IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 22, 2019, accepted May 26, 2019, date of publication June 3, 2019, date of current version June 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920397

Combining LSTM and DenseNet for Automatic Annotation and Classification of Chest X-Ray Images

FENGQI YAN[1], XIN HUANG[1,2], YAO YAO[1], MINGMING LU[1], AND MAOZHEN LI[3]
[1]Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[2]Software College, Jiangxi Agricultural University, Nanchang 3300029, China
[3]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, U.K.

Corresponding author: Xin Huang (1610466@tongji.edu.cn)

This work was supported in part by the Science and Technology Commission of Shanghai Municipality under Grant 16511102800, and in part by the Fundamental Research Funds for the Central Universities under Grant 22120180117.

ABSTRACT The chest X-ray is a simple and economical medical aid for auxiliary diagnosis and therefore has become a routine item for residents' physical examinations. Based on 40 167 images of chest radiographs and corresponding reports, we explore the abnormality classification problem of chest X-rays by taking advantage of deep learning techniques. First of all, since the radiology reports are generally templatized by the aberrant physical regions, we propose an annotation method according to the abnormal part in the images. Second, building on a small number of reports that are manually annotated by professional radiologists, we employ the long short-term memory (LSTM) model to automatically annotate the remaining unlabeled data. The result shows that the precision value reaches 0.88 in accurately annotating images, the recall value reaches 0.85, and the F1-score reaches 0.86. Finally, we classify the abnormality in the chest X-rays by training convolutional neural networks, and the results show that the average AUC value reaches 0.835.

INDEX TERMS Annotation, deep neural network, DenseNet, long short term memory.

I. INTRODUCTION

Chest X-Ray (CXR) is commonly used for early screening of diseases such as thorax, chest, lung tissue, mediastinum, heart, etc. A professional radiologist can diagnose pneumonia, aortic node protrusion, pleural thickening, pneumothorax and other diseases by observing CXR. The data reveals that the number of physical examinations in China exceeded 300 million in 2013 [1]. A large 3A hospital can perform a number of 40,000 CXRs for outpatients alone every year, and the number continues to increase annually. However in China, the imbalance of medical resources is quite serious. The number of radiologists is in short supply due to the long period of cultivation. This situation is not predicted to significantly improve in the next ten years. At the same time, working as a radiologist is labor-intensive, there is a relatively low level of diagnosis in unprivileged areas, and misdiagnoses are frequent due to excessively heavy work and inadequate diagnostic capacity. According to [2], a lung nodule diagnosis was missed or misdiagnosed 20% to 50% of the time in CXRs Even the most professional and brilliant radiologists make serious clinical mistakes in 3% to 6% of cases [3]. Therefore, it is of strong practical significance to alleviate radiologists' work, assist their diagnoses and reduce missed diagnosis or misdiagnosis as much as possible by the existing AI technology using CXRs.

At present, the study on CXR-assisted diagnosis with deep learning techniques mainly focuses on Chest X-Ray14 [4] and Open-i [5] two datasets for their high quality of disease labeling information, which is crucial to deep learning. In the study of CXR data in China, Candemir *et al.* [6] released the shenzhen dataset, which contains only 336 cases with manifestation of tuberculosis and 326 normal cases, cannot fully embody the importance of CXRs through the research in China. In terms of study on unreleased data sets, as far as we know, Dong *et al.* [7] used clustering to annotate diseases with over 16,000 CXR reports in China and evaluated the performance by multiple CNN models. Dong's work was designed with g foresight, but the following problems still exist. First, the unsupervised learning method such as

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

clustering is obviously not accurate; Secondly, it is not rigorous to rely entirely on computer to understand and classify the text. Third, diagnostic needs for CXR vary with different regions, and the information obtained in CXR can also be different. This situation is also verified when comparing the reports from Chinese and foreign radiologists. China's CXR report focuses more on characterization while the foreign radiologists are more accustomed to pointing out possible diseases.

Based on this, with the fact that radiology reports in China are generally templatized by the aberrant physical regions, we propose an annotation method according to the abnormal part in the images. Therefore, the annotation is taken as a text classification task instead of information extraction problem. First we have a small part of reports manually annotated by professional radiologists, and then exploit the semi-supervised learning to automatically annotate the remaining unlabeled data.

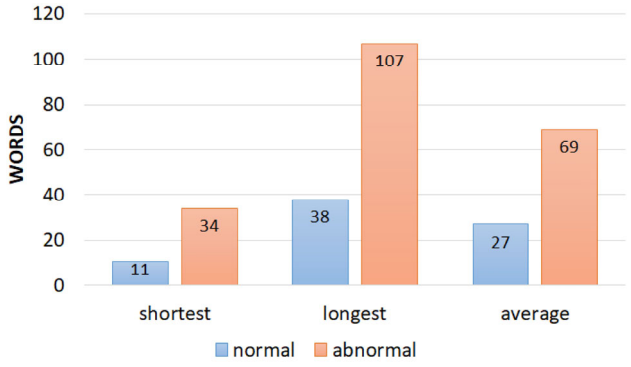


FIGURE 1. Normal report and abnormal report text statistics.

Fig. 1 statistics show that the abnormalities has a correlation with the length of report text, so we have the reports grouped by length and evenly distributed through the process of selecting the manual labeling target and the weak supervised learning. Drawing on the excellent experience of previous CXR image classification [24], we use convolutional neural networks to classify the labeled data, and the test shows a good result.

