

Automated ICD-9 Coding via A Deep Learning Approach

Min Li¹, Zhihui Fei, Min Zeng, Fang-Xiang Wu², Yaohang Li³, Yi Pan⁴, and Jianxin Wang⁵

Abstract—ICD-9 (the Ninth Revision of International Classification of Diseases) is widely used to describe a patient’s diagnosis. Accurate automated ICD-9 coding is important because manual coding is expensive, time-consuming, and inefficient. Inspired by the recent successes of deep learning, in this study, we present a deep learning framework called DeepLabeler to automatically assign ICD-9 codes. DeepLabeler combines the convolutional neural network with the ‘Document to Vector’ technique to extract and encode local and global features. Our proposed DeepLabeler demonstrates its effectiveness by achieving state-of-the-art performance, i.e., 0.335 micro F-measure on MIMIC-II dataset and 0.408 micro F-measure on MIMIC-III dataset. It outperforms classical hierarchy-based SVM and flat-SVM both on these two datasets by at least 14 percent. Furthermore, we analyze the deep neural network structure to discover the vital elements in the success of DeepLabeler. We find that the convolutional neural network is the most effective component in our network and the ‘Document to Vector’ technique is also necessary for enhancing classification performance since it extracts well-recognized global features. Extensive experimental results demonstrate that the great promise of deep learning techniques in the field of text multi-label classification and automated medical coding.

Index Terms—ICD-9 coding, multi-label classification, convolutional neural network, document to vector

1 INTRODUCTION

WITH the rapid development of medical field, the Ninth Revision of International Classification of Diseases (ICD-9) codes are widely used to describe a patient’s diagnosis including symptoms, statistical analysis of mortality rate and medical reimbursement [1]. ICD-9 codes mean that each disease has a unique code and is used in the electronic health records as a billing mechanism. Usually, ICD-9 codes are undertaken by the coders of the hospital’s Medical Record Department who assign ICD-9 codes to medical record according to a doctor’s clinical diagnosis. However, because they need to master specialized skills such as the knowledge in the field of medicine, coding rules and medical terminologies, manual coding is expensive, time-consuming and inefficient. Considering these constraints, it is urgent to develop an accurate computational approach for automated ICD-9 coding.

Developing an accurate computational system to support automated ICD-9 coding based on medical diagnosis document is still a challenging task. The automated ICD-9 coding

task usually has the following problems: 1) Patient’s clinical records is not always structured in the same way. It is very difficult to extract important and relevant knowledge from various kinds of medical records effectively. 2) The medical field has a lot of terminologies, which is difficult for non-professionals to understand the meaning of these terminologies. Additional tools are needed to interpret the some terms and symptoms and to get semantic information from medical records. 3) Each physician usually has his own way to describe symptoms. Therefore, even for the same disease, there are many different ways to describe it.

Since the early 1990s, many scientists have explored how to automatically assign ICD-9 codes based on clinical records. In recent years, more and more scientists applied machine learning methods to automatically assign ICD-9 codes such as support vector machine (SVM) [2], [3], [4], [5], Naive Bayes [6], [7], k-nearest neighbors [8], [9], topic model [10], [11], and other techniques based on Natural Language Processing [12], [13], [14], [15]. Perotte et al. [5] proposed novel evaluation metrics, which reflect the distances among gold-standard and predicted codes and their locations in the ICD-9 tree. They used a support vector machine (SVM) classifier to make prediction. Medori et al. [6] conducted four experiments based on Naive Bayes with varying sets of attributes. Their experiments obtained recall of 81.8 percent and demonstrated the importance of powerful machine learning method and stemming features. Ruch et al. [8] combined a set of machine learning (k-nearest neighbors) and data-poor methods to generate a single automatic text categorizer. The combined ranking system obtained a precision of 75 percent at high ranks and a recall of about 63 percent for the top twenty returned codes. Chen et al. [12] employed a deep-level semantic analysis approach that involves dependency parsing, parse tree matching, and semantic

- M. Li, Z. Fei, M. Zeng, and J. Wang are with the School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, P.R. China. E-mail: {limin, jxwang}@mail.csu.edu.cn, zhihuiifei@foxmail.com, zengmin1990@163.com.
- F.X. Wu is with the Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N5A9, Canada. E-mail: faw341@mail.usask.ca.
- Y. Li is with the Department of Computer Science, Old Dominion University, Norfolk, VA 23529. E-mail: yaohang@cs.odu.edu.
- Y. Pan is with the Department of Computer Science, Georgia State University, Atlanta, GA 30302. E-mail: yipan@gsu.edu.

