Pulmonary Nodule Classification Based on Heterogeneous Features Learning

Chao Tong, Baoyu Liang, Qiang Su, Mengbo Yu, Jiexuan Hu, Ali Kashif Bashir, and Zhigao Zheng,

Abstract—**Pulmonary cancer is one of the most dangerous cancers with the a high incidence and mortality. An early accurate diagnosis and treatment of pulmonary cancer can observably increase the survival rates, where computer-aided diagnosis systems can largely improve the efficiency of radiologists. In this paper, we propose a deep automated lung nodule diagnosis system based on three-dimensional convolutional neural network (3D-CNN) and support vector machine (SVM) with multiple kernel learning (MKL) algorithms. The system not only explores the computed tomography (CT) scans, but also the clinical information of patients like age, smoking history and cancer history. To extract deeper image features, a 34-layers 3D Residual Network (3D-ResNet) is employed. Heterogeneous features including the extracted image features and the clinical data are learned with MKL. The experimental results prove the effectiveness of the proposed image feature extractor and the combination of heterogeneous features in the task of lung nodule diagnosis.**

Index Terms—**Pulmonary nodule classification, lung cancer, heterogeneous features, multiple kernel learning.**

I. INTRODUCTION

PULMONARY cancer is one of the most dangerous cancers with the a high incidence and mortality. It causes 1.8 million deaths around the world [1]. A rising trend in incidence can still be observed in most countries. The early manifestations of lung cancer are generally pulmonary nodules (0-30 mm). The nodules develop into tumors of different sizes (2-50 mm) later. In advanced stage, pulmonary cancer shows different manifestations which brings great difficulties to treatment. Therefore, an early diagnosis and treatment of pulmonary cancer can observably increase the survival rates. In the early diagnosis of pulmonary cancer, lung computed tomography (CT) scanning plays a significant role. It can eliminate the overlap between the anterior and posterior tissues and show the normal anatomical structure and pathological manifestation in the cross section, providing great auxiliary effects for the diagnosis of pulmonary nodules with little risk.

To identify the malignancy of pulmonary nodules, radiologists have to interpret a huge amount of CT images and

C. Tong, B. Liang and M. Yu are with School of Computer Science and Engineering, Beihang Beijing, China, 100191

Q. Su and J. Hu are with Department of Oncology, Beijing Friendship Hospital, Capital Medical University, Beijing, China.

A. K. Bashir is with Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK., School of Electrical Engineering and Computer Science, National University of Science and Technology, Islamabad (NUST), Pakistan

Z. Zheng is with School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China

Corresponding Author: Zhigao Zheng (zhengzhigao@pku.edu.cn)

C. Tong, B. Liang contributed equally to this work.

make quick decisions, which brings about heavy burden [2]. Despite their efforts, the diagnosis results are often constrained by subjectivity. The results vary significantly among different interpreters. To help increase the diagnosis rate, several pulmonary nodule malignancy diagnosis models have been proposed, including the famous Mayo Clinic model [3] and other models [4]–[6]. These methods take features of pulmonary nodules obtained from CT images as well as information of patients including age, smoking history and other facts to identify the malignancy of nodules. They are validated as effective pulmonary nodule diagnosis methods [7] and is widely used in manual lung cancer diagnosis. Nevertheless, none of these methods saves labor.

As a result, a considerable amount of Computer-Aided Diagnosis (CAD) systems have been developed to help radiologists increase the speed and accuracy of lung nodule diagnosis. CAD systems are categorized into detection systems (CADe) and diagnostic systems (CADx) [2], where detection systems are designed to find pulmonary nodules in CT images and diagnostic systems are to identify whether one detected nodule is benign or malignant.

Many researchers have utilized morphologic characteristics of pulmonary nodules to diagnose their malignancy. Kostis *et al.* [8] developed several methods for nodule growth rate estimation in order to identify the malignancy of lung nodules. Kido *et al.* [9] calculated fractal dimensions which reflect the characteristics of the lung-nodule interfaces, to distinguish bronchogenic carcinomas from benign pulmonary nodules. Way *et al.* [10] designed a CADx system which is capable of extracting morphologic features including volume, surface area, perimeter, maximum diameter, HU values and gray-level features including the average, variance, skewness and kurtosis of the gray-level histogram. Among these low-level features, the most important ones are selected with F-statistics [11] for further classification.

In recent years, deep learning, benefiting from the easy access and huge amount of data, has achieved tremendous success in traditional fields as data handling, computer vision, natural language processing. Recent advances in deep learning have facilitated investigation of the application of this technique on diagnosis tasks, especially lung nodule malignancy classification.