The major contributions of the paper are as follows:

(1) It presents a methodology for annual annotation of X-ray images.

(2) Building on the manually labeled images, this paper presents a LSTM based machine learning model for automatic annotation of X-ray images. The average F1-score in accurately annotating images reaches 0.86, which is 0.25 higher than the K-Medoids.

(3) It further employs DenseNet for classification of abnormal X-ray images with an average AUC value of 0.835.

The reminder of the paper is organized as follows. Section 2 reviews some related work, Section 3 presents manual annotation rules, automatic annotation models, semi-supervised learning methods and abnormality classification models. Section 4 analyses the evaluation results. Section 5 concludes the paper and points out some future work.

II. RELATED WORK

A. TEXT CLASSIFICATION

Text classification has always been a hot topic in deep learning. Joulin *et al.* [8] used a model similar with word2vec [9] to classify text and the n-gram feature was used when considering the order of words. The TextCNN model proposed by Kim [10], with CNN as the basic model, achieved good performance on classification through convolution, pooling and other operations. Lai *et al.* [11] proposed to obtain representations of each word's context on the basis of word embedding, and then merge them together as a representation of the word. The HAN model proposed by Yang *et al.* [12] used the Attention mechanism and document-level classification to improve performance. Henaff *et al.* [13] proposed a new dynamic memory network that uses fixed-length memory cells to store entities. Multiple memory cells are independent of each other, which can be applied in many fields of text processing.

B. SEMI-SUPERVISED DEEP LEARNING

Lee [14] used the idea of Pseudo label to modify the loss function and added unlabeled data incrementally to achieve a regularization effect. Temporal ensembling [15] is the evolution of Pseudo, it constructed a better target through data enhancement and regularization integration. Johnson and Zhang [16] used Local Region Convolution to learn the two-view (TV) Embedding feature in unmarked text, and then used convolutional neural networks for classification. Then Johnson and Zhang [17] extended the algorithm and used LSTM to perform variable text feature learning. Rasmus *et al.* [18] added a short-circuit connection between the coding layer and the decoding layer of the auto-encoder, and then used classifiers to classify the features learned from the encoder. Dai and Le [19] spliced the self-encoders in order, and learned the hidden features of the sequence data by minimizing the reconstruction errors of these self-encoders.

C. LEARNING TO READ CHEST X-RAYS

In the automatic labeling of reports, Demner-Fushman *et al.* [20] manually labeled 3,955 reports from Indiana University by creating a small controlled vocabulary. Subsequently, Demner-Fushman *et al.* [21] automatically generated annotations using different annotation tools such as MTI and SGindexer, which achieved good results. Hassanzadeh *et al.* [22] compared the effects of different automatic annotation tools on electronic medical records. Mostafiz and Ashraf [23] first entity-marked the report and obtained better results than traditional medical annotation tools through supervised learning. However, these studies are based on the only publicly reported Open-i dataset. Wang *et al.* [4] released the Chest X-Ray14 dataset by means of text mining, but only included images and labels,

and the corresponding report was not released, which is a pity.In the diagnostic imaging of CXRs, Rajpurkar *et al.* [24] diagnosed pneumonia on the Chest X-Ray14 dataset with an accuracy rate of 88.87%, exceeding the human average. Wang *et al.* [25] attempted to generate an inspection report directly on the basis of CXR; Shin *et al.* [26] conducted research on labeling and classification in Open-i datasets. The latest research is not just staying on the choice of models, Kumar *et al.* [27] studied the most suitable loss function for the classification of chest disease and proposed an enhanced cascade network. Guan *et al.* [28] proposed a new AG-CNN network based on the distribution of case regions. The network consists of global branches and local branches. Experiments showed that it got better scores in the disease classification at the edge of ''Hernia''. Baltruschat *et al.* [29] fully considered the effects of non-image features on disease classification, and added features such as angle and gender to the model. The latest research have shown good results, but it is undeniable that due to the lack of information in the dataset itself and the differences in labeling, there is still a long way to go in the research of chest disease classification.

III. DENSENET BASED CLASSIFICATION

A. MANUAL ANNOTATION

The first step in this task is to select the data for manual labeling. We found that the normal part of the description in the report is similar in text. Meanwhile the longer the sentence is, the more likely it is to describe the abnormality. Through the analysis of the template, we found that the main descriptions in the report were centered on six parts known as thoracic, lung, aorta, heart, diaphragmatic surface and partition angle. Therefore, we selected these six parts as the main research objects, except for the abnormalities in the six major parts, including the stomach and the clavicle, we marked them as other parts. In the process of selecting manual labeling data, we divided all the data into 5 groups according to the length of the finding sentence, and selected the manual target in different groups. We have manually labeled 2,500 reports, accounting for 2.3%, 5.5%, 7.7%, 9.7% and 10.7%, respectively. Table 1 shows the details of the grouping.

TABLE 1. Manually annotated grouping information.

	1	2	3	4	5
word	11-30	31-44	45-60	60-80	>80
total	17356	9024	7816	6216	3755
selected	400	500	600	600	400
ratio	2.3%	5.5%	7.7%	9.7%	10.7%

In the process of manual labeling, we only considered whether there is an abnormality, the severity of the abnormality and the characterization of different anomalies were not considered. Specifically, for description ''a little blurry shadow in the bottom left of lung'' and ''dense shadow in the bottom right of lung'', both of them were annotated as ''Abnormalities in both lungs''. In the same way, when there are two pathologically different symptoms as ''right phrenic surface blur'' and ''right phrenic surface uplift'', we annotated ''Abnormalities in phrenic surface'' on both. Table 2 shows the number of abnormal parts after manual statistics we have counted. Among them, the number of abnormal lungs was the highest, reaching 1,489, accounting for 59.6% of all manual markers (not only pulmonary abnormalities). These hand-labeled reports will be the seeds of semi-supervised learning.

