically, we perform a simple t -test for each feature individually and when the resulting p -value is larger than pre-defined threshold, we regard the corresponding feature is not informative for classification. Based on the selected features, we finally train a linear Support Vector Machine (SVM), which has already proved its efficiency in various applications [12], as a classifier.

III. EXPERIMENTAL RESULTS

A. Experimental Settings

We designed our SDAE to have 5 layers with 3 hidden layers. The number units in the three hidden layers were 300, 200, and 100. For SDAE training, we used a with back-propagation algorithm [13] in DeepLearnToolbox² with a mini-batch size of 50. As for the non-linear sigmoid function in AE, we used a hyperbolic tangent function. To maximally utilize our data samples and to better utilize the unsupervised characteristic of DAE training, we used all samples in our dataset during pre-training³, i.e., 178 malignant and 3,420 benign samples.

² Available at <https://github.com/rasmusbergpalm/DeepLearnToolbox>

³ We set the number of epochs 200.

It should be noted that because of the imbalance of training samples between malignant and benign classes, we randomly selected 200 out of 3,420 samples of the benign class. In our performance evaluation, we utilized 178 malignant and 200 benign (randomly selected) samples only and conducted 5-fold cross-validation. In other words, we set aside samples of one fold for testing only and used samples of the remaining 4 folds. We should emphasize that during fine-tuning of our SDAE and SVM learning, we used the samples in 4 folds only with never involving the hold-out testing samples.

After fine-tuning of our SDAE, we augmented our feature vectors to be 196-dimension by concatenating output values of the top hidden units, i.e., 100 values, and the original 96-dimensional features. In the feature selection step, we set the threshold of p -value 0.001. Lastly, the model hyper-parameter C of our SVM was determined by a 5-fold nested cross-validation in the space of $\{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$ using a libSVM library⁴. To validate the effectiveness of the proposed feature augmentation with deep-learned features, we compare our method with the conventional method that uses only the original raw features. Note that all the other setting for the competing method is the same with the proposed method.

B. Performance

We used four metrics for performance measurement, namely, accuracy, sensitivity, specificity, and Area Under receiver operating characteristic Curve (AUC). Figure 1 compare with average value of performance measurement. AUC use different measure unit. Another performances use percentage. Original + SDAE features achieved better performance in every average value. Specially, average accuracy and sensitivity are improved 2.1% and 3.4%, respectively.

IV. CONCLUSION

In this paper, we proposed to use a deep architecture to find latent non-linear information in morphological features for pulmonary nodule classification in CT scans. Clinically, finding malignant nodule is more important in the early stage. Our deep learned feature boosted the power of discriminating between malignant and benign nodules, with high improvement in sensitivity.

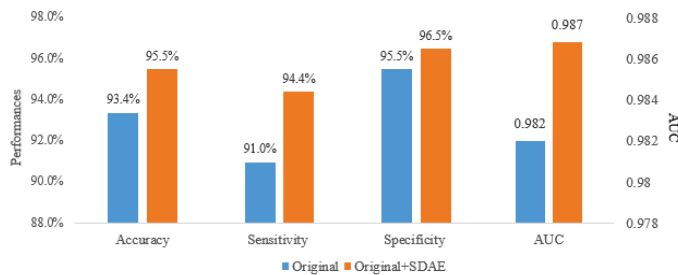


Figure 2. Performance comparison.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1C1A1A01052216). The authors gratefully acknowledge technical supports from Biomedical Imaging Infrastructure, Department of Radiology, Asan Medical Center.

REFERENCES

- [1] B. W. Stewart and C. P. Wild, editors, *World Cancer Report 2014*. Lyon, France: International Agency for Research on Cancer, Feb. 2014.
- [2] M. K. Gould, J. Fletcher, M. D. Iannettoni, W. R. Lynch, D. E. Midthun, D. P. Naidich, and D. E. Ost, "Evaluation of Patients with Pulmonary Nodules: When is it lung cancer?: ACCP evidence-based clinical practice guidelines, 2ed edition" *Chest*, Vol. 132, No. 3, pp. 108-130, Sep. 2007.
- [3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, Nov. 1998.
- [4] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation," arXiv:1505.03046, Sep. 2015.
- [5] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. Th. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, "Automatic Classification of Pulmonary Peri-Fissural Nodules in Computed Tomography Using an Ensemble of 2D views and a Convolutional Neural Network Out-of-the-Box," *Medical Image Analysis*, Vol. 26, No. 1, pp. 195-202, Dec. 2015.
- [6] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using Deep Learning to Enhance Cancer Diagnosis and Classification," *Proc. of the ICML Workshop on the Role of Machine Learning in Transforming Healthcare (WHEALTH)*, Vol. 28, June 2013.
- [7] D. Kumar, A. Wong, and D. A. Clausi, "Lung Nodule Classification Using Deep Features in CT Images," *Proc. 12th Conference on Computer and Robot Vision (CRV)*, pp. 133-138, June 2015.
- [8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *The Journal of Machine Learning Research*, Vol. 11, pp. 3371-3408, Mar. 2010.
- [9] X. Tang, "Texture Information in Run-length Matrices," *IEEE Transactions on Image Processing*, Vol. 7, No. 11, pp. 1602-1609, Nov. 1998.
- [10] R. M. Hatalick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, No. 6, pp. 610-621, Nov. 1973.
- [11] H.-I. Suk, S.-W. Lee, and D. Shen, "Latent Feature Representation with Stacked Auto-Encoder for AD/MCI Diagnosis," *Brain Structure & Function*, Vol. 220, No. 2, pp. 841-859, Mar. 2015.
- [12] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, Vol. 313, No. 5786, pp. 504-507, July 2006.
- [13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-propagating Errors," *Nature*, Vol. 323, No. 9, pp. 533-536, Oct. 1986.

⁴ Available at <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>