Hua *et al.* [12] proposed two models which use a deep belief network (DBN) and a convolution neural network (CNN), respectively, for nodule classification. These two models are compared with two conventional methods with feature computing steps. The experimental results have proved the advantage of using deep learning methods in lung nodule

diagnosis. Kumar *et al.* [13] introduced a CADx system using deep features extracted from a Stacked AutoEncoder (SAE) for nodule classification. The proposed system achieved an overall accuracy of 75.01% with a sensitivity of 83.35%. StochasticNet radiomic sequencers were leveraged by Shafiee *et al.* [14] which consist of three stochastically-formed convolutional layers and achieved an accuracy of 84.49% with a sensitivity of 91.07% and a specificity of 75.98%. Sun *et al.* [15] implemented three deep learning methods, CNN, DBN and stacked denoising autoencoder (SDAE) for nodule classification. They compared the results of these three methods with a traditional CADx system which is made up of a scheme with 28 manually designed features and a support vector machine (SVM). The accuracies of CNN, DBN and SDAE methods are 79.76%, 81.19% and 79.29%. Murillo *et al.* [16] introduced the Structural Co-occurrence Matrix technique for feature extraction and different classifiers including multilayer perceptron, SVM and k-nearest neighbors for classification.

Based on CNN, many researchers have noticed the variety of nodule sizes and morphologies and proposed methods with multi-scale features. Shen *et al.* [17] proposed a multi-scale CNN (MCNN) method which utilizes multi-scale nodule patches to learn a set of class-specific features. The proposed method was evaluated on the dataset of Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) and achieved an accuracy of 86.84%. Later in 2017, they [18] extended their methods from MCNN to multi-crop CNN (MCCNN). MCCNN applies a new strategy based multi-pooling strategy. It was evaluated with 880 benign nodules and 495 malignant nodules from LIDC-IDRI and achieved an accuracy of 87.14%. Juan *et al.* [19] proposed a multi-level convolutional neural network for ternary nodule classification and achieved 84.81% accuracy without hand-craft preprocessing algorithm.

In addition to the deep features extracted by deep learning methods, there are still several work on the combination of deep features and traditional morphologic features of lung nodules. Kim *et al.* [20] combined deep learned representations extracted using an SDAE with 76 raw hand-crafted imaging features and constructed a linear SVM for nodule classification. The result shows its superiority over methods using only the original raw features. Xie *et al.* [21] fused texture, shape and deep model-learned information at decision level for nodule classification.

In both CADe systems [22], [23] and CADx systems, the utilization of three-dimensional CNNs (3D-CNN) is still in an early stage. Hussein *et al.* [24] proposed a 3D-CNN based approach for rich feature representation of lung nodules. Liu *et al.* [25] presented a 3D-CNN for nodule classification and explored the classifier ensembles of 3D-CNN and traditional machine learning models. Kuang *et al.* [26] proposed a 3D multi-view CNN for nodule classification in which a 3D Inception network is utilized.

Studies over the past years has provided important information and experience on automatic lung nodule diagnosis with CT images. While some research has found the positive effects of the combination of deep features extracted by deep learning methods and the morphologic features of pulmonary nodules

in CT images, there have been few empirical investigations into the combination of image features and other features of a patient. These features, including age, smoking history, cancer history and other information, provide considerable reference value in clinical diagnosis of lung nodules. In addition, as CT scans are naturally 3D and the 3D contexts provide abundant information for nodule diagnosis, two-dimensional convolutional neural networks (2D-CNN) can cause dimensional loss and thus fail to fully use the 3D context information. Although several 3D-CNN methods have been proposed to deal with the volumetric representation provided by CT scans, the structure can still go deeper.

In this work, we develop a deep automated CADx system based on 3D-CNN and support vector machine (SVM) with multiple kernel learning (MKL) algorithm to manage heterogeneous features. Not only are the lung nodule CT scans processed in this system, but also some other heterogeneous information of patients like age, smoking history and other features. Experiments are conducted on a private data set whose images and patient information are collected from a hospital. As the data amount in the private data set is not adequate, we use the strategy of transfer learning, firstly pretraining the 3D-CNN model with CT images from the LIDC-IDRI data set and then finetuning the pretrained model with the private data set.

Our main contributions are as follows:

(1) To make fully use of the spatial information of CT scans and to adequately extract deep features, a 3D-ResNet is used for automatic lung nodule binary classification (benign and malignant). The strategy of transfer learning is employed to solve the problem of insufficient data in the private data set. Our model for image feature extraction is pretrained with the famous LIDC-IDRI data set and finetuned to be adjusted to the private data set.

(2) Heterogeneous features including deep features extracted from images and patient information are combined to form a fused description of the object. To the best of our knowledge, this is the first study to introduce into the CADx system the clinical priori knowledge about the impacts of different kinds of information other than morphologic features on nodule diagnosis.