TABLE 2. Manually annotated numbers of abnormal objects.

normal	thoracic	lung	aorta
387	29	1489	751
heart	diaphragmatic	partition	others
487	621	508	67

B. ANNOTATION MODEL BASED ON LSTM

We designed a text classification model based on Long Short Term Memory (LSTM) for annotation. The model consists of the Embedding layer, LSTM layer, the Pooling layer and the full linear layer. The model structure is shown in Fig 2.

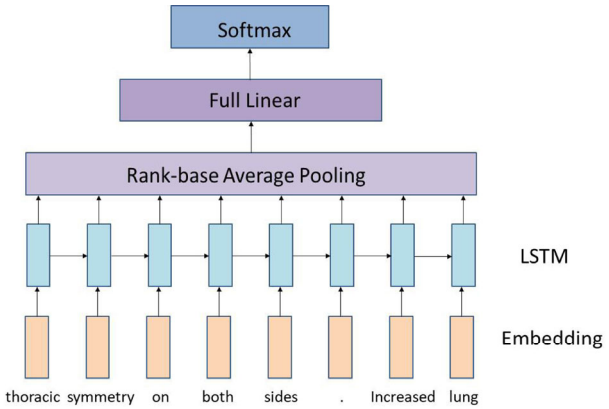


FIGURE 2. Annotate model architecture.

The first layer of the model is the Embedding layer, which is used to construct a thesaurus of high-frequency terms with N words. The model combines the word vector of each word in the thesaurus, then maps each text into an N-dimensional vector. Each dimension of the vector is represented by the maximum value of the similarity between each word corresponding to the dimension and the text in the thesaurus. We use R to represent the thesaurus of the text vector, $R = [r_1, r_2, \ldots r_i, \ldots, r_n]$,$r_i$ represents the word vector of the i^{th} word in the thesaurus. Q represents the text vector after the word segmentation, $Q = [q_1, q_2, \ldots q_j, \ldots, q_m]$,$q_j$, where qj represents the jth text vector after word segmentation

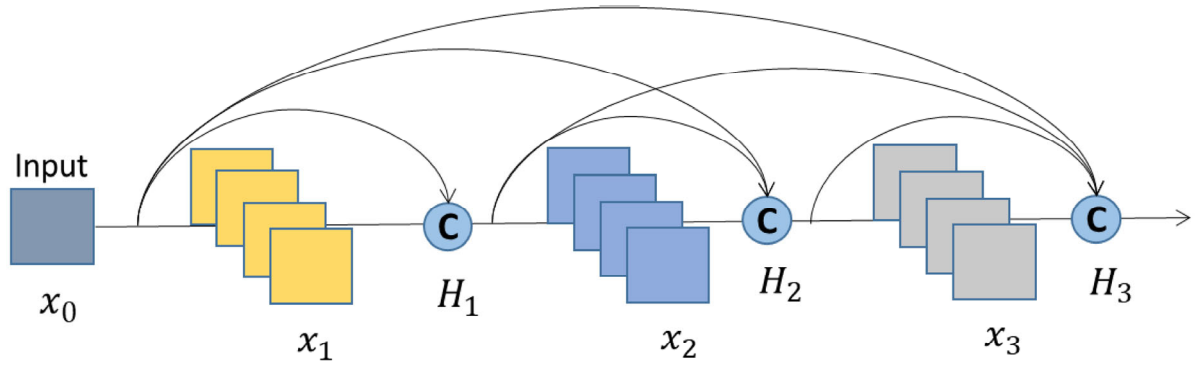


FIGURE 3. DenseNet network architecture diagram.

The second layer of the model is the LSTM layer. LSTM is a time recurrent neural network suitable for processing and predicting important events with relatively long intervals and delays in time series. The core of LSTM lies in the cell design. There are three gates in one cell, known as the input gate, the forgetting gate and the output gate. A message in the LSTM network can be judged if available according to the rules. Only information that complies with the algorithm's certification can be retained, otherwise it will be discarded by the forgotting gate. In our model, the LSTM layer input is the output of the Embedding layer. The calculations of the LSTM layer are shown in Equations 1,2, and 3.

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ u_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot \begin{bmatrix} s_{t-1} \\ y_{t-1} \\ b \end{bmatrix} + \begin{bmatrix} b_i \\ b_f \\ b_o \end{bmatrix} \tag{1}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot u_t \tag{2}$$

$$s_t = o_t \cdot \tanh(c_t) \tag{3}$$

where i_t,f_t,o_t are input gates, forgetting gates and output gates respectively. c_t is LSTM's internal cell. W is the weight matrix. $\cdot$ refers to element multiplication.

