Deep Regression via Multi-Channel Multi-Modal Learning for Pneumonia Screening

QIULI WANG¹, DAN YANG¹, ZHIHUAN LI¹, XIAOHONG ZHANG¹, AND CHEN LIU²

¹School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China
²Radiology Department, The First Affiliated Hospital of Army Medical University, Chongqing 400032, China

Corresponding authors: Xiaohong Zhang (xhongz@cqu.edu.cn) and Chen Liu (liuchen@aifmri.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772093, in part by the National Key Research and Development Project of China under Grant 2018YFB2101200, in part by the Chongqing Major Theme Projects under Grant cstc2018jszx-cyztzxX0017, and in part by the Fundamental Research Funds for the Central Universities under Grant 2019CDYGYB014.

ABSTRACT

Pneumonia screening is one of the most crucial steps in the pneumonia diagnosing system, which can improve the work efficiency of the radiologists and prevent delayed treatments. In this paper, we propose a deep regression framework for automatic pneumonia screening, which jointly learns the multi-channel images and multi-modal information (i.e., clinical chief complaints, age, and gender) to simulate the clinical pneumonia screening process. We demonstrate the advantages of the framework in three ways. First, visual features from multi-channel images (Lung Window Images, High Attenuation Images, Low Attenuation Images) can provide more visual features than single image channel, and improve the ability of screening pneumonia with severe diseases. Second, the proposed framework treats chest CT scans as short video frames and analyzes them by using Recurrent Convolutional Neural Network, which can automatically extract multiple image features from multi-channel image slices. Third, chief complaints and demographic information can provide valuable prior knowledge enhancing the features from images and further promote performance. The proposed framework has been extensively validated in 900 clinical cases. Compared to the baseline, the proposed framework improves the accuracy by 2.3% and significantly improves the sensitivity by 3.1%. To the best of our knowledge, we are the first to screen pneumonia using multi-channel images, multi-modal demographic and clinical information based on the large scale clinical raw dataset.

INDEX TERMS

Computed tomography, clinical diagnosis, biomedical imaging, pneumonia screening.

I. INTRODUCTION

Pneumonia is a prevalent thoracic disease that affects many people. In clinical practice, radiologists need to consider multi-modal information to screen pneumonia cases from massive clinical data. Conventionally, this task relies on experts’ manual operations, which is time-consuming and inhibits fully automatic assessment. Thus, developing a fast, robust, and accurate Computer-Aided Diagnosis (CAD) system to perform automated screening of pneumonia is meaningful and vital.

Roughly speaking, works related to pneumonia screening, detection, monitoring, and diagnosing [1]–[6] can be classified into three categories, including chest X-Ray based methods, chest CT based methods, and multi-modal based methods. Our study is related to the second and third categories.

The first category builds frameworks by analyzing chest X-Rays. In 2017, Wang et al. [5] provided hospital-scale chest X-Ray database ChestX-ray8, which contained eight common thoracic diseases. Later in 2017, Yao et al. [7] achieved AUROC of 0.713 in classifying pneumonia using DenseNet Image Encoder. Rajpurkar et al. [8] developed CheXnet with 121 convolutional layers and achieved AUROC 0.7680 in pneumonia classification.

The second category focuses on analyzing chest CT. For example, Shin et al. [4] exploited three important, but previously understudied factors of employing deep convolutional neural networks to computer-aided detection problems. Gao et al. [9] presented a method to classify interstitial lung disease imaging patterns on CT images and introduced normal lung rescale, high attenuation rescale, and low attenuation rescale. In this paper, we use CT as the image inputs. Rather than manual image selection, we apply a Recurrence Convolutional Neural Network to perform automatic extraction of multiple image features. Moreover, based on
researches [4], [9], we further demonstrate the effect of each image window during the deep learning process.

The third category, which is based on multi-modal analysis, has attracted considerable attention in the literature [10], [12]–[16], and our method is closely related to this category. In 2016, Shin et al. [11] proposed a method used CNN to extract features from chest X-Rays and used LSTM to generated MeSH (Medical Subject Headings) terms for chest X-Rays. In 2018, Wang et al. [3] proposed TieNet, which could learn visual image features and semantic features of textual reports at the same time. This study was innovative and proved that textual information is closely related to the images. Instead of using reports written by the radiologists, we use textual clinical complaints to enhance the features extracted from images and exploit expressions in chief complaints that are related to pneumonia. Meanwhile, we also use prior knowledge from the demographic information, which is seldom incorporated in the conventional models. In the field of Alzheimer’s disease diagnosis, Liu et al. [18] proposed a framework for joint classification and regression and proved that age and gender are related to brain states. Invoked by this study, we assume that age and gender are also related to the states of lungs, so that they can provide useful prior knowledge and improve the performances of the learning methods.

It is well accepted that radiologists need to consider several information sources in clinical practice, which is shown in Figure 1. First of all, image information is necessary because most pneumonia cases have direct visual evidence. Second, the radiologists also need to check the patient’s age, and gender, which can provide valuable prior knowledge. Last but not least, clinical chief complaints are very important because chief complaints can reflect patients’ direct feelings and conditions and provide prior knowledge too. After considering all the information above, the radiologists will then decide further treatments. In this paper, we propose a deep regression framework for pneumonia screening, which takes multi-channel images, demographic information (i.e., age and gender), and clinical information (chief complaints) into the deep learning process. Our framework aims to simulate clinical practice and improve screening accuracy using multi-channel images and multi-modal information. As far as we know, we are the first to screen pneumonia var multi-channel multi-modal learning using a large scale clinical raw dataset.