(3) To learn from the heterogeneous features, an SVM with MKL is proposed for nodule classification which is able to learn an optimal linear combination of basis kernel to deal with the heterogeneous features with different notions of similarity.

(4) The experimental results can prove the effectiveness of the combination of heterogeneous features in the task of lung nodule diagnosis.

II. METHODS

IN this section, we present the proposed automated pulmonary nodule diagnosis system based on 3D-CNN and MKL algorithm that combines the heterogeneous features from CT images and the related information of patients including age, smoking history, cancer history in five years, hypertension, heart disease, diabetes, tuberculosis, hepatitis and drinking thistory.

TABLE I
ARCHITECTURE OF THE CHAIN-LIKE 3D-CNN MODEL.

Module	Kernel	Input	Output
3D Convolution	$3 \times 3 \times 3$	$40 \times 40 \times 20 \times 1$	$38 \times 38 \times 18 \times 32$
3D Convolution	$3 \times 3 \times 3$	$38 \times 38 \times 18 \times 32$	$36 \times 36 \times 16 \times 32$
Max Pooling	$2 \times 2 \times 1$	$36 \times 36 \times 16 \times 32$	$18 \times 18 \times 16 \times 32$
3D Convolution	$3 \times 3 \times 3$	$18 \times 18 \times 16 \times 32$	$16 \times 16 \times 14 \times 64$
3D Convolution	$3 \times 3 \times 3$	$16 \times 16 \times 14 \times 64$	$14 \times 14 \times 12 \times 64$
Max Pooling	$2 \times 2 \times 2$	$14 \times 14 \times 12 \times 64$	$7 \times 7 \times 6 \times 64$
3D Convolution	$3 \times 3 \times 3$	$7 \times 7 \times 6 \times 64$	$5 \times 5 \times 4 \times 128$
3D Convolution	$3 \times 3 \times 3$	$5 \times 5 \times 4 \times 128$	$3 \times 3 \times 2 \times 128$
Max Pooling	$2 \times 2 \times 2$	$3 \times 3 \times 2 \times 128$	$1 \times 1 \times 1 \times 128$
FC	\	128	512
FC	\	512	512
FC	\	512	2
Softmax	\	2	2

A. Three-dimensional Convolutional Neural Networks

3D-CNNs are feed-forward neural networks that are used to extract different levels of features of the input image. Compared to 2D-CNNs, 3D-CNNs, whose channels are 3D feature volumes instead of 2D feature maps, are capable of extracting features in a cubic manner and preserving the spatial information of the inputs. CNNs are typically made up of convolutional, activation, pooling and fully-connected layers. These layers are trained in an end-to-end mode to extract features and to make predictions.

3D Convolutional Layers. A 3D convolutional layer provides the 3D convolutional function that extracts the 3D feature volumes using multiple convolution kernels $\mathbf{K}$ with a shape of $C_{out} \times C_{in} \times D_k \times H_k \times W_k$, where C_{in} and C_{out} refers to the number of input channels and output channels (or the amount of kernels), D_k denotes the depth of one kernel while H_k and W_k the height and width of one kernel, respectively. Given an input $\mathbf{x}$ with size $B \times C_{in} \times D \times H \times W$, where B denotes the batch size, C is the number of channels, D represents the depth of the feature volume, and H and W are the height and wdith of the feature volume, respectively. The output of the 3D convolutional can be computed as (1):

$$\mathbf{y}_{i,j} = \mathbf{b}_i + \sum_{k=1}^{C_{in}} \mathbf{K}_{j,k} \star \mathbf{x}_{i,k}, \tag{1}$$

where $\star$ denotes the 3D cross-correlation operator, $\mathbf{b}$ is the bias, i represents the i-th sample among B samples and j is the j-th channel among C_{out} channels. The size of output y should be $B \times C_{out} \times D_{out} \times H_{out} \times W_{out}$. D_{out}, H_{out} and W_{out} are determined according to the parameters in the cross-correlation operator and the kernel size.

Non-linear Activation Layers. To increase the non-linearity of 3D-CNN, after a 3D-CNN layer follows a non-linear activation layer. In this paper, the rectified linear unit (ReLU) is adopted, which can be expressed as (2):

$$\mathbf{z} = \max(0, \mathbf{y}), \tag{2}$$

where y denotes the output of the previous 3D-CNN layer and z is the output of the activation layer.