Manuscript received 17 Oct. 2017; revised 4 Mar. 2018; accepted 15 Mar. 2018. Date of publication 20 Mar. 2018; date of current version 5 Aug. 2019. (Corresponding author: Min Li.)

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TCBB.2018.2817488

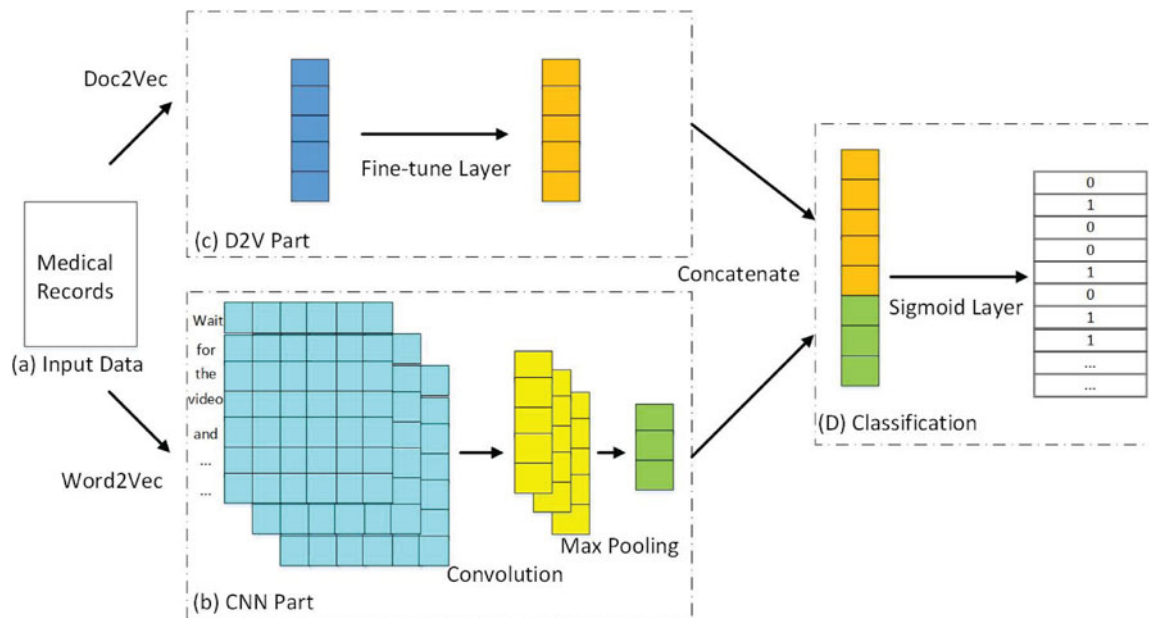


Fig. 1. Overview of proposed DeepLabeler methodology. The input is a discharge summary ((a) input data), through D2V part ((b) D2V part, which is used to extract local features) and CNN part ((c) CNN part, which is used to extract global features), then (d) outputs of two parts concatenated to a whole vector. The vector is fed into the fully connected layer for classification.

matching score calculation between training and testing corpus, which overcomes the knowledge acquisition bottleneck in an unsupervised method. Pereira et al. [13] proposed a text mining approach by using processed electronic medical records based on K-Nearest Neighbor classifier. Their results suggest a good performance proposing a diagnosis from electronic medical records. Chen et al. [15] presented an improved approach based on the Longest Common Subsequence and semantic similarity for automatic diagnoses, their method can increase the accuracy of processing in Chinese disease mapping. Ira Goldstein et al. [16] applied three systems (search engine, boosting algorithm and rule-based model) for automatically predicting the ICD-9 codes. Their evaluation shows that semantic information significantly contributes to ICD-9 coding with lexical elements.