Figure 2 illustrates the schematic diagram of the proposed framework. Herein, (a) each CT image will be transformed into a multi-channel image with three windows: Lung Window (LW), High Attenuation (HA), and Low Attenuation (LA). LW provides visual features of overall lung tissues. HA provides visual features of abnormal increase in lung density. LA provides visual features of abnormal decrease in lung density. Three channels complement each other, which not only maintains the ability to extract information from overall lung tissues but also increases the ability to capture information from abnormal lung tissues; (b) We include clinical chief complaints in our framework. Chief complaints can provide information on the location of pain, and symptoms, which is related to the CT image and enhances the visual features extracted from CT; (c) Demographic information about age and gender can provide prior knowledge, which is related to the lungs states; (d) All information mention above will be fed into a regression framework to simulate the clinical screening process for pneumonia.

In summary, the main contributions of this work are as follows.

i) We formulate the clinical pneumonia screening process as a regression problem. The proposed framework takes advantage of the multi-channel images, multi-modal information, and overcomes the shortcomings of a single source of information.

ii) We propose to treat chest CT scans as short video frames and analyze them using Recurrent Convolutional Neural Network, which can make the most of CT slice information and automatically extract multiple visual features from multi-channel images.

iii) We demonstrate that different CT windows can provide different density information during the convolutional process and provide supplementary information for each other, which can improve the classification accuracy.

iv) We verify that certain expressions in the chief complaints are statistically related to the diagnosing of pneumonia. Chief complaints, along with prior demographic information, can improve the converging speed during training the framework and the overall performances of the deep learning framework.

The rest of this paper is organized as follows: Section II describes our research questions and studied information. Section III describes the details of our dataset. Then, the architecture of the proposed framework is presented in
II. STUDY SETUP

A. RESEARCH QUESTIONS

We formalize our study with the following three research questions:

RQ1: How the image windows affect the extraction of visual features during the convolutional process? In this RQ, we demonstrate the effects of image windows on CNN. To answer this RQ, we will conduct several experiments with different image windows and compare their experimental performances. Moreover, we output the convolutional features to display the effect of image windows intuitively.

RQ2: Can the expressions in chief complaints provide discriminative semantic features for the deep learning frameworks? In the second RQ, we would like to know the statistical relationships between chief complaints and pneumonia. To answer this RQ, we calculate the word frequency and find out the keywords play essential roles in pneumonia screening.

RQ3: How the multi-channel multi-modal information affects the performances of deep learning networks? In this RQ, we would like to discuss the effect that multi-channel multi-modal information brings to the deep learning framework. Therefore, we will build three regression networks: one framework with multi-channel images information, one framework with multi-modal demographic and clinical information, and the proposed framework with multi-channel multi-modal information. Then we demonstrate how these data affect the performances of the deep learning networks.

B. INFORMATION STUDIED

Generally speaking, demographic information, clinical information, and multi-channel CT images are all essential in the decision-making process. In our study, we decide to focus on CT images, demographic information, and clinical information.

1) CT IMAGE WINDOWS

Visual information from the CT images with different image windows aims to give direct visual features of pneumonia. With different windows, CT enhances different characteristics (like Ground Glass, Fibrosis), which are critical features during diagnose. Lung window is the most common CT window in clinical diagnosis for pneumonia. Under some circumstances, the radiologists need the mediastinal window (low attenuation) to observe the tissue details that cannot be displayed clearly in the lung window. Studies like [4], [9] have introduced normal lung rescale, high attenuation rescale, and low attenuation rescale. These studies invoke us to further exploit the effect of each image window during the learning process.

2) DEMOGRAPHIC AND CLINICAL INFORMATION

Features from demographic and clinical information aim to provide prior knowledge related to pneumonia. According to the data from the US CDC (Centers for Disease Control and Prevention), adults over 65 years or children younger than 5 years old are more likely to get pneumonia\textsuperscript{1}, since people in these age groups have weaker immune systems than middle-aged and young people. CDC also points out that people who smoke are more likely to get pneumonia than these people who do not. According to the 2018 China Adult Tobacco Survey, among people over 15 in China, 50% ...

\textsuperscript{1}https://www.cdc.gov/pneumonia/prevention.html
of males smoked, while only 2.1% of females smoked. Since the number of male smokers is significantly higher than that of females, this phenomenon may lead to gender differences in pneumonia distribution. As far as we know, conventional CAD systems for pneumonia seldom incorporate such demographic information into the learning process.

Clinical information, especially patients’ chief complaints are valuable in clinical practice. Understanding the chief complaints is the key to pass the National Medical Licensing Examination in China [19]. In this study, we introduce textual clinical chief complaints into the learning framework and exploit how chief complaints affect the performance of the deep learning framework.

### III. EVALUATION DATASETS

There are many datasets for thoracic disease research, but datasets contain all CT images, demographic and clinical information are rare. To evaluate our framework, we build a private dataset with the raw data collected from the Radiology Department of The First Affiliated Hospital of Army Medical University. All data have been reviewed and provided corresponding ethics approval. All sensitive information has been removed. In this study, we collect over a thousand community-acquired pneumonia cases and healthy cases from hospital PACS (Picture Archiving and Communication Systems) in the last three years (2016 - 2019) randomly.

Raw data from the hospital are stored in DICOM files, which might have more than one series of images, and each series has specific data types, image windows, or view angles. Generally speaking, radiologists use the series under the lung window with the smallest ‘Slice Thickness’. So we design a protocol to pick up the specific series for us:

(a) We choose the series with the specific ‘Convolution Kernel’. Different ‘Convolution Kernel’ names indicate different image windows. The names of ‘Convolution Kernel’ vary between hospitals and CT equipment. In our study, we choose ‘B31f’, ‘I31f 3’, ‘B70f’, ‘B80f’, ‘B70s’. We notice that in the Radiology Department of The First Affiliated Hospital of Army Medical University, ‘B70s’, which can be referred to lung window, is the most common parameter used in clinical practice.

(b) We remove the series like ‘Patient Protocol’, ‘Topogram’. These series contain some basic parameters and information about CT equipment, which are not suitable for deep learning.