3D Pooling Layers. A 3D max-pooling layer follows the activation layer to improve the robustness of extracted features
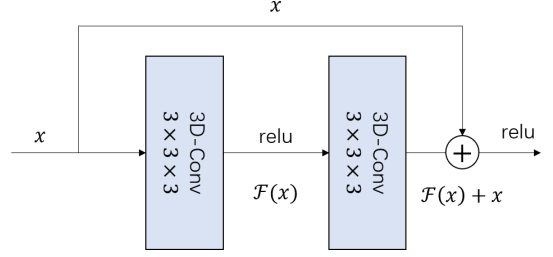

Fig. 1. A residual block.

and to cope with nodule images that are not aligned well. The operation of 3D max-pooling layer can be defined as (3):

$$\mathbf{u}_{i,j,d',h',w'} = \max\{\mathbf{z}_{i,j,d,h,w} : d' \leq d < d' + p, \\ h' \leq h < h' + p, w' \leq w < w' + p\}, \tag{3}$$

where p denotes the pooling size, the meaning of i, j are the same as in (1), d, h, w represent the 3D position of the input and d', h', w' represent the 3D position of the output.

Fully-Connected Layers. A fully-connected layer is composed of neurons each of which is connected with all neurons in the previous layer. It is generally followed by a softmax classifier to make predictions.

B. CNN Architecture

Huge amounts of CNN architectures have been proposed for feature extraction. The structure of traditional CNNs, such as LeNet [27], Alexnet [28] and VGGNet [29], are chain-like, which means that the computational blocks in these structures form a simple chain. [26] In this section, we introduce a chain-like 3D-CNN model which takes the input shape $width \times height \times depth \times channel$ of $40 \times 40 \times 20 \times 1$ for feature extraction. Its architecture is shown in TABLE I.

Besides, it is well-known that the deeper CNN architectures are capable of extracting higher levels of image features. However, as these chain-like networks go deeper, a degradation problem appears. [30] Several works have used some traditional 3D-CNN architectures for lung nodule classification, which indicates that their architectures cannot be so deep. In this paper, we explore a 34-layer 3D residual net (ResNet) [31] for feature extraction of pulmonary nodules. The 3D-ResNet model is composed of 3D residual blocks. A 3D residual block contains several convolutional layers and a shortcut connection. After the convolution operations, the input and the output of a 3D residual block are connected with a shortcut connection skipping the layers in the block to realize an identity mapping. In other words, through the connection, the input signal and the output signal of the convolutional layers are summed at the tail of the residual block. The structure of a residual block is shown in Fig 1, where 3D-Conv represents a 3D convolutional layers with the kernel size of $3 \times 3 \times 3$. In this paper, we use a 34-layer 3D-ResNet consisting of 4 residual blocks whose 3D convolutional layer numbers are 3, 4, 6, 3, respectively. To make a prediction, the neuron number of the last fully-connected layer is set to two. The detailed structure of the proposed 34-layer 3D-ResNet module is shown in Fig 2. In Fig 2, 3D-Conv means a 3D convolutional
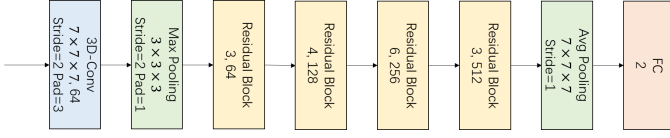
Fig. 2. The 34-layer 3D-ResNet.

Fig. 3. The overall structure of our system.

layer, whose parameters include kernel size, channels, stride and padding. The pooling layers following the convolutional leyers are set with the kernel size, stride and padding. The parameters in Residual Block are the convolutional layers in the bolck and the channel size, respectively. FC represents a fully-connected layer that has two neurons in the network to predict the malignancy of nodules.

C. Classifier with Multiple Kernel Learning

The 3D-ResNet estimates the malignancy of lung nodules barely by CT images, paying no attention to the influence of other relevant heterogeneous information. For instance, in Mayo Clinic model [3], the probability of nodule malignancy p is comuputed as follows:

$$p = \frac{e^x}{1 + e^x}, \qquad (4)$$

where

$$\begin{aligned} x = &-6.8272 + (0.0391 \times age) + (0.7917 \times smoke) \\ &+ (1.338 \times cancer) + (0.1274 \times diameter) \\ &+ (1.0407 \times spiculation) + (0.7838 \times location). \end{aligned} \qquad (5)$$

In (5), age, $smoke$ and $cancer$ are three attributes related to a patient to identify the age of the patient, whether he/she smokes and whether he/she suffered from extrathoracic cancer in the last five years. The other three variables are associated with morphologic features of nodules. It can be concluded that apart from the features that can be extracted from chest radiological data, clinical data can also affect the classification nodules.

