Unsupervised Deep Learning Features for Lung Cancer Overall Survival Analysis

Shuo Wang, Zhenyu Liu, Xi Chen, Yongbei Zhu, Hongyu Zhou, Zhenchao Tang, Wei Wei, Di Dong, Meiyun Wang, Jie Tian, *Fellow, IEEE*

***Abstract*— Lung cancer overall survival analysis using computed tomography (CT) images plays an important role in treatment planning. Most current analysis methods involve hand-crafted image features for survival time prediction. However, hand-crafted features require domain knowledge and may lack specificity to lung cancer. Advanced self-learning models such as deep learning have showed superior performance in many medical image tasks, but they require large amount of data which is difficult to collect for survival analysis because of the long follow-up time. Although data with survival time is difficult to acquire, it is relatively easy to collect lung cancer patients without survival time. In this paper, we proposed an unsupervised deep learning method to take advantage of the unlabeled data for survival analysis, and demonstrated better performance than using hand-crafted features. We proposed a residual convolutional auto encoder and trained the model using images from 274 patients without survival time. Afterwards, we extracted deep learning features through the encoder model, and constructed a Cox proportional hazards model on 129 patients with survival time. The experiment results showed that our unsupervised deep learning feature gained better performance (C-Index = 0.70) than using hand-crafted features (C-Index = 0.62). Furthermore, we divided the patients into two groups according to their Cox hazard value. Kaplan-Meier analysis indicated that our model can divide patients into high and low risk groups and the survival time of these two groups had significant difference ($p < 0.01$).**

***Index Terms*— Lung cancer, survival analysis, deep learning, unsupervised feature learning, convolutional neural networks**

I. Introduction

Lung cancer is the leading cause for cancer related deaths with a 5-year survival rate of only 18% [1], and lung cancer overall survival analysis can provide a personalized treatment plan for each patient. Recently, researchers found that computed tomography (CT) images of lung cancer patients are associated with overall survival time [2]–[4]. In these studies, thousands of hand-crafted CT image features such as first-order statistic values and texture features are extracted. Afterwards, feature-selection algorithm is applied to reserve only less than 100 features. Finally, a multivariate Cox proportional hazards model [5] is used to assign a score for each patient indicating their relative hazard. In the Cox regression model, a patient gets relatively low score if he/she lives longer than other patients. To quantitatively measure the performance of Cox model, Harrell's concordance index (C-Index) [4] was used as a measurement whose range is between [0, 1].

From the published literatures, hand-crafted features (called radiomic feature) showed prognostic value in lung cancer survival analysis [3], [4]. However, these generalized features reflected limited information and may not be adaptive to lung cancer since they can only describe low-level visual features instead of high-level abstract features. To extract discriminative features that are adaptive to specific task, deep learning (especially convolutional neural networks, CNN) is broadly used in many lung cancer analysis tasks [4], [6].

Despite the advantage of CNN, few works utilized deep learning to lung cancer survival analysis because of the limited data amount. Unlike other classification tasks [6], survival analysis requires long follow-up time to record the death date of patient which is difficult to acquire. Therefore, applying supervised deep learning model for survival analysis may suffer from over-fitting due to the limited data amount.

Although data with survival time is difficult to collect, it is relatively easy to collect images of lung cancer patients without survival time (unlabeled data). In this work, we proposed an unsupervised feature learning method to learn features from unlabeled data, and then utilize these deep learning features for survival analysis. These unlabeled data may not be directly used in supervised model training, but they can be used for mining lung cancer-related features instead of using hand-crafted features. In this work, we proposed a residual convolutional neural networks (RCAE) to learn deep features from unlabeled lung cancer images. Afterwards, these deep features were used to build Cox model for overall survival analysis. More specifically, our contributions are three-folds: 1) we learned features from unlabeled lung cancer images (from 274 patients) instead of hand-crafted features. This makes our features more adaptive to lung cancer. 2) When extracting deep learning features,

This work was supported by the National Natural Science Foundation of China [81227901, 61231004, 81501616, 81671851, 81527805, and 81501549], the Science and Technology Service Network Initiative of the Chinese Academy of Sciences [KFJ-SW-STS-160], the special program for science and technology development from the Ministry of science and technology, China [2017YFA0205200, 2017YFC1308701, 2017YFC1309100, 2016CZYD0001], the Instrument Developing Project [YZ201502], the Beijing Municipal Science and Technology Commission [Z161100002616022], and the Youth Innovation Promotion Association CAS.