To further improve machine learning methods for automatically assigning ICD-9 codes, we borrow the ideas from very recent breakthrough in deep learning [17]. In the past few years, deep learning techniques, particularly convolutional neural networks (CNN), have demonstrated to be effective for computer vision tasks such as medical image classification [18], [19], object detection [20] and image segmentation [21] and so on. The main advantage of CNN is its capability of automatic detection of image local features and simplification of the conventional image analysis pipelines. It is well known that local context feature is critical for automatic ICD-9 code assignments from the discharge summary contents. Specifically, medical terminologies which consist of words are probably the most effective features for ICD-9 code assignments. CNN can be applied to extracting local contextual features from medical texts. In this study, we used CNN to extract local features from patient records.

On the other hand, global features also are vital for automatic ICD-9 code assignments. We innovatively apply the 'Document to Vector' (D2V) technique for automatically assigning ICD-9 codes. The 'Document to Vector' is an unsupervised algorithm, which was proposed by Le et al. [22] to

learn fixed-length feature representation from variable-length documents. In D2V training processing, each document and word is encoded as a unique dense vector. We call the document vector as document embedding (DE) and word vector as word embedding (WE). These vectors are concatenated together to predict the next word in its context. D2V can capture effective global features of given medical text. Specifically, D2V generates document vectors and word vectors which are stored in the training dataset. Then, each document vector in the training dataset can be regarded as a summary of a medical text. Finally, we can use them to train CNN for prediction. To the best of our knowledge, D2V technique has never been utilized in the task of ICD-9 assignment.

In this paper, we propose an end-to-end deep learning framework, so-called DeepLabeler, which combined CNN with D2V technique to automatically assign ICD-9 codes. DeepLabeler combines the advantages of CNN and D2V technique, which captures not only complicated local text relationships, but also models document topics in a neural network. We perform extensive experiments on MIMIC-II and MIMIC-III datasets [23] and the results demonstrate that our DeepLabeler outperforms the hierarchy-based SVM [5] with a relatively strict Micro F-measure of 0.335 and 0.408, respectively. Furthermore, we find that DeepLabeler performs better when the number of dataset samples becomes larger.

2 METHODS

2.1 Overview: DeepLabeler

As shown in Fig. 1, DeepLabeler consists of two steps for automatic ICD-9 code assignment. The first step is feature extraction while the second one is multi-label classification. Feature extraction has two parallel parts: (a) the CNN part to extract rich local features, and (b) the D2V part to capture semantic information in a whole document by taking the ordering of the words into consideration. The classification step uses a fully connected layer with the sigmoid activation

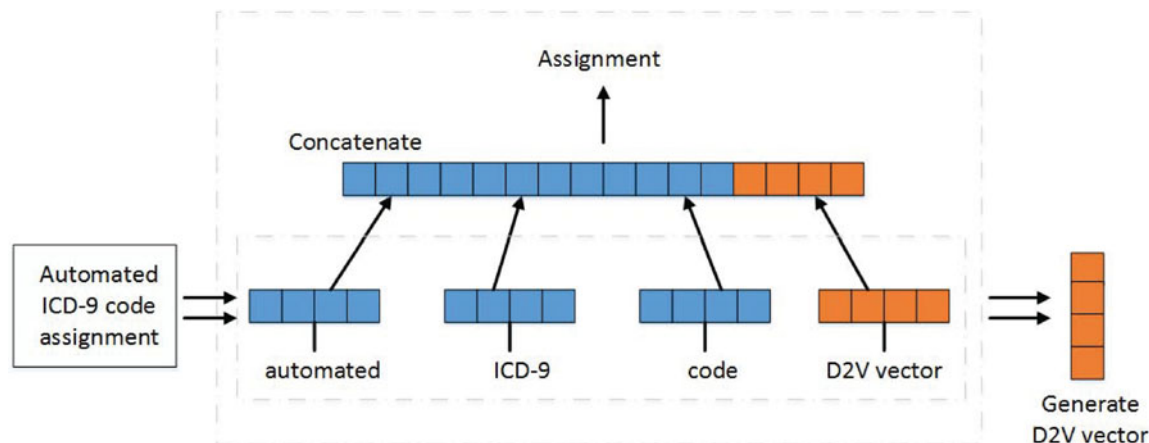


Fig. 2. The network architecture of generating a document vector.

function to predict the probability of each ICD-9 code. Specifically, we need to pre-train the word vectors (WE) and document vectors (DE) before training DeepLabeler. Given a document, we use the word vectors to train the CNN part and obtain a highly informative feature vector which can represent the whole document. At the same time, the document vector is fine tuned by one fully connected layer of D2V part and then is concatenated with another feature vector generated by CNN. Finally, a full connection layer is stacked to output a result vector which can represent the probability of each ICD-9 code. The sigmoid activation function is applied in the output layer which is the practice of multi-label learning. Finally a code is assigned when the output score is greater than a pre-determined threshold.