In this paper, we consider the features of chest radiological data , i.e., CT images, and the clinical data including age, the history of extrathoracic cancer, smoking history, hypertension, heart disease, diabetes, tuberculosism, hepatitis, drinking history. For each of the clinical data except age, we use 0 to represent negative and 1 to positive. In order to combine the heterogeneous features, the simplest way is to concatenate all of them into a long vector. With the long vector, classifiers such as SVM can be used to make a prediction. SVM is a classical classification algorithm whose goal is to find an optimal separating hyperplane setting samples from different classes apart. SVM is a kernel-based statistical classifier that can be applied to nonlinear problems and has a hign generalization ability.

However, as the features are from different Hilbert space and have distinct implications, to feed them directly to a classifier is obviously inappropriate. Fortunately, the effectiveness of MKL on improving the interpretability of decision function has been validated by some research [32], [33].
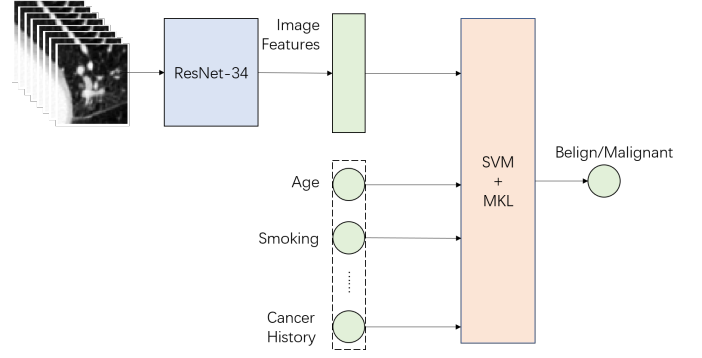
In MKL, different kernel functions are applied. The purpose of kernel learning is to learn the kernel matrix K_r by optimizing an objective function that makes the agreement between the kernel and data. [34] In kernel learning, linear kernel, polynomial kernel and radial basis function (RBF, also called Gaussian function) kernel are some of the most frequently used kernels. Based on kernel learning, MKL methods use a linear combination of various kernels other than a single one for kernel-based methods, which can be expressed as follows:

$$K(x, z) = \sum_{r=1}^{R} \beta_r K'_r(x, z), \quad \beta_r \geq 0. \qquad (6)$$

In this equation, K_r, $r = 1, 2, \ldots, R$ denotes the r-th kernel used to map the r-th type of features. β is an optimal weight.

To deal with the heterogeneous features, we adopt RBF kernel for the classification of image features and polynomial kernel for those features extracted from the clinical data. The function of RBF kernel is described as follows:

$$K(x, z) = exp(-\frac{\|x - z\|}{2\gamma^2}), \qquad (7)$$

where x, z are two input features and γ is a hyperparameter. The function of polynomial kernel is defined as:

$$K(x, z) = (x \cdot z + 1)^p, \qquad (8)$$

where p is a hyperparameter. With these two kernels, we obtain the combination of them by EasyMKL [34], which learns the optimal weight β.

D. Overall Architecture

Fig 3 shows the overall architecture of our system. The CT scans of nodules are fed to a ResNet-34 model for feature extraction. We use the features before the last fully-connected layer as image features. Along with the features extracted, clinical data encoded with 0/1 labels, including age, smkoing history and other features, are classified with a SVM that employs MKL to estimate the malignancy of lung nodules, where a polynomial kernel is used for clinical features and a RBF kernel works for image features.

To train the parameters in the ResNet-34 model, we use the strategy of transfer learning. The model is firstly trained with the LIDC-IDRI data set that has more data than the private

TABLE II
DISTRIBUTIONS OF THE CLINICAL DATA.

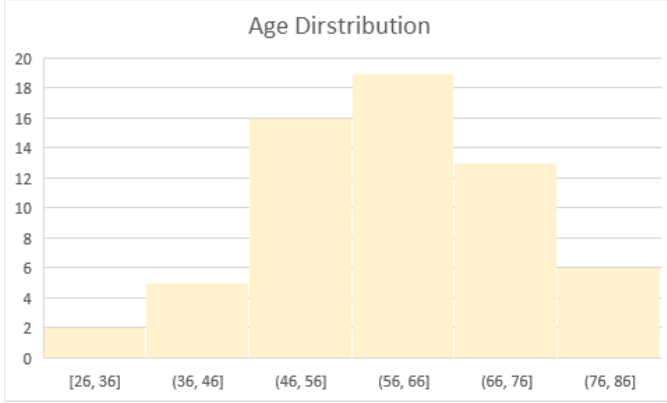| Attributes	Number(out of 61)
Smoking history	14
Extrathoracic cancer history	2
Hypertension	15
Heart Disease	5
Diabetes	1
Tuberculosis	2
Hepatitis	1
Alcohol consumption habits	11

Fig. 4. Distribution of age.

data set but no clinical information. After obtaining the well-trained parameters, it is finetuned with a private data set which has fewer CT images but clinical data of patients.