$^{+}$S. Wang (wangshuo2014@ia.ac.cn), $^{+}$Z. Liu, Y. Zhu, Z. Tang, W. Wei, D. Dong, Y. Zang, and *J. Tian (tian@ieee.org) are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing 100190, China. * are corresponding authors. $^{+}$ are co-first authors.

*M. Wang is with the Department of Radiology, Henan Provincial People's Hospital.

H. Zhou is with the Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

X. Chen is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China.

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

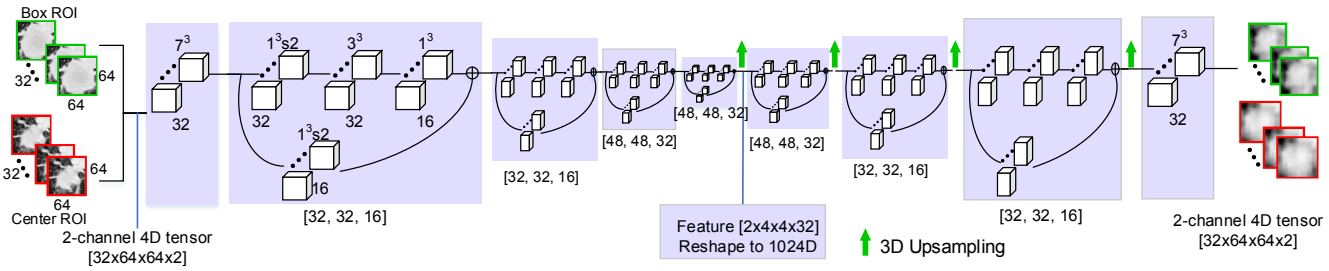


Fig. 1. Illustration of the residual convolutional auto encoder. The network contains nine residual blocks. Each residual block shares the same structure including four 3D convolutional layers. The numbers in brackets under residual blocks are the kernel numbers of each convolutional layer inside this block. The numbers on convolutional layers indicate the convolutional kernel size (1^3s2 represent convolutional kernel size $1 \times 1 \times 1$ with stride 2). The output of the fourth residual block is used to extract deep learning feature. The input image is a two-channel 4D tensor where one channel represents the ROI cropped on tumor center (center volume), and the other channel is the resized tumor (box volume).

we extracted two ROIs on each lung tumor (a center volume and a box volume) to enable multi-scale feature learning. For each patient, we extracted a fixed-size ROI on tumor center (center volume) and a ROI containing the whole tumor (box volume). This strategy enables our feature leaning network to extract multi-scale features through multi-scale input. 3) We tested our method on 64 patients with survival time, and achieved higher C-Index (0.70) than using hand-crafted features (0.62).

II. Method

Our method includes: 1) unsupervised feature learning and 2) supervised Cox model training. The feature learning phase aims at learning lung cancer features from large amount of unlabeled CT images of lung tumor. Afterwards, we extracted the learned features on labeled dataset and built a supervised Cox model for overall survival analysis.

A. Unsupervised Feature Learning

In this phase, we built a residual convolutional auto encoder (RCAE) to learn features from unlabeled lung cancer images (Fig. 1). This model includes nine residual blocks [7]. The first five residual blocks construct an encoder while the last four residual blocks construct a decoder. Each residual block contains four 3D convolutional layers which can be formulated as $f_j = ReLU\left(BN\left(\sum_i c_{ij} * f_i + b_j\right)\right)$. In this equation, c_{ij} denotes 3D convolutional kernel between feature map f_i and f_j (* is 3D convolution operation). b_j is the bias of convolutional kernel c_{ij}. After convolution, batch normalization ($BN()$) is applied, and then, $ReLU(u) = max(0, u)$ is used as activation function. Inside residual block, three convolutional layers (kernel size 1^3, 3^3, 1^3) are stacked as forward path while another layer of kernel size 1^3 is a shortcut path. These two paths are finally added together for residual learning. In the encoder part of this network, we used convolutional stride = 2 for feature map down sampling instead of max pooling [7], [8]. In the decoder part, 3D up-sampling is used between residual blocks.