2.2 Preliminaries

Automated ICD-9 coding is a multi-label classification problem. CNN is well-accepted for document classification recently as it can capture phrase-level or sentence-level features [24], [25], [26], [27]. Besides, D2V technique, as a text representation approach, can obtain diverse semantic representation for documents from different clinical free texts [22]. These two approaches are combined to use for constructing the feature selection part of deep learning model. D2V technique uses the entire database to train the deep learning model and obtains document vectors which capture global features and semantic information. CNN utilizes convolutional kernels to extract the features of sentences and obtains feature vectors which capture local features of documents. The two types of vectors are complementary in our proposed model. The details of these two approaches are as follows.

2.2.1 Document to Vector

Word representation technique plays an important role in many natural language processing tasks, e.g., text classification, sentiment analysis. Recently, more and more word representation technique, such as Word2Vec, and Document2Vec (D2V), has been proposed to handle natural language understanding task [22], [28], [29]. These dense vectors have potential to improve the performance of automated ICD-9 code assignments. D2V, a deep learning method, learns dense vector representation of documents at the semantic level and achieve good performance in many researches [30], [31], [32]. Fig. 2 illustrates the network

architecture of generating a document vector. In this framework, every word and the whole document is represented a unique dense vector. Then the document and words vectors are concatenated to predict the next word in its context. In this way, the word order is kept, and thus better vector representations can be obtained. A fully connected neural network with one hidden layer is used to train the model. Back propagation algorithm with stochastic gradient descent is used in the fully connected neural network. The learning methods of document vector and word vectors are regarded as unsupervised preprocessing of clinical free texts.

2.2.2 Convolutional Neural Network

CNN is a class of deep, feed-forward artificial neural networks that have successfully been applied to computer vision [33], [34]. CNN models are widely used in various Natural Language Processing (NLP) problems and have achieved promising results in semantic parsing [35], search query retrieval [36], sentence modeling [25], sentence classification [24], prediction [37] and other traditional NLP tasks [38]. CNNs utilizes layers with convolving filters that are applied to local features [39]. This means that the network automatically learns the features that were hand-engineered in traditional algorithms. The effectiveness of local features and independence from human effort in feature design are two major advantages in CNN. Fig. 3 illustrates how CNN is used for document classification and local features extraction from a document. In NLP task, local features can be regarded as word-level or phrase-level features for a given document.

Before using the CNN model, it is general to translate each word in the documents into a low dimensional dense word vector using W2V technology, and then all the word vectors in the same document can make up a matrix. The document matrix is finally fed to the CNN model for generating a fixed length vector. The output of this process is a feature vector which can represent the document in a low dimensional dense form. Unlike D2V, it is a process of supervised learning based on stochastic gradient descent and back propagation.

2.3 The Benefits of DeepLabeler

CNN can extract rich local features, but has two major defects in extracting features of medical data in our task.

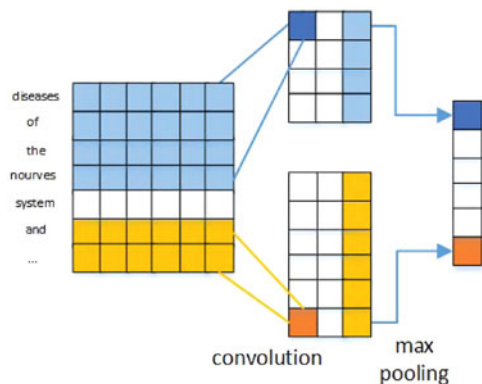


Fig. 3. The process of extracting features from CNN.