III. EXPERIMENTS AND RESULTS

TO evaluate the performance of the proposed model and to validate the effectiveness of the 3D-ResNet architecture and the heterogeneous features fusion, a series experiments are described in this section.

A. Data Processing

Our experiments involve two data sets. The first one is LIDC-IDRI data set which contains 1,018 lung cancer CT cases annotated by 4 experienced radiologists. The other one is a private data set obtained from a hospital. It contains 14415 CT images of 61 patients as well as their clinical information including age, the history of extrathoracic cancer, smoking history, hypertension, heart disease, diabetes, cerebrovascular disease, tuberculosis, hepatitis and alcohol consumption habits. The feature distributions are exhibited in Table II and Fig. 4.

In LIDC-IDRI, the malignancy levels are classified into 5 levels reflecting the degree of malignancy with 1 being benign and 5 being malignant. In the experiments, we consider level 1 and 2 as benign nodules while 4 and 5 as malignant nodules. Level 3 is ignored as the malignancy of nodules of this level cannot be determined. As each of the nodules is annotated by multiple radiologists, we take the mean malignancy level of a nodule annotated by radiologists as the real label. In the private dataset, lung nodules are directly labeled as 0 (benign nodules) and 1 (malignant nodules).

TABLE III
ARCHITECTURE OF THE DNN MODEL.

Module	Input	Output
Input	$28 \times 28 \times 1$	784
FC	784	512
FC	512	256
FC	256	64
FC	64	2
Softmax	2	2

TABLE IV
ARCHITECTURE OF THE CNN MODEL.

Module	Kernel	Input	Output
Convolution	5×5	$28 \times 28 \times 1$	$24 \times 24 \times 32$
Max Pooling	2×2	$24 \times 24 \times 32$	$12 \times 12 \times 32$
Convolution	5×5	$12 \times 12 \times 32$	$8 \times 8 \times 32$
Max Pooling	2×2	$8 \times 8 \times 32$	$4 \times 4 \times 32$
FC	\	$4 \times 4 \times 32$	512
FC	\	512	2
Softmax	\	2	2

To generate the 3D lung nodule crops, we resampled the CT scans and cropped the nodules into volumes with a shape of $40 \times 40 \times 20$ according to the annotated locations of each nodule. In LIDC-IDRI data set, we obtain 1181 benign nodules and 420 malignant nodules after cropping. To balance the amont between them, we rotate all malignant nodules randomly and finally get 1181 benign nodules and 1260 malignant nodules. In the private data set, we obtain 228 malignant and 200 benign nodules.

For convenience of the comparative experiments, we also generated the 2D crops of nodules in LIDC-IDRI according to the annocations and obtained 3879 benign nodule crops and 3103 malignant nodule crops. All 2D nodule crops are resized to 28×28 according to the work of Song *et al.* [35]. The 2D crops of benign and malignant nodules in the private data set are 200 and 228, respectively, each of which is the center slice of the corresponding 3D nodule volume.

For each of the clinical data except age, we use 0 to represent negative and 1 to positive. In addition, the features of age are normalized to [0, 1].

B. Evaluation Metrics

For measurement criterions, accuracy, sensitivity and specificity are employed. Accuracy measures the correctness of one model, sensitivity measures the correctness of the positive

TABLE V
ARCHITECTURE OF THE SAE MODEL.

Module	Input	Output
Input	$28 \times 28 \times 1$	784
FC	784	256
FC	256	64
FC	64	2
Softmax	2	2

TABLE VI
CLASSIFICATION RESULTS FOR CT SCANS

Methods	LIDC-IDRI			Private Data Set		
	Accuracy(%)	Sensitivity(%)	Specification(%)	Accuracy(%)	Sensitivity(%)	Specification(%)
DNN	79.01	80.19	78.03	62.35	76.47	41.18
SAE	76.36	76.41	76.31	58.82	68.63	44.12
CNN	77.08	71.54	81.71	67.06	70.59	61.76
3D-CNN	**91.29**	**91.01**	91.40	**84.70**	**86.05**	83.33
ResNet-34	89.68	75.28	**95.47**	83.52	81.39	**85.71**

samples while specificity measures the correctness of the negative samples. Thir defination are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (9)$$

$$Sensitivity = \frac{TP}{TP + FN}, \quad (10)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (11)$$

where TP is the number of malignant nodules being correctly classified; TN is the number of benign nodules being correctly classified; FP is is the number of benign nodules being wrongly classified and FN is is the number of malignant nodules being wrongly classified.