The third layer of the model is the Pooling. Considering the serialization characteristics of LSTM, we use the Rank-based Average Pooling (RAP) method. RAP can be averaged by taking the activation values of the top t values. In the pooling domain, the weight coefficients of the first t activation values are set to $1/t$, and the others are set to 0. Therefore, the results of the pooling can be obtained by Equation 4:

$$rap_j = \frac{1}{t} \sum_{i \in R_j, r_i \leq t} \alpha_i \tag{4}$$

where t represents the priority threshold for selecting the activation value to start pooling. R_j represents the pooling domain of the j^{th} feature, while i represents the index value of the activation value within this pooling domain. r_i and α_i respectively represents the priority and activation values of the activation value i. The last two layers of the model are the fully connected layer and the output layer. We use softmax as the activation function for the model output to predict the category of text, then output the result. Softmax maps the output of the fully connected layer to the interval (0,1), and generates the probability of each category. Equation 5 shows the calculation of the softmax activation function. We use the 7-dimensional vector $L = [l_1, l_2, \ldots, l_c]$ to represent the label, where $l_c \in 0, 1$, $c = 7$. l_c indicates whether there is an exception, 1 indicates existence, and 0 indicates no existence.

$$P(y = j|x) = \frac{e^{x^T w_j}}{\sum_{c=1}^{7} e^{x^T w_j}} \tag{5}$$

where $P(y = j|x)$ indicates that when the input is x, the probability of being predicted as category j is P. Obviously the sum of each category's probability is 1.

C. ABNORAML CLASSIFICATION MODEL

DenseNet [32] is used as the base model, which was proposed by Huang et al. in 2017. The core idea of DenseNet is feature reuse. Compared to traditional networks, DenseNet has a novel Dense Block module. In the Dense Block module, each layer can be directly connected. This connection enhances the reuse of features, so that the final classifier makes decisions based on all the features of the entire network. Fig 3 shows the structure of the Dense Block.

The input picture of the model is x_0, the model is composed of L layers, the nonlinear conversion function of each layer is $H_l()$, and l^{th} is the serial number of the layer. The output of the l^{th} layer is recorded as x_l. Equation 6 shows the x_l calculation method.

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \tag{6}$$

DenseNet confirmed in a previous study by Rajpurkar *et al.* [24] that it has a good effect on the classification of X-ray chest disease. So we made three changes to the basic DenseNet model with reference to Rajpurkar's method. First, the output of the model is modified to the vector c of the binary label, indicating whether the following seven exceptions exist. Second, we replace DenseNet's original fully connected layer with a fully connected layer which has a 7-dimensional vector output, and classify it with the sigmoid function. Third, the model's loss function is modified to be a

unweighted binary cross entropy. The modified loss function is shown in the formula 2:

$$L(X, y) = \sum_{c=1}^{7}[-y_c \log p(Y_c = 1|X) - (1 - y_c) \log p(Y_c = 0|X)] \quad (7)$$

where $p(Y_c = 1|X)$ is the predicted probability that abnormality c is contained, and $p(Y_c = 0|X)$ is the predicted probability that abnormality c is not included.

IV. EXPERIMENTAL

A. DATA COLLECTION AND PREPROCESSING

1) DATA COLLECTION

We obtained 46,711 CXRs and 42,316 copies of reports from the Picture Archiving Communication System of Tongji Affiliated Hospital, where each report has at least one CXR. CXRs are in DICOM format [31], and the reports are written in Chinese and all of them are confirmed by a peer reviewer. In the reports, the Finding section records the description of the CXR, and we select this part to generate the CXR label. At the same time, we preprocess the obtained data.

2) PREPROCESSING

First, we removed the CXRs on the lateral and selected the positive CXRs as the research object. For a report containing multiple positive CXRs, we chose the one with the highest image quality. Then, we strip out the report with less than 10 words in the Finding section. Third, we segmented the text with a word segmentation tool called jieba.[1] Finally, we obtained 40,167 copies of CXRs and reports which was one-to-one correspondence. Fig 5 is an example of CXR and report. Secondly, we convert the original X-ray chest of Dicom format to the PNG image format, and adjust the converted image size to 256*256. Then all the data is divided into training/validation/test according to the ratio of 80%/10%/10%, with the distribution of abnormal data of different parts of the validation set and the test set is guaranteed. Finally we get a training set with 32167 X-ray films, a validation set with 4000 and a test set also with 4000 films.

B. BASELINE AND DETAILS

1) UNSUPERVISED BASELINE

For the Unsupervised annotation baseline, we refer to the method of [7]. First, define the similarity of clauses based on the edit distance [35]. The edit distance is defined by the minimum operations (insert, delete, and replace) that convert one clause to another.At the same time, the k-medoids algorithm [36] is used to perform clustering on clauses. K-medoids is related to the k-means [37] algorithm and selects points in the dataset as cluster centers.

[1] https://github.com/fxsjy/jieba

2) TRAINING OF SEMI-SUPERVISED

The entire process is divided into six steps. Step 1, The manually labeld data is divided into two parts: train set and validation set; Step 2, Use the annotate model to train the optimal parameters; Step 3, Predict the unlabeled data using the model trained in step 2, The label predicted is called pseudo-labeled; Step 4, Extract a part of the training set to make a new validation set, make sure validation comes for original train; Step 5, Merge the rest of the original train set with the pseudo-labeled part into a new train set, and use the annotate model again to train the optimal parameters; Step 6, Use the training model in step 5 to predict the unlabeled data to get the final result label. Fig 4 shows the flow chart of the proposed annotation model.