First, CNN ignores the semantic features of the full text because it does not take into account the order of words or phrases. It can be considered as a variant of the ‘bag of word’ representation while capturing more informative features than unigram or bigram models. Second, the number of words in a discharge summary note varies from dozens to thousands, but CNN model only takes a unchanged shape matrix as input. This means that a large amount of document matrix must be either truncated or padded by some zeros, which may lose a certain proportion of the original information in the document. Therefore, the performance of the CNN model may be seriously affected when the length of each document varies significantly. Foreseeing these problems, we realize that the D2V part can be an important complement to the CNN part. Fortunately, D2V is able to avoid the above two problems. D2V captures semantic information of a whole document by taking the order of words into consideration. More importantly, the D2V part put all words in a document for training, thus it would not discard any useful word information, which is not possible for CNN part. Therefore, we are inspired to integrate the CNN and D2V parts to achieve the better performance in this multi-label classification and natural language understanding task.

3 EXPERIMENTS

3.1 Dataset

MIMIC (Multi-parameter Intelligent Monitoring in Intensive Care) is a real-world Intensive Care Unit (ICU) database and it is a popular database for studying automated ICD-9 coding. It is a publicly available database developed by the MIT Lab for Computational Physiology, comprising de-identified health data associated with tens of thousands critical care patients [23]. It includes discharge summaries, diagnostic codes, vital signs, laboratory measurements, etc. The latest version of MIMIC is MIMIC-III (<http://mimic.physionet.org/>), which comprises over 58,000 hospital admissions [23]. The data spans June 2001-October 2012. The MIMIC-II is an older version from 2001 to 2008. Because these medical records are from the ICU patients, the distribution of the diseases has a certain tendency that serious diseases are likely to be overrepresented. In MIMIC, there is a large amount of data that is not relevant to disease classification. In order to facilitate the study, we focus on understanding the clinical free texts, i.e., discharge summaries only.

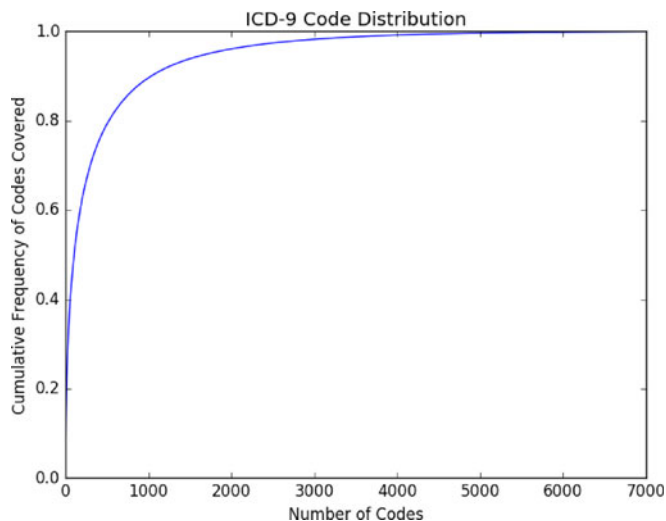


Fig. 4. Cumulative distribution of ICD-9 codes of MIMIC-III.

Each patient in the MIMIC database is associated with a list of ICD-9 codes that are the labels in the experiments. One ICD-9 code can indicate a classification of a disease, diagnostic or symptom, injury or a treatment procedure. This is a typical NLP task of text semantic parsing and multi-label classification [40], [41]. We use the discharge summary of each patient as the training data for our model.

In MIMIC-III, the longest discharge summary contains 7,980 words while the shortest one has only 9 words before data preprocessing. The database currently contains 6,984 ICD-9 codes, showing a highly skewed distribution, with the top 105 codes making up 50 percent of the total codes. Fig. 4 shows a cumulative distribution function curve of the ICD-9 codes. A brief description of the statistics of MIMIC-III dataset is summarized in Table 1.