C. Results

We have implemented four models for comparison. Among them, three models are from the work of Song *et al. et al.* using three kinds of deep learning methods, CNN, SAE and deep neural network (DNN) respectively for 2D nodule classification. Their model configuration are shown in IV, V, III, respectively. The other model is a conventional chain-like 3D-CNN, whose configuration is shown in I. The experiments are organized as follows.

(1) Image feature extractor. In order to show the effectiveness of the proposed image features extractor, the ResNet-34 model, together with the implemented models is validated on the LIDC-IDRI data set and the private data set without using the clinical data of patients.

All models are trained on the LIDC-IDRI firstly and finetuned with the private data set. The training data and testing data are splited into 4:1. The learning rate is initialized to 0.0001 and an optimizer of Adam is employed to adaptive adjust the learning rate. The training is finished with 300 epoches under *PyTorch* framework. The loss function used in these models is cross entropy.

TABLE VI presents the performance of classifiers using different image feature extractors for nodule classification with CT images. From this table, we can see that the 3D-CNN model is more competitive than the other deep learning methods with a higher accuracy of 91.29%, sensitivity of 91.01% and specification of 91.40% on LIDC-IDIR data set and a higher accuracy of 84.70%, sensitivity of 83.33% and specification of 86.65% on the private data set for nodule classification.

TABLE VII
CLASSIFICATION RESULTS FOR HETEROGENEOUS FEATURES

Methods	Accuracy(%)	Sensitivity(%)	Specification(%)
3D-CNN	84.70	86.05	83.33
3D-CNN+MKL	89.72	86.96	91.80
ResNet-34	83.52	81.39	85.71
ResNet-34+MKL	**90.65**	**87.50**	**94.12**

Surprisingly, the deeper network 3D ResNet-34 does not achieve a better performance due to a lower sensitivity. We suppose it is the classifier layers, a global average pooling and a fully-connected layer that is not adequately non-linear compared with the 3D-CNN model. This inspiration comes from the results of the next experiment introduced below.

In addition, both of the two 3D networks outperform the 2D models to a large extent, which validates that the 3D structures are capable of utilizing the 3D nature of CT scans and lead to better performances.

(2) Classifier with MKL. We test the influence of the heterogeneous feature learning by combining the image features extracted by ResNet-34 and the chain-like 3D-CNN with the clinical data of patients on the private dataset. We save the output of the last layer before the fully-connected layer as the input image features of the classifier. The hyper-parameters in SVM and MKL are obtained through with grid search.

The classification results of the above methods are shown in Table VII. The accuracy of ResNet-34 model with MKL has reached 89.72% while the accuracy of ResNet-34 model without heterogeneous features is 83.52%.

Note that while the ResNet-34 achieves an unsatisfactory result in the classification using only CT scans compared with the 3D-CNN model, in this experiment it achieves better performance. As the mothods of 3DCNN+MKL and ResNet-34+MKL share the same classifier and clinical data in this experiment, it can be inferred that the features extracted by ResNet-34 is more representative. It provides us with the inspiration that the reason why ResNet-34 cannot outperform the 3D-CNN model in the last experiment is probably because of the different classifiers of the two models.

By comparing the results between those models using heterogeneous features learning and the methods barely using the image features, the significance of heterogeneous features is proved.

IV. Conclusion

In this paper, an efficient deep CADx system based on 3D-ResNet and MKL for pulmonary nodule classification is proposed.To fully utilize the 3D nature of the CT scans and to extract deep features, we use a 34-layer 3D-ResNet for feature extraction. Not only the image features, but also the clinical features of the patients are exploed in our work. To effectively learn the heterogeneous featuers, we employ a SVM suing MKL technique for nodule classification. The experimental results can prove the improvements from the fusion of heterogeneous features. As the clinical data in this work are still insufficient and there exists some imbalance problems, it lacks of sense to directly defer the importance of each feature using data mining algorithms. However, mining out the potential relationships between these features and lung cancer may bring about new inspiration. In addition, the pulmonary nodules are cropped directly from the labels and the clinical features used in this work are artiffitially extracted from the medical records which cost labours. To improve the level of automation and to promote the efficiency of radiologists, it is worthy of realizing an overall nodule diagnosis system including the modules of nodule detection, clinical feature extraction and nodule classification in the future.

Acknowledgment

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. Tavares, "Automatic 3d pulmonary nodule detection in ct images: a survey," *Computer methods and programs in biomedicine*, vol. 124, pp. 91–107, 2016.

[3] S. J. Swensen, M. D. Silverstein, D. M. Ilstrup, C. D. Schleck, and E. S. Edell, "The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules," *Archives of internal medicine*, vol. 157, no. 8, pp. 849–855, 1997.