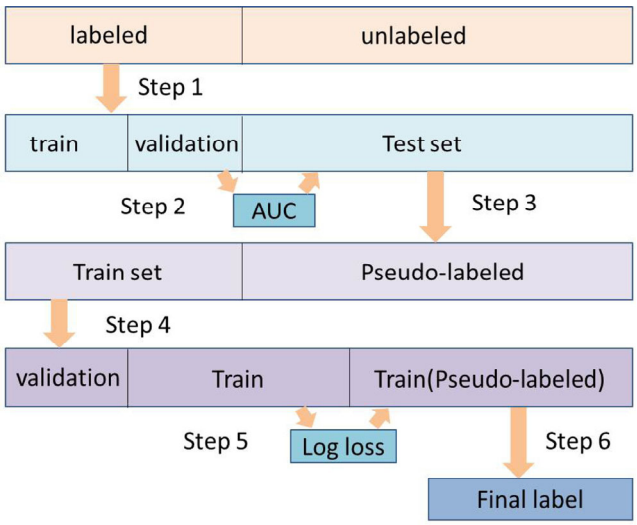


FIGURE 4. Training process of annotation model base on semi-supervised.

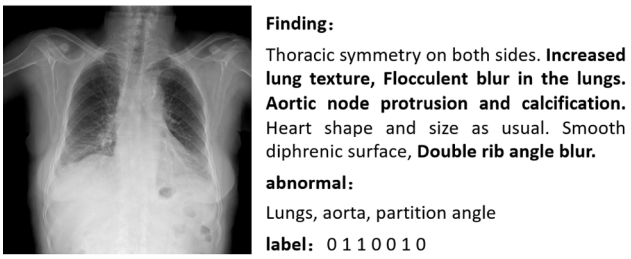


FIGURE 5. A sample of CXR.

3) TRAINING OF DENSENET

We trained DenseNet end-to-end utilizing Adam [33] with standard parameters ($\beta 1 = 0.9$ and $\beta 1 = 0.999$) by using mini batches of size 32. We adopted an initial learning rate of 0.001 which is delayed by a factor of 10 every time the validation loss plateaued after an epoch and chose the model with the least validation loss. Also, the GPU we adopted for training were two Tesla K40 launched by NVIDIA.

TABLE 3. Evaluation of image labeling results on dataset. Performance is reported using P, R, F1-score.

	Semi-Supervised			K-Medoids		
	precision	recall	F1-score	precision	recall	F1-score
normal	0.91	0.86	**0.89**	0.83	0.80	0.81
thoracic	0.80	0.84	**0.81**	0.66	0.58	0.61
lungs	0.94	0.82	**0.90**	0.56	0.48	0.54
aorta	0.96	0.79	**0.87**	0.74	0.71	0.72
heart	0.93	0.90	**0.91**	0.78	0.73	0.75
diaphragmatic	0.86	0.88	**0.87**	0.58	0.56	0.56
partition	0.84	0.89	**0.85**	0.66	0.57	0.59
others	0.63	0.81	**0.77**	0.37	0.26	0.33
average	0.88	0.85	**0.86**	0.65	0.52	0.61