The early studies [6], [9] suggest that multi-scale input image enables network to learn multi-scale information and shows better performance than single-scale input. Therefore, we extract two tumor ROIs at two scales as the input of our proposed RCAE. For a given lung tumor, we extract two ROIs called center volume and box volume. The center volume is cropped on tumor center with a $32 \times 64 \times 64$ window. The box volume contains the whole tumor and is resized to $32 \times 64 \times 64$ by third-order spline interpolation. Finally, these two 3D volumes construct a 2-channel 4D tensor ($32 \times 64 \times 64 \times 2$) and are fed into the RCAE network. Following the auto encoder structure, the output of this network is a 4D tensor that has the same shape with the input of the network.

The goal of the network training is to encode image to low-dimensional features and finally reconstruct the input image from limited features. This can be achieved by minimizing the "mean squared error" loss function $Loss\,(W) = \frac{1}{N}\sum_{n=1}^{N}\|O - I\|_2^2 + \lambda|W|$. In this equation, O and I represent the output and input images of the network. W is the weight of the network. λ controls the regularization strength, and is set to 3×10^{-5}. N is training sample numbers. This loss function is minimized through Adam [10] algorithm with learning rate of 1×10^{-3}. Finally, the output of the fourth residual block is extracted as the deep learning feature which is 1024-dimensional.

To avoid over-fitting, we augmented the training samples by mirroring the tumor image on horizontal and vertical directions on the axial plane. In addition, drop out [11] was also applied after every convolutional layer with dropout rate of 0.25. The network was implemented using keras, and the training process converged after 20 epochs of training.

B. Overall Survival Analysis

After unsupervised feature learning, we extracted the fourth residual block of the RCAE network as a feature encoder for lung cancer images. For a given patient, we extracted two ROIs (center volume and box volume) on the tumor, and fed them into the feature encoder, and finally we acquired 1024 features as the deep learning feature of this patient. This feature dimensionality is too large compared with the amount of labeled data. Therefore, feature reduction is necessary before building survival analysis model. In this study, we used LASSO-Cox model that integrated both feature selection and Cox model building. Compared with single Cox model, LASSO-Cox added 1-norm regularization on the

parameters of Cox model, and therefore achieved feature selection purpose. Specifically, the LASSO-Cox added a penalization term $\lambda[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1]$ on the loss function of Cox model during training. In this equation, β is the parameter of the Cox model, and α and λ controls the regularization strength. In our experiment, these two parameters were selected through a five-fold cross validation on the training set. Finally, The LASSO-Cox model outputs a score for each patient indicating the relative hazard of this patient compared with other patients.

Since the early studies also used CT image features for overall survival prediction [3], [4], we compared the performance of our unsupervised deep learning features with the published hand-crafted features which are defined as radiomic feature. In our study, we extracted 1108 radiomic features including tumor intensity, shape, texture and wavelet-decomposed features through the public pyradiomics library [12]. Afterwards, we used LASSO-Cox for model building, and the parameters were selected using five-fold cross validation on the training set.

III. Experiments and Results

A. Dataset

We collected 403 images from the Department of Radiology at Henan Provincial People's Hospital during 2012 to 2015. Chest contrast-enhanced CT was performed on every patient using one of the two multi-detector row CT (MDCT) systems (Philips Brilliance 16 slices CT, Phillips Medical System, the Netherlands or 64-slice LightSpeed VCT, GE Medical systems, Milwaukee, USA), with the following acquisition parameters: 120 kV; 160 mAs; 0.5 or 0.4 s rotation time; detector collimation: 16 1.25 mm or 64 0.625 mm; field of view, 350 350 mm; matrix, 512 512. The CT image was reconstructed with a standard kernel. The slice thickness of the CT images was in the range between 0.625 mm and 1.25 mm.

In this dataset, 274 patients are unlabeled data and 129 patients are labeled data which have survival time (months). In the 129 patients, 20 cases are censored data which means the death did not happen when the follow-up time ended, but the follow up time is longer than one year. In the labeled dataset, we sorted the patients in scanning time order, and the first 65 patients were used as training set while the last 64 patients were used as testing set which is a commonly used way to split dataset in survival analysis and radiomic analysis [13], [14].

B. Unsupervised Model Training

The unlabeled dataset is used for unsupervised RCAE network training. After 20 epochs of training, this network converged. Fig. 2 illustrated four CT images and the corresponding outputs of the RCAE network.