The automated ICD-9 coding is a multi-label classification problem [42], [43], [44]. Each medical record has multiple ICD-9 codes and each ICD-9 code is one of the multiple labels in classification. Although some problems of automated ICD-9 coding were mentioned in the Introduction section, now some specific ICD-9 coding problems related to the MIMIC dataset need to be taken care of. There are three factors that make automated ICD-9 coding difficult. First, the number of ICD-9 codes is large and the distribution of ICD-9 codes is a serious biased distribution. Table 2 shows six ICD-9 codes, ranked as 1st, 10th, 50th, 100th, 500th and 1,000th in terms of their frequencies in all 58,929 medical records from our dataset. The most frequent ICD-9 code, 401.9 (Hypertension), appears in 35.1 percent medical records, while the 1,000th frequent ICD-9 code, 999.8 (Other

TABLE 1
Basic Statistics of the MIMIC-III Dataset

Total # of discharge summary	52,962	Total # of codes	6,984
Avg. # of words per discharge summary	1,524	Avg. # of codes per patient	11
Max # of words per discharge summary	7,980	Max # of codes per patient	39
Min # of words per discharge summary	9	Min # of codes per patient	1

TABLE 2
The 1st, 10th, 50th, 100th, 500th and 1000th ICD-9 Code
in Terms of the Frequencies of Appearances in
58,929 Medical Records from MIMIC-III

Rank	ICD-9 Code	Frequency of Patients with Code
1	401.9: Hypertension	0.3513
10	530.81: Esophageal reflux	0.1073
50	276.7: Hyperpotassemia	0.0368
100	V10.46: Personal history of malignant neoplasm of prostate	0.0205
500	785.2: Undiagnosed cardiac murmurs	0.0037
1,000	999.8: Other and unspecified transfusion reaction not elsewhere classified	0.0014

and unspecified transfusion reaction not elsewhere classified), appears in 0.14 percent of the medical records only. This means that most of ICD-9 codes only have a few annotated medical records and thus it causes a serious imbalance of dataset distribution. Second, each patient has different number of ICD-9 codes which varies greatly. Some patients may have up to 39 associated codes, while the others may have only one code. Third, the average of number of samples of each code is small. The number of codes is 6,984 and the number of samples is 52,962, thus the average number of samples of each code is only about 7.58, which indicates most of them lack enough samples for training and leads to a poor performance of classification.

3.2 Implementation

We used an open-source tool, Tensorflow [45] (<https://www.tensorflow.org/>), to implement the CNN model. Document embedding and word embedding were implemented by using Gensim (<http://radimrehurek.com/gensim>). The skip-gram model was used to train the word embedding vectors in Gensim.

3.3 Evaluation Metrics

Choosing appropriate evaluation metrics in multi-label classification is important to evaluate the performance of classifiers. There are many types of evaluation metrics in multi-label classifications [46]. We used micro-average precision, micro-average recall and micro-average F-measure to evaluate the performance of our model. They are defined as follows:

$$MiP = \frac{\sum_{m=1}^M \sum_{i=1}^N y_i^m \hat{y}_i^m}{\sum_{m=1}^M \sum_{i=1}^N \hat{y}_i^m} \quad (1)$$

$$MiR = \frac{\sum_{m=1}^M \sum_{i=1}^N y_i^m \hat{y}_i^m}{\sum_{m=1}^M \sum_{i=1}^N y_i^m} \quad (2)$$

$$Micro\ F\text{-measure} = \frac{2 \times MiP \times MiR}{MiP + MiR}, \quad (3)$$

where M is the number of all ICD-9 codes, N is the number of all samples, y_i and $\hat{y}_i \in \{0, 1\}^M$ are the true and predicted label for sample i , respectively.

F-measure is the harmonic mean of precision and recall, which is a good indicator for the overall predictive power of models. It is the most important metric in automated ICD-9

TABLE 3
Model Performance of the ICD-9 Coding on Test Sets of
MIMIC-III, Which Contains 6,984 Associated Codes

	Micro-average Precision	Micro-average Recall	Micro-average F-measure
flat-SVM	0.635	0.158	0.253
hierarchy-based SVM	0.415	0.280	0.335
D2V+CNN (threshold = 0.2)	0.486	0.351	0.408
D2V+CNN (threshold = 0.3)	0.555	0.292	0.383

coding task[47]. Our model returns the probability for each ICD-9 code, which allows for further tuning to optimize the precision and recall. This tuning is carried out by specifying a threshold in this study, as the models have already been optimized for F-measure. Micro-averaged F-measure is chosen for our interest in predicting correct ICD-9 codes for as many patients as possible, rather than ensuring good coverage of the different classes.