(c) We calculate ‘Slice Thickness’ of each series, and keep the series with the smallest ‘Slice Thickness’, since small thickness may keep more detailed information about lung structures.

(d) If there is more than one series meeting the last three requirements, we keep the series with the largest number of slices, which has a larger span of view.

As a result, we pick up 552 pneumonia cases and 450 healthy cases (1002 cases total), and each CT scan has a case file, which contains necessary patient information: patient ID, gender, age, and chief complaints. Since our data are collected from the Radiology Department, the proportion of pneumonia is higher than usual, we randomly choose 450 cases from 552 healthy cases and balance two classes. We show in Table.1 the details of this dataset.

### IV. DEEP REGRESSION FRAMEWORK FOR PNEUMONIA SCREENING

#### A. OVERVIEW

In this section, we will discuss the details of the proposed framework. The architecture of the proposed framework is shown in Figure.3, which can be described as a multi-channel multi-modal regression framework for classification. As shown in this figure, there are three kinds of inputs: (1) multi-channel images, (2) semantic features of clinical chief complaints, (3) demographic information. The regression progress can be formulated as:

$$P(X) = \text{Softmax}(F(V(X) \otimes C(X) \otimes A(X) \otimes G(X)))$$

$X$ is the input case, $\otimes$ is the concatenation operation, $V(X)$ is visual features captured from multi-channel images, $C(X)$ is semantic information captured from clinical chief complaints, $A(X)$ and $G(X)$ indicate patient age and gender. $F$ is a function to fit the regression framework of $V(X)$, $C(X)$, $A(X)$ and $G(X)$. $P$ characterizes the likelihood of being pneumonia.

In our framework, an RCNN is used to learn visual features from multi-channel CT images and generate $V(X)$, and an LSTM network is used to learn the semantics of clinical chief complaints.

#### B. RECURRENT CONVOLUTIONAL NEURAL NETWORK

RCNN (Recurrent Convolutional Neural Network) has been proved to be very effective in video caption, description, and classification [22]–[25], some studies have applied RCNN to medical image analysis. Zreik et al. [26] recently used RCNN for automatic detection and classification of coronary artery plaque. They used CNN extracts features out of $25 \times 25 \times 25$ voxels cubed and used an RNN to process the entire sequence using gated recurrent units (GRUs) [27]. Tseng et al. [28]...
exploited convolutional LSTM to model a sequence of 2D slices, and jointly learned the multi-modalities and convolutional LSTM in an end-to-end manner to segment 3D biomedical images.

As mentioned before, CT allows visualization of lung structures, which brings a large amount of redundant information, like muscle, vessels, and bones. It will cost lots of calculation resources if we use 3D CNN directly. Conventional methods often require manual image selection so that they can move out images without disease tissues. In clinical practice, the radiologists slide all CT images as watching short video frames. It is a 2D operation but can get 3D visual features. It invokes us to analyze CT as short video frames and use Recurrent Convolutional Neural Network. In RCNN, each slice will be fed into CNN in sequence and get a sequence of visual features. Then this sequence will be fed into RNN so that we can perform automatical extraction of multiple visual features from 3D structures. We treat this process as an encoder $V$ for CT multi-channel slice visual features, which is calculated as:

$$V_t = LSTM(\varepsilon_{x_t}, V_{t-1}, z_{t-1})$$  \hspace{1cm} (1)

$\varepsilon_{x_t}$ is the visual features of the $t$-th CT slice, which is captured by CNN. $V_{t-1}$ is the output of $t - 1$ step. $z_{t-1}$ is the hidden state of $t - 1$ step. $t$ is the slice number. In this study, each scan has 32 slices.

1) CONVOLUTIONAL NEURAL NETWORK
In this study, we use ResNet-50 as the visual extractor in RCNN. We compare three kinds of classic CNN frameworks: VGG-16 [29], ResNet-50 [30] and GoogLeNet with Inception-V3 [31] and our experiments demonstrated that ResNet-50 performed the best. The experiments of this part will be discussed later in Section V. We use ResNet-50 without fully-connected layers as a feature extractor. Due to the radiologists’ clinical requirements, the input size of CNN is $512 \times 512$.

2) GLOBAL AVERAGE POOLING
Since the input size of CNN is $512 \times 512$, the output size will be extensive. Using fully-connected layers directly will lead to a heavy calculation burden. To reduce this burden, we use the global average pooling instead. The global average pooling [32] is a technology that can reduce the number of neurons significantly. It is a replacement of fully-connected layers to enable the summing of spatial information of feature maps. After global average pooling, the outputs with size $[32, 16, 16, 2048]$ will be transformed into vectors with size $[32, 1, 1, 2048]$. Then a fully-connected layer is used to reduce dimensions to 128 and fit the number of LSTM units. We will discuss the reason we use 128 LSTM units in Section VI.

3) LONG SHORT TERM MEMORY
Recurrent neural networks (RNNs) [33], [34] are an abundant class of dynamic models that have been used to generate sequences in domains as diverse as text and motion capture data. There have been several kinds of RNN units like GRU [35] and LSTM. GRU and LSTM are both widely used...
in medical image analysis [27], [28]. There is no significant difference in performance between the two models. In this study, we use LSTM as our RNN cells, because it has been demonstrated to be capable of large-scale learning of sequence data.

C. MULTI-CHANNEL IMAGES REPRESENTATION

As mentioned in Section III, each series of CT has one specific ‘Convolution Kernel’. Nevertheless, it may lead to data inconsistent between different cases. So we transform raw data into HU(Hounsfield Unit) values. The Hounsfield Unit is a quantitative scale for describing radio-density. After transformed into HU value matrices, all slices form CT scans will have the same unit of measure.