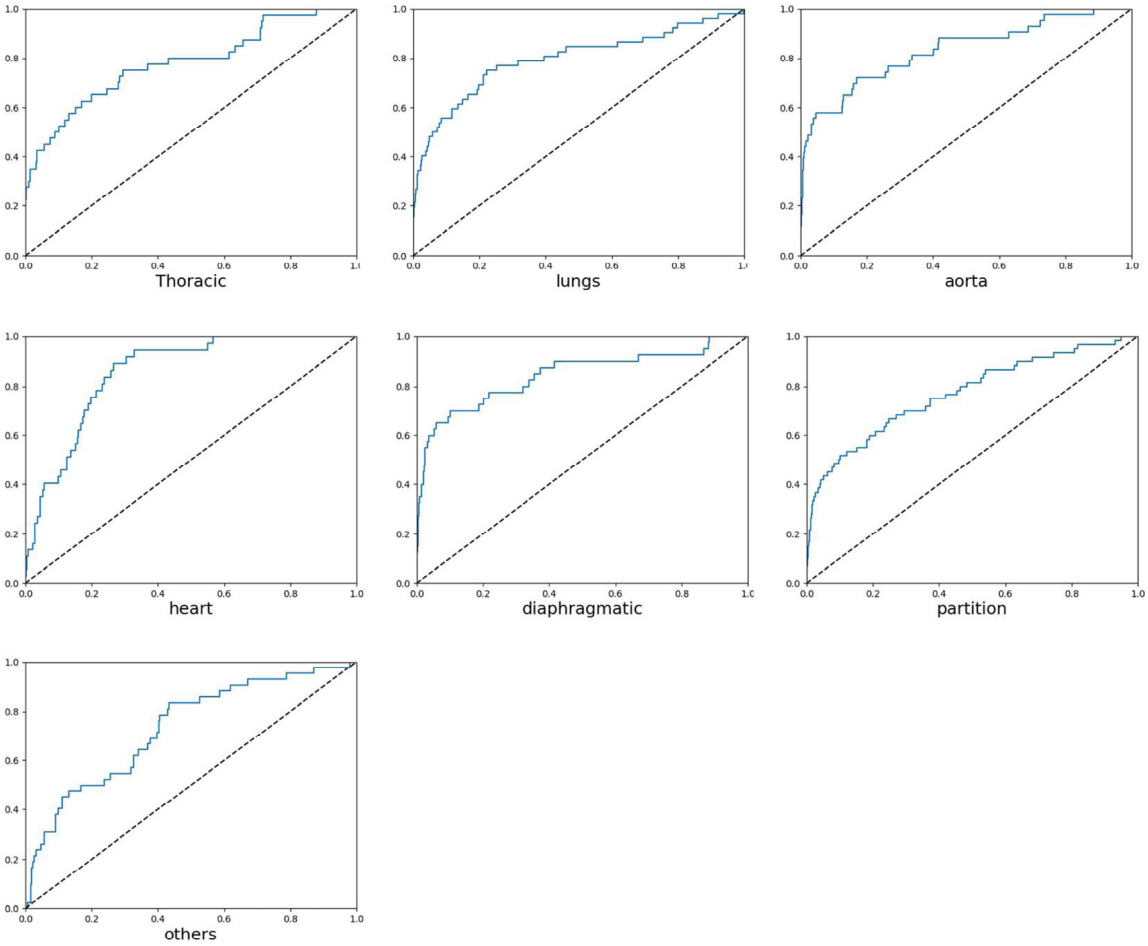


FIGURE 6. The ROC curve of abnoraml classification.

C. RESULTS

1) ANNOTATION RESULTS

We present results of our annotation experiments in Table 3, which evaluated by precision, recall and F1-score. The results of using annotation model base on semi-supervised are all better than the unsupervised method. This can be understood, and the K-Medoids method using unsupervised learning can only be clustered by the edit distance of the clause. The limitation of this method is that the effect on short sentences is significantly better than that on long sentences. This has been verified in our experiments, and the better-performing normal clause has fewer description words. Using the semi-supervised learning annotation method, guided by a professional radiologist's manual labeling, the average F1-score can reach 0.86, which is 0.25 higher than K-Medoids.

2) MULTIPLE DISEASE DETECTION RESULTS

The evaluation results show that the seven abnormalities of our models have achieved a high level of accuracy.

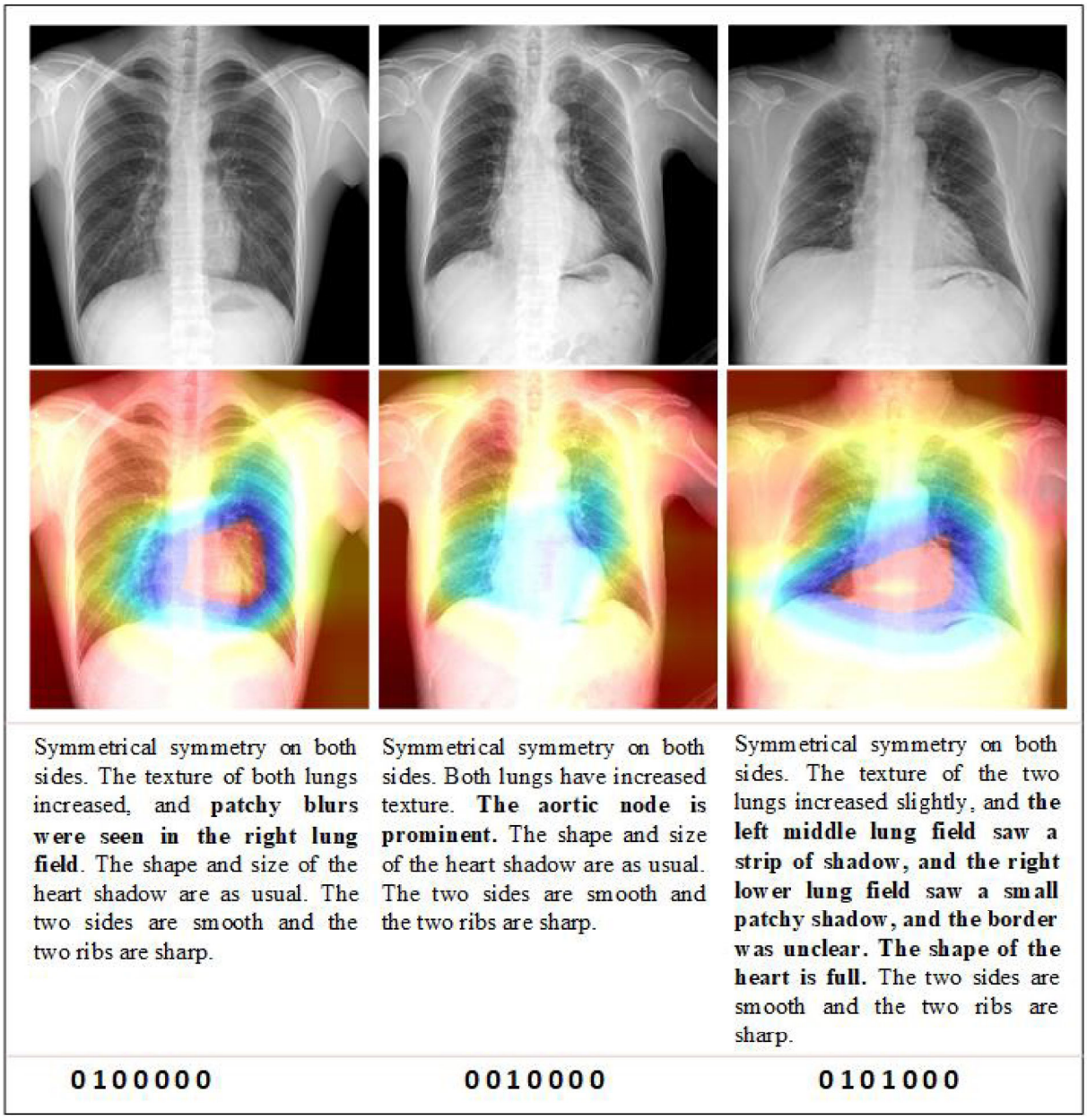


FIGURE 7. Comparison of heat map and inspection report.