[4] M. K. Gould, L. Ananth, and P. G. Barnett, "A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules," *Chest*, vol. 131, no. 2, pp. 383–388, 2007.

[5] A. McWilliams, M. C. Tammemagi, J. R. Mayo, H. Roberts, G. Liu, K. Soghrati, K. Yasufuku, S. Martel, F. Laberge, M. Gingras *et al.*, "Probability of cancer in pulmonary nodules detected on first screening ct," *New England Journal of Medicine*, vol. 369, no. 10, pp. 910–919, 2013.

[6] G. J. Herder, H. van Tinteren, R. P. Golding, P. J. Kostense, E. F. Comans, E. F. Smit, and O. S. Hoekstra, "Clinical prediction model to characterize pulmonary nodules: validation and added value of 18 f-fluorodeoxyglucose positron emission tomography," *Chest*, vol. 128, no. 4, pp. 2490–2496, 2005.

[7] A. Al-Ameri, P. Malhotra, H. Thygesen, P. K. Plant, S. Vaidyanathan, S. Karthik, A. Scarsbrook, and M. E. Callister, "Risk of malignancy in pulmonary nodules: a validation study of four prediction models," *Lung Cancer*, vol. 89, no. 1, pp. 27–30, 2015.

[8] W. Kostis, A. Reeves, D. Yankelevitz, and C. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical ct images," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1259–1274, 2003.

[9] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, and C. Kuroda, "Fractal analysis of small peripheral pulmonary nodules in thin-section ct: Evaluation of the lung-nodule interfaces," *J Comput Assist Tomogr*, vol. 26, no. 4, pp. 573–578.

[10] T. W. Way, B. Sahiner, H. Chan, L. Hadjiiski, P. N. Cascade, A. Chughtai, N. Bogot, and E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on ct scans: Improvement of classification performance with nodule surface features," *Medical Physics*, vol. 36, no. 7, 2009.

[11] H.-P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space," *Physics in Medicine & Biology*, vol. 40, no. 5, p. 857, 1995.

[12] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and therapy*, vol. 8, 2015.

[13] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in ct images," in *2015 12th Conference on Computer and Robot Vision*. IEEE, 2015, pp. 133–138.

[14] M. J. Shafiee, A. G. Chung, D. Kumar, F. Khalvati, M. Haider, and A. Wong, "Discovery radiomics via stochasticnet sequencers for cancer detection," *arXiv preprint arXiv:1511.03361*, 2015.

[15] W. Sun, B. Zheng, and W. Qian, "Automatic feature learning using multichannel roi based on deep structured algorithms for computerized lung cancer diagnosis," *Computers in biology and medicine*, vol. 89, pp. 530–539, 2017.

[16] M. B. Rodrigues, R. V. M. Da Nóbrega, S. S. A. Alves, P. P. Rebouças Filho, J. B. F. Duarte, A. K. Sangaiah, and V. H. C. De Albuquerque, "Health of things algorithms for malignancy level classification of lung nodules," *IEEE Access*, vol. 6, pp. 18 592–18 601, 2018.

[17] 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.

[18] 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.

[19] J. Lyu and S. H. Ling, "Using multi-level convolutional neural network for classification of lung nodules on ct images," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 686–689.

[20] B.-C. Kim, Y. S. Sung, and H.-I. Suk, "Deep feature learning for pulmonary nodule classification in a lung ct," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2016, pp. 1–3.

[21] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct," *Information Fusion*, vol. 42, pp. 102–110, 2018.

[22] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2016.

[23] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 559–567.

[24] 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.

[25] S. Liu, Y. Xie, A. Jirapatnakul, and A. P. Reeves, "Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 4, p. 041308, 2017.

[26] K. Liu and G. Kang, "Multiview convolutional neural networks for lung nodule classification," *Plos One*, vol. 27, no. 1, pp. 12–22.

[27] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[28] 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.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *computer vision and pattern recognition*, pp. 770–778, 2016.

[31] K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3d residual networks for action recognition," *arXiv: Computer Vision and Pattern Recognition*, 2017.

[32] F. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," p. 6, 2004.

[33] C. Tong, B. Liang, J. Li, and Z. Zheng, "A deep automated skeletal bone age assessment model with heterogeneous features learning," *Journal of Medical Systems*, vol. 42, no. 12, p. 249, 2018.

[34] F. Aiolli and M. Donini, "Easymkl: A scalable multiple kernel learning algorithm," *Neurocomputing*, vol. 169, pp. 215–224, 2015.

[35] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using deep learning for classification of lung nodules on computed tomography images." *Journal of Healthcare Engineering*, vol. 2017, p. 8314740, 2017.