Following the study in [4], [9], HU value matrices will be transformed into images using different HU windows. Let X denote the multi-channel image input, then X can be calculated as:

\[ X = \sum_{q \in X} \text{Threshold}_q[\text{Min}, \text{Max}] \]

\[ \chi = \{\text{LW}, \text{HA}, \text{LA}\} \]

\[ \text{LW} = \text{Threshold}_\text{LW}[-1000, 400\text{HU}] \]

\[ \text{HA} = \text{Threshold}_\text{HA}[-160, 240\text{HU}] \]

\[ \text{LA} = \text{Threshold}_\text{LA}[-1400, -950\text{HU}] \]

Our thresholds are handcrafted and different from studies [4], [9]. Each slice will generate three one-channel gray-scale images, which will be concatenated into one three-channel image later. The ‘Slice Thickness’ between each slice is adjusted into 10mm, and each case will keep 32 slices so that they can cover the whole lung structures.

As shown in Figure 4, three-channel images can show more information about lung density. Original CT images are gray-scale images; high-density tissues are white; normal lung tissues and low-density tissues tend to be black. Relatively, three-channel false-color images have a larger scale of colors. First of all, high-density tissues will still tend to be white, like bones, high-density lung tissues. Second, normal lung tissues will tend to be red, and low-density tissues tend to be black, which is very useful when patients have severe lung diseases, which will lead to abnormal changes in lung density.

D. MULTI-MODAL INFORMATION REPRESENTATION

Studies like [20], [21] treated the demographic information as confounding factors. The main disadvantage of this strategy is that the original representations of subjects will be modified, because it adds up several steps of engineered pre-processing in a directed and engineered way [18]. Following the study [18], we concatenate demographic factors to the visual feature vectors learnt by RCNN. This method is simple yet effective. The demographic information of all studied patients has been listed in Table 1.

For patients’ chief complaints, since all chief complaints are written in Chinese, we have to do Chinese word segmentation. Chinese word segmentation is a challenging problem, so we will take a short cut and use a mature tool: Jieba Text Segmentation to segment Chinese sentences into Chinese word sequences.

After segmentation, we use Word2vec [36], [37] to embed word sequences into vectors and use CBOW(Continuous Bag-of-Words) to capture relationship between words. Since our corpus is small, the embedding size is set to 50, and the window size for CBOW is set to 3. We set length of Chinese word sequence to 16 since 16 is the maximum length among all chief complaint sequences. For those sequences whose length is less than 16, we add ‘None’ to fill up the voids and increase the length to 16. Then the sequences will be fed into LSTM:

\[ C_{ct} = \text{LSTM}(\text{Word}_c, C_{ct-1}, z_{ct-1}) \] (2)

\[ \text{Word}_c \] is word embedding matrix of the \( ct \)-th word in chief complaint. \( C_{ct-1} \) is the output of \( ct - 1 \) step. \( z_{ct-1} \) is the hidden state of \( ct - 1 \) step. \( ct \) is the current number of step.

E. ENSEMBLE OF DECISIONS

The multi-channel images will be encoded into a vector of length 128. The clinical chief complaints will be encoded into a vector of length 50. These two vectors, along with two demographic factors, will be fed into two fully-connected layers to fit a regression framework and calculate the likelihood \( P \). We use cross-entropy function as our loss function:

\[ \arg \min_w -\frac{1}{Q} \sum_{q=1}^{Q} \sum_{X_n \in \chi} \log(P(y^q_n = q|X_n; W)) \]

\[ \chi = \{X_n\}_{n=1}^{N} \] denotes the training set, \( X_n \) represents \( n \)-th case of training set. \( y^q_n \) is label vector. In this study, the class labels are used in a back-propagation procedure to update the network weights in the convolutional layers, LSTM units, and learn the most relevant features in the fully-connected layers. \( W \) denotes the trainable parameters of the proposed framework. \( Q \) is equal to 2, which indicates two classes in our study.

V. EXPERIMENTS

A. EXPERIMENTAL SETUP

As mentioned in Section III, the analyzed dataset contains 450 healthy cases and 450 pneumonia cases. We divide the dataset into five folds. In the RQ1, we use four folds of data, three folds for training, and one fold for testing. In the RQ3, we provide results of five-cross validation using all data for a complete review about system performance.

1https://github.com/fxsjy/jieba
TABLE 2. Comparison of all RCNNs.

| Structure    | Input Channel | Accuracy | Sensitivity | Specificity | F1 Score |
|--------------|---------------|----------|-------------|-------------|----------|
| RCNN(VGG)    | Three Channels| 0.828    | 0.889       | 0.767       | 0.857    |
| RCNN(ResNet) | Three Channels| 0.894    | 0.867       | 0.922       | 0.880    |
| RCNN(ResNet) | Lung Window   | 0.928    | 0.933       | 0.922       | 0.930    |
| RCNN(ResNet) | Low Attenuation| 0.900   | 0.889       | 0.911       | 0.894    |
| RCNN(ResNet) | High Attenuation| 0.783   | 0.711       | 0.856       | 0.745    |
| RCNN(ResNet) | Original HU Values| 0.839   | 0.900       | 0.778       | 0.868    |

The initial learning rate is $10^{-4}$ and drops 50% every 3000 training steps. The dropout rate in fully-connected layers is 0.5. Each experiment contains three training epochs. All the experiments are run on a GPU of NVIDIA Tesla V100. Each experiment needs about 12-16G of GPU memory. The network implementation is based on the Tensorflow framework [38].

Moreover, we use CNN frameworks pre-trained on ImageNet [39]. Experiments demonstrate that using pre-trained frameworks can significantly improve the converging speed. Even though the data in ImageNet has a different size with CT data, we believe it is still helpful using pre-trained frameworks. Pre-training aims to initialize the convolutional kernels, so that they can get preliminary ability to extract features. The size of images will not affect the preliminary ability to extract visual features since the convolutional process is a sliding process over the images.