Among them, the aorta, heart and lungs provided the best result, with precisions of 0.96, 0.93 and 0.94, respectively. For all samples, average precision, recall and F1-score are 0.88, 0.85, and 0.86 respectively.

Tabel 4 shows the AUC value of classification on CXR abnormalities by DenseNet, and the horizontal direction of the table is the comparison between different pre-trainings of the same part. By comparison, we discover that the selection of pre-trained data sets has significant differences for the final classification results. The pre-training of the large image dataset ImageNet did not result in a significant improvement in the classification performance, while with the pre-training on Chest X-14, the average AUC value increased by about 4%. The column of the table shows the classification effect of different parts in the same pre-training scene. Overall, the ''other parts'' gave the worst result with the lowest AUC. This is because we marked all the other factors such as ribs, clavicles, stomach and even surgical externalities as ''other parts'', and these features are not similar. Compared with the other three parts, the ''thoracic'' and ''lungs'' have significantly lower AUC values. The average AUC values of the three different pre-training for the ''thorax'' are 0.752, 0.74 and 0.804, respectively. The average AUC values for the three different pre-training exercises for ''lungs'' were 0.749, 0.765 and 0.826, respectively. The number of samples with abnormalities only in ''thoracic'' was the

TABLE 4. The AUC of abnoraml classification.

	None	ImageNet	Chest X-14	average
thoracic	0.752	0.745	0.804	0.767
lungs	0.749	0.765	0.826	0.780
aorta	0.831	0.842	0.898	0.857
heart	0.815	0.859	0.867	0.847
diaphragmatic	0.809	0.826	0.827	0.821
partition	0.791	0.808	0.844	0.814
others	0.724	0.711	0.782	0.739
average	0.782	0.794	0.835	0.804

least in dataset, so it was difficult to train the model in the case of a small range of samples. Figure 6 shows the ROC curve on the testing set.

D. DISCUSSION

The ''lungs'' with the largest sample size also gave a low AUC value. The reason is that although there are more abnormalities in the ''lungs'', the abnormal parts are very complicated since the ''lungs'' make up the largest percentage of the entire X-ray. However, we did not distinguish the location of abnormalities when annotating X-ray films. For ''a little patchy blur in the bottom left lung'', ''small nodule dense shadow in the right upper lung'' we annotated both of them as an abnormality in the ''lungs'', which also made the classification difficult. The ''aortic'' has the highest AUC since it has as many abnormalities as the lungs. In addition, the ''aortic'' is a rather small area and the abnormality is relatively simple, thus achieved the best result.

We use heatmap to further understand and evaluate the performance of models on different diseases' classification. Fig. 7 shows the results. The first row in the figure is the original image of the chest X-ray, the second row is the generated heatmap, and the third row is the original report description corresponding to the X-ray chest. We found that the abnormality of the ''lung'' part is indeed wider, which confirms the reason for the low AUC value in Table 2. We also found that when a single abnormality occurs, the effect of the model is very obvious, while it performs worse when multiple abnormalities are accompanied by overlapping parts.

V. CONCLUSIONS

This paper is based on the real situations of radiologists' diagnosis with X-ray films in China. Through the analysis of chesX-rays and the corresponding report, we proposed an annotation method according to the abnormal part in the images, and automatically generated the training label by semi-supervised learning. According to the increasing characteristics of the training set in semi-supervised learning base on professional radiologist's manual labeling, the average F1-score can reach 0.86, which is 0.25 higher than K-Medoids. At the same time, we compare the AUC values of the three different pre-training methods for the abnormality classification between ImageNet and ChestX-14. By comparison, we discover that the model using the pre-training result of the large dataset ChestX-14 is significantly better than other datasets. We found that the AUC values of ''thoracic'' and ''lung'' were significantly lower than the other three parts. The reason was that the number of ''thoracic'' samples was small, which made the model training difficult. The abnormal parts of the ''lungs'' are also complicated, the distribution area is wide, and it is easy to overlap with other parts. This also points out the direction for future research. One direction is to achieve a finer granularity in annotation in different areas of the lungs and complement the analysis of overlap area cutting. Another direction is to use the target detection methods [34] in other fields to apply migration learning to chest disease detection.

REFERENCES

[1] *Health Statistical Bulletin*, (in Chinese). Accessed: Jun. 10, 2019. [Online]. Available: http://www.nhc.gov.cn/guihuaxxs/s10742/201405/886f82dafa344c3097f1d16581a1bea2.shtml
[2] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, Oct. 2012, pp. 590–598.
[3] M. A. Bruno, E. A. Walker, and H. H. Abujudeh, "Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction," *Radiographics*, vol. 35, no. 6, pp. 1668–1676, 2015.
[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2097–2106.
[5] *Indiana University Chest X-Rays*. Accessed: Jun. 10, 2019. [Online]. Available: https://openi.nlm.nih.gov/gridquery?it=xg&coll=cxr&m=1&n=100
[6] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, Feb. 2014.
[7] Y. Dong, Y. Pan, J. Zhang, and W. Xu, "Learning to read chest X-ray images from 16000+ examples using CNN," in *Proc. 2nd IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol.*, Jul. 2017, pp. 51–57.
[8] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," 2016, *arXiv:1607.01759*. [Online]. Available: https://arxiv.org/abs/1607.01759
[9] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[10] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882
[11] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. 29th AAAI Conf. Artif. Intell.*, Feb. 2015, pp. 1–7.
[12] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.
[13] M. Henaff, J. Weston, A. Szlam, A. Bordes, and Y. LeCun, "Tracking the world state with recurrent entity networks," 2016, *arXiv:1612.03969*. [Online]. Available: https://arxiv.org/abs/1612.03969
[14] D. H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn. (ICML)*, vol. 3, Jun. 2013, p. 2.
[15] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*. [Online]. Available: https://arxiv.org/abs/1610.02242
[16] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 919–927.