The input image of RCAE is $32 \times 64 \times 64 \times 2$, however, it will be compressed into 1024D which is only %0.4 of input image size. The output image in Fig. 2 showed that the encoded 1024D feature contained most tumor information since the sketch of the tumor can be reconstructed using the 1024D features. Compared with hand-crafted radiomic features, these deep learning features are learned from lung cancer directly which makes them more adaptive to lung cancer.

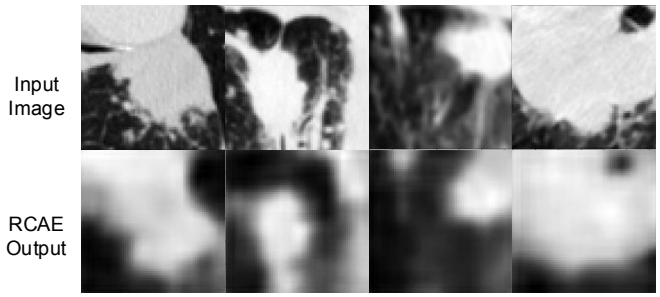


Fig. 2. Input and output of the RCAE network. The first row illustrates four CT image slices, and the second row is the corresponding output of the RCAE network.

C. Quantitative Results

When using the LASSO-Cox model for overall survival analysis, 8 features were selected from the deep learning features, and 7 features were reserved from the radiomic features. The parameters were $\alpha = 0.6$, $\lambda = 0.43$ for deep learning feature and $\alpha = 0.6$, $\lambda = 0.37$ for radiomic feature.

To evaluate the performance of the Cox model, we used Harrell's concordance index (C-Index) as a main measurement. In addition, the Cox model can predict a survival function for each patient indicating the survival probability of this patient at each time point. We further used the predicted 1-year survival probability from Cox model to predict the 1-year survival rate in this dataset. Area under the receiver operating curves (AUC) and accuracy were used to measure the classification performance.

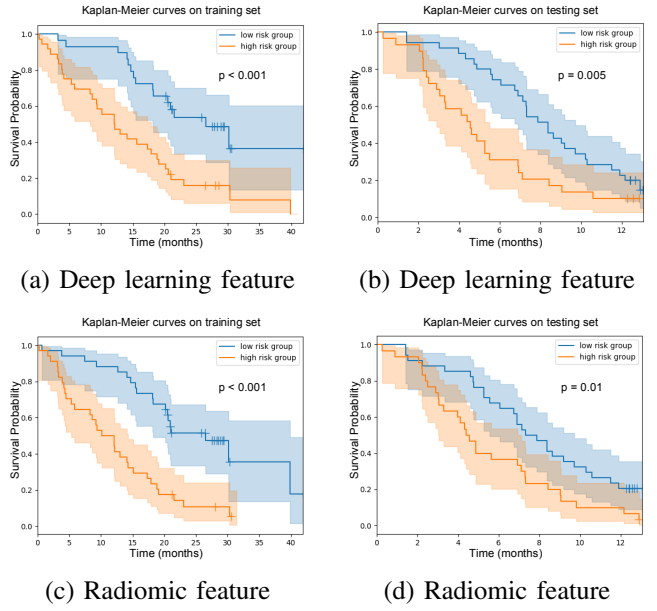
Table I showed the C-Index, AUC, and accuracy of the Cox model on the training and testing set using deep learning and radiomic features. By comparing the C-Index of deep learning feature and radiomic feature, we found that the deep learning features generalized better than radiomic features, because the deep learning feature generated higher C-Index on the testing set than radiomic features. In terms of the 1-year survival rate prediction, the deep learning features gained higher AUC and accuracy than traditional radiomic features.

To further illustrate the performance of the deep learning feature, we introduced Kaplan-Meier analysis. After the Cox model predicting hazard score for each patient, we used a threshold to divide the dataset into two groups where one group had patients with shorter survival time (high risk group) and the other group included patients with longer survival time (low risk group). The threshold is set to the mean hazard value of the training set. After dividing the dataset into two groups, we used Kaplan-Meier (K-M) curve to visualize the probability of patient being alive as time increases. If the two K-M curves in different groups are far from each other, the Cox model is effective. Fig. 3 (a, b) showed the K-M curves on the training and

TABLE I

C-INDEX ON THE LABELED DATASET (N=129) USING OUR PROPOSED DEEP LEARNING FEATURES AND TRADITIONAL RADIOMIC FEATURES.