The source code for this framework can be found at https://github.com/wangqiuli0102/MCMM_Pneumonia. We will also release the trained parameters and some sample cases. We cannot release dataset because of the privacy of patients.

B. BASELINE SYSTEM

To manifest the advantages offered by the proposed framework, we construct a baseline framework for comparison. Generally speaking, we should have a 3D CNN as our baseline because it is the most common method to analyze the 3D data. However, the input size is so large that it is impossible to train a 3D CNN using raw data under our conditions. Meanwhile, the lung window is the most common choice in clinical practice. As a result, we choose the RCNN, which contains ResNet-50 and trained with lung window images, as our baseline.

C. RQ1: HOW THE IMAGE WINDOWS AFFECT THE EXTRACTION OF VISUAL FEATURES DURING THE CONVOLUTIONAL PROCESS?

Approach: To answer this research question, we conduct several experiments with different image windows and compare their experimental performances. We output the middle convolutional features of CNNs to display the effect of image windows during the training. Additionally, we run an experiment using original CT values ([-2000, 2000HU]) as a comparison of other image windows. By default, the experiments in this section use RCNN, so that the influences of clinical and demographic information are removed. We use three folds of data for training, and one fold data for testing.

Results: Table.2 provides a comprehensive performance comparison between different combinations of RCNN frameworks and input channels. We analyze this table in two aspects: (1) The effect of image windows. (2) The impact of CNN frameworks.

1) THE EFFECT OF IMAGE WINDOWS

To exploit the effect of image windows, we run experiments on RCNN(ResNet) with five image windows: lung window, high attenuation window, low attenuation window, original HU window, and three-channel window. As can be seen, RCNN(ResNet) with three-channel images performs the best in accuracy, sensitivity, specificity, and F1. It means three-channel images can provide the most comprehensive visual features than other window settings. RCNN(ResNet) trained with HU images is slightly better than RCNN(ResNet) trained with lung window images, and both frameworks have weaker performances than RCNN(ResNet) with three-channel images. RCNN(ResNet) trained with low attenuation images performs the worst. RCNN(ResNet) trained with HU images is the second-best choice since HU images contain all visual features that exist in other image windows. On one side, its large scale of HU values (-2000, 2000HU) makes it challenging to learn the necessary visual features. On the other side, three-channel images help the deep learning frameworks focus on different parts of the lungs with different image windows. So it is not difficult to understand that original HU is worse than three-channel images but better than the others.

2) THE IMPACT OF CNN FRAMEWORKS

We test three classic CNN frameworks: VGG-16, GoogLeNet, and ResNet-50 with three-channel images. As can be seen, ResNet-50 has better performances than VGG-16 and GoogLeNet. RCNN with ResNet-50 achieves 0.930 in F1 score, 5% higher than GoogLeNet with InceptionV3, and 7.3% higher than VGG-16. This result is similar to the conclusion drawn in [5], their experiments also show that ResNet-50 outperforms GoogLeNet and VGG-16.

In order to exploit the effect of three-channel images further, we output the convolutional feature maps of five pneumonia cases in Figure.4. More specifically, we output the feature maps after one convolutional layer, one max-pooling layer, and three ResNet blocks, the size of feature maps...
FIGURE 4. Convolutional feature maps from CNN frameworks trained by different images. We show five pneumonia examples in this figure. The leftmost column is three-channel input. The four columns on the right are convolutional feature maps from CNN frameworks trained with lung window images, high attenuation images, low attenuation images. Areas in white blocks are abnormal low-density lung tissues; areas in yellow blocks are abnormal high-density lung tissues.

As can be seen: When fed in images with the HA window, CNN can capture high-density features but has difficulty in capturing the differences between low-density tissues and normal tissues. Contrarily, when fed in images with the LA window, CNN can keep low-density features well, but has difficulty in capturing high-density features. Meanwhile, with three-channel images, CNN has better discrimination for normal tissues, high-density tissues, and low-density tissues.

In general, multi-channel images can take advantage over single-channel images: the low attenuation channel and high attenuation channel can provide density features that are difficult to be captured in the single-channel images and improve the ability to extract lung visual features.

D. RQ2: CAN THE EXPRESSIONS IN CHIEF COMPLAINTS PROVIDE DISCRIMINATIVE SEMANTIC FEATURES FOR THE DEEP LEARNING FRAMEWORKS?

Approach: To answer this research question, we count the frequencies of Chinese words (we only keep the words about symptoms, feelings) in chief complaint and compare the top-10 keywords of pneumonia cases and healthy cases.

Results: Table.3 and Table.4 show the top 10 keywords in PC (pneumonia cases) and HC (healthy cases). As can be seen, chief complaints from pneumonia cases contain more keywords about intuitive symptoms, like ‘Productive Coughing’, ‘Coughing Blood’, and ‘Trouble Breathing’. Meanwhile, chief complaints from healthy cases contain more keywords about feelings, like ‘Pain’, ‘Backache’, or ‘Uncomfortable’. We also find ‘Cough’ is the most frequent keyword in both HC and PC. It appears 165 times (36.7%) in PC, 145 times (32.2%) in HC, which means it is a symptom with minimal discriminative information. Moreover, we count the number of words that can provide
TABLE 3. Top 10 frequent key words in pneumonia cases.

| Key Words          | Frequency in PC | Percentage | Frequency in HC | Percentage |
|--------------------|-----------------|------------|-----------------|------------|
| Coughing           | 163             | 0.367      | 145             | 0.322      |
| Productive Coughing| 88              | 0.196      | 47              | 0.104      |
| Recurrence Condition| 44            | 0.098      | 36              | 0.080      |
| Trouble Breathing  | 41              | 0.091      | 13              | 0.029      |
| Coughing Blood     | 34              | 0.075      | 1               | 0.002      |
| Fever              | 34              | 0.075      | 12              | 0.027      |
| Aggravation        | 33              | 0.073      | 11              | 0.024      |
| Pain               | 29              | 0.064      | 110             | 0.244      |
| Uncomfortable      | 14              | 0.031      | 46              | 0.102      |
| Infection          | 14              | 0.031      | 1               | 0.002      |

Percentage is frequency divided by number of cases. PC is pneumonia Cases, HC is Healthy Cases.