[17] R. Johnson and T. Zhang, "Supervised and semi-supervised text categorization using LSTM for region embeddings," 2016, *arXiv:1602.02373*. [Online]. Available: https://arxiv.org/abs/1602.02373

[18] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3546–3554.

[19] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3079–3087.

[20] D. Demner-Fushman, M. D. Kohli, M. B. Rosenman, S. E. Shooshan, L. Rodriguez, S. Antani, G. R. Thoma, and C. J. McDonald, "Preparing a collection of radiology examinations for distribution and retrieval," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 2, pp. 304–310, 2015.

[21] D. Demner-Fushman, S. E. Shooshan, L. Rodriguez, S. Antani, and G. R. Thoma, "Annotation of chest radiology reports for indexing and retrieval," in *Proc. Int. Workshop Multimodal Retr. Medical Domain*. Cham, Switzerland: Springer, Mar. 2015, pp. 99–111.

[22] H. Hassanzadeh, A. Nguyen, and B. Koopman, "Evaluation of medical concept annotation systems on clinical records," in *Proc. Australas. Lang. Technol. Assoc. Workshop*, 2016, pp. 15–24.

[23] T. Mostafiz and K. Ashraf, "Pathology extraction from chest X-ray radiology reports: A performance study," 2018, *arXiv:1812.02305*. [Online]. Available: https://arxiv.org/abs/1812.02305

[24] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning." 2017, *arXiv:1711.05225*. [Online]. Available: https://arxiv.org/abs/1711.05225

[25] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "TieNet: Text-image embedding network for common thorax disease classification and reporting in chest X-rays," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9049–9058.

[26] H. C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2497–2506.

[27] P. Kumar, M. Grewal, and M. M. Srivastava, "Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 546–552.

[28] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," 2018, *arXiv:1801.09927*. [Online]. Available: https://arxiv.org/abs/1801.09927

[29] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," 2018, *arXiv:1803.02315*. [Online]. Available: https://arxiv.org/abs/1803.02315

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] C. Parisot, "The DICOM standard," *Int. J. Cardiac Imag.*, vol. 11, no. 3, pp. 171–177, 1995.

[32] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

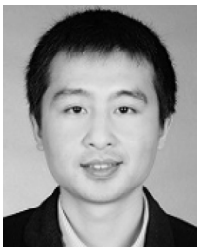
[34] M. Xu, H. Fang, P. Lv, L. Cui, S. Zhang, and B. Zhou, "D-STC: Deep learning with spatio-temporal constraints for train drivers detection from videos," *Pattern Recognit. Lett.*, vol. 119, pp. 222–228, Mar. 2019.

[35] G. Navarro, "A guided tour to approximate string matching," *ACM Comput. Surv.*, vol. 33, no. 1, pp. 31–88, 2001.

[36] L. Kaufman and P. J. Rousseeuw, "Clustering by means of medoids," Dept. Math. Inform., Delft Univ. Technol., Delft, The Netherlands, Tech. Rep., 1987, vol. 87003.

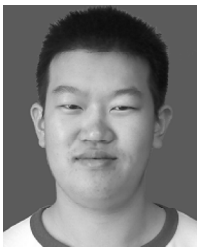
[37] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *J. Roy. Stat. Soc. C (Appl. Statist.)*, vol. 28, no. 1, pp. 100–108, 1979.

FENGQI YAN was born in 1978. He received the M.S. degree from the Shandong University of Science and Technology, in 2007. He is currently pursuing the D.Eng. degree in electronics and information with Tongji University, China. His research interests include medical big data and medical information services.



XIN HUANG was born in 1984. He received the M.S. degree from Nanchang University, in 2010. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, Tongji University, China. He is currently with the Software College, Jiangxi Agricultural University, Nanchang, China. His research interests include image processing, data fusion, and machine learning.

YAO YAO received the bachelor's degree in computer science and technology from Tongji University, in 2017, where she is currently pursuing the M.S. degree. Her main research interests include natural language processing and text retrieval serving for medical field.



MINGMING LU was born in 1991. He received the bachelor's degree from the China University of Mining and Technology, in 2013. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, Tongji University. His main research interests include machine learning, natural language processing, and intelligent systems with applications to start medicine.

MAOZHEN LI received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences, in 1997. He is currently a Professor with the Department of Electronic and Computer Engineering, Brunel University London, U.K. He has over 160 research publications in these areas, including four books. His main research interests include high-performance computing, big data analytics, and intelligent systems with applications to smart grid, smart manufacturing, and smart cities. He has served over 30 IEEE conferences. He is a Fellow of the British Computer Society and the IET. He serves on the Editorial Board of a number of journals.