	Deep learning feature		Radiomic feature	
	Training (n=65)	Testing (n=64)	Training (n=65)	Testing (n=64)
C-Index (95% CI)	0.71 (0.68,0.74)	0.70 (0.68,0.72)	0.71 (0.68,0.74)	0.62 (0.59,0.65)
AUC (95% CI)	0.71 (0.67,0.75)	0.71 (0.68,0.74)	0.70 (0.64,0.72)	0.61 (0.60,0.70)
Accuracy (95% CI)	0.80 (0.75,0.85)	0.75 (0.71,0.79)	0.71 (0.67,0.75)	0.58 (0.53,0.63)



(a) Deep learning feature (b) Deep learning feature

(c) Radiomic feature (d) Radiomic feature

Fig. 3. Kaplan-Meier curves in the training and testing sets. (a) and (b) are the K-M curves in the training and testing sets using deep learning feature. (c) and (d) are the K-M curves in the training and testing sets using radiomic feature.

testing sets using deep learning features; while Fig. 3 (c, d) illustrated the K-M curves on the training and testing sets using radiomic features. To quantitatively analyze the performance, we used log-rank test to compute the survival time difference between two groups. On the training set, both the deep learning feature and radiomic feature have significant difference ($p<0.001$), however, the deep learning feature showed better performance on the testing set ($p = 0.005$) than the radiomic feature ($p = 0.01$).

IV. CONCLUSION

In this paper, we proposed a method for lung cancer overall survival analysis which utilized unlabeled images for feature learning. Deep learning shows good performance in many medical image analysis tasks, however, they are not well-applied in survival analysis because of the limited data amount. Although it's difficult to acquire images with survival time, it's relatively easy to collect lung cancer images without follow-up information. We proposed a residual convolutional auto encoder to learn lung cancer-related features from unlabeled data. Afterwards, these features were used for supervised survival time analysis. We used Cox proportional hazards model and Kaplan-Meier analysis to test the performance of our proposed feature. Table I and Fig. 3 indicated that our proposed features were effective in overall survival analysis of lung cancer and showed better performance than traditional image features (radiomic features).

REFERENCES

[1] R. Siegel, K. Miller, and A. Jemal, "Cancer statistics, 2016," *CA-Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, 2016.

[2] J. E. van Timmeren, R. T. Leijenaar, W. van Elmpt, B. Reymen, C. Oberije, R. Monshouwer, J. Bussink, C. Brink, O. Hansen, and P. Lambin, "Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam ct images," *Radiotherapy and Oncology*, 2017.

[3] 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.

[4] C. Shen, Z. Liu, M. Guan, J. Song, Y. Lian, S. Wang, Z. Tang, D. Dong, L. Kong, M. Wang *et al.*, "2d and 3d ct radiomics features prognostic performance comparison in non-small cell lung cancer," *Translational oncology*, vol. 10, no. 6, pp. 886–894, 2017.

[5] N. E. Breslow, "Analysis of survival data under the proportional hazards model," *International Statistical Review/Revue Internationale de Statistique*, pp. 45–57, 1975.

[6] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, and J. Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Medical Image Analysis*, vol. 40, pp. 172–183, 2017.

[7] 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.

[8] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[9] S. Wang, M. Zhou, O. Gevaert, Z. Tang, D. Dong, Z. Liu, and J. Tian, "A multi-view deep convolutional neural networks for lung nodule segmentation," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017, pp. 1752–1755.

[10] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[12] V. G. Jjm, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, B. T. Rgh, J. C. Fillion-Robin, S. Pieper, and A. Hjwl, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, p. e104, 2017.

[13] Y. Huang, Z. Liu, L. He, X. Chen, D. Pan, Z. Ma, C. Liang, J. Tian, and C. Liang, "Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (i or ii) nonsmall cell lung cancer," *Radiology*, vol. 281, no. 3, pp. 947–957, 2016.

[14] Z. Liu, X. Y. Zhang, Y. J. Shi, L. Wang, H. T. Zhu, Z. C. Tang, S. Wang, X. T. Li, J. Tian, and Y. S. Sun, "Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer," *Clinical Cancer Research*, p. clincanres.1038.2017, 2017.