The frequency of bold keywords is much higher than that in HC.

TABLE 4. Top 10 frequent key words in healthy cases.

| Key Words          | Frequency in HC | Percentage | Frequency in PC | Percentage |
|--------------------|-----------------|------------|-----------------|------------|
| Coughing           | 145             | 0.322      | 163             | 0.367      |
| Pain               | 110             | 0.244      | 29              | 0.064      |
| Productive Coughing| 47              | 0.104      | 88              | 0.196      |
| Uncomfortable      | 46              | 0.102      | 14              | 0.031      |
| Recurrent Condition| 36              | 0.080      | 44              | 0.098      |
| Backache           | 22              | 0.049      | 6               | 0.013      |
| Chest Tightness    | 15              | 0.033      | 10              | 0.022      |
| Trouble Breathing  | 13              | 0.029      | 41              | 0.091      |
| Fever              | 12              | 0.027      | 34              | 0.075      |
| Aggravation        | 11              | 0.024      | 33              | 0.073      |

The frequency of bold keywords is much higher than that in PC.

information about the location. In 900 cases, every 2 cases have one word that related to the location (about 450 times). This phenomenon demonstrates that information from chief complaints can provide useful prior knowledge for the learning framework.

We output the activation of LSTM neurons and observe some interesting phenomenons. Among those neurons in LSTM, some of them are almost the same during training, and impossible to tell which one is more active. However, some neurons are more active when fed in pneumonia cases and influence the performances of our framework. It means certain expressions can activate certain neurons during the learning process.

In summary, words in chief complaints are statistically related to pneumonia. Chief complaints from pneumonia patients contain more keywords about intuitive symptoms; Contrarily, chief complaints from healthy cases contain more keywords about feelings.

E. RQ3: HOW THE MULTI-CHANNEL MULTI-MODAL INFORMATION AFFECTS THE PERFORMANCES OF DEEP LEARNING NETWORKS?

Approach: To answer this research question, we build three frameworks: the proposed framework with multi-channel multi-modal information, the framework with multi-channel images, the framework with multi-modal demographic and clinical information, to evaluate the impacts of multi-channel images and multi-modal information separately. We use RCNN trained with lung window images as the baseline. It should be noted that ‘RCNN’ in this section refers to RCNN(ResNet). In this section, we use five-cross validation to evaluate the proposed framework.

Results: We show experimental results of the five-cross validation in Table.5. As can be seen, the average accuracy, sensitivity, specificity, and F1 Score of the baseline are 0.923, 0.911, 0.936, 0.917. And if we add multi-channel images into this framework, the average accuracy, sensitivity, specificity, and F1 Score increase to 0.936, 0.931, 0.940, and 0.933. It means that multi-channel images do provide more complement visual features than single-channel images and improve all four indicators. Moreover, when we add multi-modal clinical and demographic information, the average accuracy, sensitivity, specificity, and F1 Score will increase to 0.936, 0.931, 0.940, and 0.933. It should be noted that sensitivity was significantly improved by 3.1%, and the performances of this framework are very balanced. We also train a regression framework using demographic and clinical information only, and it achieves poor performances. However, experimental results prove that demographic and clinical information can provide useful prior knowledge. Image is the decisive factor, and prior knowledge provides supplementary information.

The study of [3] achieved an AUROC of 0.947 using reports only and achieved an AUROC of 0.969 using reports and X-Rays, which was the highest among all studies based on X-Rays. We assume that it is because reports contain conclusions drawn by the radiologists, which are very helpful when training a framework. Unlike the study in [3], our framework can assist the radiologists before they making any decisions since the clinical and demographic information is provided before the decision-making process.
TABLE 5. Experimental results of five-cross validation.

| Data                      | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Ave |
|---------------------------|--------|--------|--------|--------|--------|-----|
| Lung Window Image (Baseline) | Accuracy | 0.917  | 0.922 | 0.922 | 0.944 | 0.911 | 0.923 |
|                           | Sensitivity | 0.922  | 0.889 | 0.911 | 0.933 | 0.900 | 0.911 |
|                           | Specificity | 0.911  | 0.956 | 0.933 | 0.956 | 0.922 | 0.936 |
|                           | F1 Score | 0.919  | 0.905 | 0.917 | 0.939 | 0.906 | 0.917 |
| Three-Channel Image       | Accuracy | 0.917  | 0.944 | 0.936 | 0.939 | 0.922 | 0.936 |
|                           | Sensitivity | 0.922  | 0.933 | 0.944 | 0.933 | 0.922 | 0.931 |
|                           | Specificity | 0.911  | 0.956 | 0.967 | 0.944 | 0.922 | 0.940 |
|                           | F1 Score | 0.919  | 0.933 | 0.950 | 0.936 | 0.922 | 0.933 |
| Complaints & Age & Gender | Accuracy | 0.739  | 0.789 | 0.667 | 0.833 | 0.722 | 0.750 |
|                           | Sensitivity | 0.767  | 0.744 | 0.756 | 0.844 | 0.756 | 0.773 |
|                           | Specificity | 0.711  | 0.833 | 0.578 | 0.822 | 0.689 | 0.727 |
|                           | F1 Score | 0.752  | 0.766 | 0.708 | 0.839 | 0.738 | 0.761 |

FIGURE 5. Validation loss during training.

Four crucial points should be noticed from Table 5 and Table 2. Firstly, lung window images can deal with 92.3% cases. That is why this window is adopted widely in clinical practice. Secondly, the high attenuation window is the second choice since it can deal with normal pneumonia cases. Thirdly, only a few patients have such severe pneumonia that causes abnormal decreases in lung density, which need a low attenuation window for observation. Fourthly, clinical chief complaints, and demographic information further improve the overall performance of the proposed framework.

The validation loss during one fold training is shown in Figure 5. According to this figure, the loss of the proposed framework drops more quickly than RCNN (ResNet), which means that clinical and demographic information can accelerate the converging speed.

VI. DISCUSSION

A. EFFECT OF DEMOGRAPHIC INFORMATION

In this section, we further discuss the effect of demographic information. Data distributions of demographic information are listed in Table 1 of Section III. According to this table, we observe some interesting patterns:

(i) A male patient has a more significant chance of getting pneumonia. In 547 male cases, about 56.1% of them get pneumonia; Meanwhile, in 353 female cases, only 40.5% of cases get pneumonia. This phenomenon might be related to smoking since males in China suffer a severe smoking problem.

(ii) The table shows that age is also associated with the chances of getting pneumonia. We can observe that people older than 40 have a much larger chance of getting pneumonia. Moreover, about 16.7% of people younger than 10 get pneumonia, higher than that in the age group of 10-20. According to the CDC, people older than 65 or younger than 5 are more likely to get pneumonia, and people who smoke are more likely to get pneumonia too. Although the details like age groups are different, the patterns we observe are basically consistent with the CDC data.

B. NUMBER OF LSTM UNITS IN RCNN

We vary the number of LSTM units in RCNN (ResNet) in the set {64, 128, 256, 512} to investigate its influence by running tests with three-channel images. According to our experiments, when the LSTM neuron number is set to 64, the accuracy is 0.925; When the LSTM number is set to 256 and 512, the accuracy drops to 0.910; When the LSTM neuron number is set to 128, the accuracy reaches the highest 0.93. As a result, we adopt 128 as our LSTM neuron number in Section V.

C. WEIGHTS OF DIFFERENT CHANNELS DURING LEARNING PROCESS

We further discuss the influences of different image channels by adding weights. The initial weights for different channels are set to 1, and these weights can be updated during the back-propagation procedure. Experiments demonstrate that weights for the different channels are slightly different, and have no effect on the final results. The weight for the low attenuation window drops to 0.993; The weight for high attenuation increases to 1.002; The weight for the lung window increases to 1.006. In general, all channels provide necessary features during the deep learning process, but the lung window is a little more important than other image windows. As a result, we remove weights for channels in Section V.
VII. LIMITATIONS AND FUTURE WORK

Even if our framework achieves promising results on the clinical data, there are still some shortcomings in our work. Firstly, we collect 900 cases in this study. Generally speaking, 900 cases are far smaller than ‘big data’, so the performance of our framework is restricted by data distribution and quality. Secondly, we only consider chest CT scans, chief complaints, gender, and age. In clinical practice, besides the tests mentioned above, patients usually need to take blood pressure measurements, blood tests, heartbeat measurements, and some other medical examinations. These examinations can help doctors gain a more objective and comprehensive understanding of the patient’s condition.

It is challenging to overcome these two shortcomings mentioned above because data collected from PACS are disorder. Constructing a big scale medical dataset with consistent data is a very challenging task since raw data is affected by radiologists’ habits, data acquisition equipment, and hospital workflows. Our future work will focus on finding a method that can perform accurate diagnose on disorder data and include more multi-modal information from medical tests.

VIII. CONCLUSION

In this study, we propose a novel multi-channel multi-modal deep regression framework for pneumonia screening, which combines multi-channel CT visual features with patients’ age, gender, and chief complaints to simulates clinical practice. The proposed framework extracts visual features from three-channel images, semantic features from chief complaints, and fuses these features with prior knowledge provided by age and gender.

We analyze the dataset collected from the First Affiliated Hospital of Army Medical University. Compared to the baseline, the proposed framework improves the accuracy by 2.3%, and significantly improves the sensitivity by 3.1%. Experiments demonstrate that the multi-channel images and multi-modal information have advantages over single-channel image data only, and the proposed framework has achieved very balanced performances.

REFERENCES

[1] N. Deepika, P. Vinupriha, and D. Kathirvelu, “Classification of lobar pneumonia by two different classifiers in lung CT images,” in Proc. Int. Conf. Commun. Signal Process. (ICCSSP), Apr. 2018, pp. 0552–0556.

[2] D. K. Iakovidis, S. Tsevas, M. A. Savelonas, and G. Papamichalis, “Image analysis framework for infection monitoring,” IEEE Trans. Biomed. Eng., vol. 59, no. 4, pp. 1135–1144, Apr. 2012.

[3] 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/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 9049–9058.

[4] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, May 2016.

[5] 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. (CVPR), Jul. 2017, pp. 3462–3471.

[6] T. Cherian, E. K. Mulholland, J. B. Carlin, H. Ostensen, R. Amin, M. De Campo, D. Greenberg, R. Lagos, M. G. Lucero, and S. A. Madhi, “Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies,” Bulletin World Health Org., vol. 83, no. 5, pp. 353–359, 2005.

[7] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, “Learning to diagnose from scratch by exploiting dependencies among labels,” 2017, arXiv:1710.10501. [Online]. Available: http://arxiv.org/abs/1710.10501

[8] P. Rajapurkar, 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: http://arxiv.org/abs/1711.05225

[9] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, and D. J. Mollura, “Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks,” Comput. Methods Biomech. Biomed. Eng., Imag. Vis., vol. 6, no. 1, pp. 1–6, Jan. 2018.

[10] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, “Multimodal neuromaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer’s disease,” IEEE J. Biomed. Health Informat., vol. 22, no. 1, pp. 173–183, Jan. 2018.

[11] 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. (CVPR), Jun. 2016, pp. 2497–2506.

[12] W. Hu, B. Cai, A. Zhang, V. D. Calhound, and Y.-P. Wang, “Deep collaborative learning with application to the study of multimodal brain development,” IEEE Trans. Biomed. Eng., vol. 66, no. 12, pp. 3346–3359, Dec. 2019.

[13] L. Xiao, J. M. Stephen, T. W. Wilson, V. D. Calhound, and Y.-P. Wang, “Alternating diffusion map based fused of multimodal brain connectivity networks for IQ prediction,” IEEE Trans. Biomed. Eng., vol. 66, no. 8, pp. 2140–2151, Aug. 2019.

[14] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, “Triple-classification of respiratory sounds using optimized S-Transform and deep residual networks,” IEEE Access, vol. 7, pp. 32845–32852, 2019.

[15] Z. Li, X. Li, Z. Zhu, S. Zeng, Y. Wang, Y. Wang, and A. Li, “Signal analysis of electrocardiogram and statistical evaluation of myocardial enzyme in the diagnosis and treatment of patients with pneumonia,” IEEE Access, vol. 7, pp. 113751–113759, 2019.

[16] M.-J. Lyu and S.-M. Yuan, “Cloud-based smart dog music therapy and pneumonia detection system for reducing the difficulty of caring for patients with dementia,” IEEE Access, vol. 8, pp. 20977–20990, 2020.

[17] Mesh: Medical Subject Headings. Accessed: 2019. [Online]. Available: https://www.nlm.nih.gov/mesh/meshhome.html

[18] M. Liu, J. Zhang, E. Adeli, and D. Shen, “Joint classification and regression via deep multi-task multi-channel learning for Alzheimer’s disease diagnosis,” IEEE Trans. Biomed. Eng., vol. 66, no. 5, pp. 1195–1206, May 2019.

[19] J. Wu, X. Liu, X. Zhang, Z. He, and P. Lv, “Master clinical medical knowledge at certificated-doctor-level with deep learning model,” Nature Commun., vol. 9, no. 1, p. 4352, Dec. 2018.

[20] J. Dukart, M. L. Schroeter, and K. Mueller, “Age correction in dementia–matching to a healthy brain,” PLoS ONE, vol. 6, no. 7, 2011, Art. no. e22193.

[21] M. de Bruijne, “Machine learning approaches in medical image analysis: From detection to diagnosis,” Med. Image Anal., vol. 33, pp. 94–97, Oct. 2016.

[22] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, “Long-term recurrent convolutional networks for visual recognition and description,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 2625–2634.

[23] N. Aafaq, N. Akhtar, W. Liu, S. Zulqarnain Gilani, and A. Mian, “Spatio-temporal dynamics and semantic attribute enriched visual encoding for video captioning,” 2019, arXiv:1902.10322. [Online]. Available: http://arxiv.org/abs/1902.10322

[24] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in Proc. Neural Inf. Process. Syst., 2014, pp. 568–576.
Q. Wang et al.: Deep Regression via Multi-Channel Multi-Modal Learning for Pneumonia Screening

[25] C. Feichtenhofer, A. Pinz, and A. Zisserman, “Convolutional two-stream network fusion for video action recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1933–1941.

[26] M. Zeirek, R. W. van Hamersveld, J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, “A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography,” IEEE Trans. Med. Imag., vol. 38, no. 7, pp. 1588–1598, Jul. 2019.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014, arXiv:1412.3555. [Online]. Available: http://arxiv.org/abs/1412.3555

[28] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, “Joint sequence learning and cross-modality convolution for 3D biomedical segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 6393–6400.

[29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Representations, 2015, pp. 1–14.

[30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.

[32] M. Lin, Q. Chen, and S. Yan, “Network in network,” in Proc. Int. Conf. Learn. Representations, 2014, pp. 1–10.

[33] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157–166, Mar. 1994.

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

[35] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” 2014, arXiv:1406.1078. [Online]. Available: http://arxiv.org/abs/1406.1078

[36] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 2013, arXiv:1301.3781. [Online]. Available: http://arxiv.org/abs/1301.3781

[37] 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.

[38] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” 2016, arXiv:1603.04467. [Online]. Available: http://arxiv.org/abs/1603.04467

[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.

CHEN LIU received the M.D. degree in medical imaging from Army Medical University, China, in 2015. He is currently an Attending Physician with the Radiology Department, Southwest Hospital, which is The first affiliated hospital of Army Medical University. He has hosted more than six research, including the National Natural Science Foundation and got funded more than 1.6 million. He has published six articles as first author. His current research interests include medical imaging, medical imaging, and deep learning.

XIAOHONG ZHANG received the M.S. degree in applied mathematics from Chongqing University, China, and the Ph.D. degree in computer software and theory, in 2006. He is currently a Professor and the Vice Dean of the School of Big Data and Software Engineering, Chongqing University. His current research interests include data mining of software engineering, topic modeling, image semantic analysis, and video analysis.

DAN YANG received the B.Eng. degree in automation, the M.S. degree in applied mathematics, and the Ph.D. degree in machinery manufacturing and automation from Chongqing University, Chongqing. From 1997 to 1999, he held a post-doctoral position at the University of Electro-Communications, Tokyo, Japan. He is currently the President of Southwest Jiaotong University. He is also a Professor with the School of Big Data and Software Engineering, Chongqing University. He has authored over 100 scientific articles and some of them are published in some authoritative journals and conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, CVPR, and BMVC. His research interests include computer vision, image processing, pattern recognition, software engineering, and scientific computing.

CHEN LIU received the M.D. degree in medical imaging from Army Medical University, China, in 2015. He is currently an Attending Physician with the Radiology Department, Southwest Hospital, which is The first affiliated hospital of Army Medical University. He has hosted more than six research, including the National Natural Science Foundation and got funded more than 1.6 million. He has published six articles as first author. His current research interests include brain functional MRI, clinical data mining, medical imaging, and deep learning.