The F-measure averaged over single codes has been widely used in multi-label learning. For instance, Perotte et al. used it to evaluate the performance of the flat and hierarchy-based SVM model [13]. BioASQ challenge, an international competition about automatically annotating new MEDLINE citations using MeSH, took the micro F-measure as the main evaluation metric [48].

3.4 Baseline Model

Perotte et al. [13] used flat SVM as a baseline model on MIMIC-II dataset (22,815 documents) with documents represented using a ‘bag of words’ (BoW) model [49] and proposed a hierarchy-based SVM which obtained a higher F-measure value than flat SVM. The flat SVM treats each ICD-9 code independently of each other, but hierarchy-based SVM leverages the hierarchical nature of ICD-9 codes into its modeling. The hierarchy-based SVM tends to achieve a higher recall and F-measure at the tradeoff of precision in their conclusions. Both models will be compared on MIMIC dataset in this study.

3.5 Experiment on MIMIC-III

In order to illustrate the effectiveness of our proposed model, we study automated ICD-9 coding by using discharge summaries from the publicly available MIMIC-III dataset. We extracted 52,962 discharge summaries from the MIMIC-III, and a total number of ICD-9 codes are 6,984. We split the total discharge summaries into a training set which consists of 47,665 documents and a test set which has 5,297 documents, the experimental results are shown in Table 3.

It is important to note that the CNN architecture we used in this medical coding problem is the classic single-layer multi-channel convolutional network, which has been proved to have a good performance in text classification and sentiment analysis [24], [26], [50]. To obtain the best performance of DeepLabeler, we searched for a set of various parameters of network architectures to find the best parameters for automated ICD-9 coding. These parameters include: 1) CNN part: word embedding size, the

TABLE 4
The Best Parameters of Network Architecture (CNN Part) on MIMIC-III

	CNN part
word embedding size	100
Maximum length	700
convolutional kernel size	3, 4, 5
convolutional filter number	64, 64, 64
convolutional layer number	1
dropout rate	0.75

TABLE 5
The Best Parameters of Network Architecture (D2V Part) on MIMIC-III

	D2V part
document embedding size	128
the number of neurons in hidden layer	64
dropout rate	0.75

maximum length of notes, convolutional kernel size, convolutional filter number, convolutional layer number and dropout rate; 2) D2V part: document embedding size, the number of neurons in the hidden layer. The optimal structure (see Tables 4 and 5) was found to be: word embedding size (CNN part): 100, the maximum length of notes, (CNN part): 700, convolutional kernel size (CNN part): 3, 4, 5, convolutional filter number (CNN part): 64, 64, 64, convolutional layer number (CNN part): 1, dropout rate (CNN part): 0.75, document embedding size (D2V part): 128, the number of neurons in hidden layer (D2V part): 64, and dropout rate (D2V part): 0.75.

Besides tuning different parameters of network architectures, choosing an appropriate threshold is crucial for classification when using deep learning models. A code is classified as a positive sample only if its score generated by a model is greater than the threshold. In order to obtain the best F-measure, our model has run with different threshold values. Figs. 5c and 5d represent the effect of threshold parameter on the classification of MIMIC-III where threshold takes the values between 0.1 and 0.95. As a result of experiments, the best result is obtained with the threshold value of 0.2. When the threshold is equal to 0.2, our model outperforms the hierarchy-based SVM with F-measures of 0.408 and 0.335, respectively. Notice that the evaluation metrics here do not use the hierarchical structure of ICD-9 code tree but instead only use the exact gold-standard matches count as true positives. This is usually stricter than using other hierarchical evaluation approaches [46], [51].

3.6 Experiment on MIMIC-II

Because the MIMIC-III dataset is a relatively newer dataset, most researchers still use MIMIC-II dataset which is widely used in the study of ICD-9 coding in the last decade. In order to better evaluate the performance of our model, we also applied DeepLabeler on MIMIC-II dataset. There are 22,815 discharge summaries with diagnosis code were collected from MIMIC-II. A total number of diagnosis codes are 5,031. We split the data into a training set and a test set consisting of 20,533 and 2,282 documents of the full dataset, respectively.

The same as the operation carried out on MIMIC-III, we first tune different parameters of network architectures (see Tables 6 and 7). The best model parameters are found to be: word embedding size (CNN part): 100, the maximum length of notes (CNN part): 700, convolutional kernel size

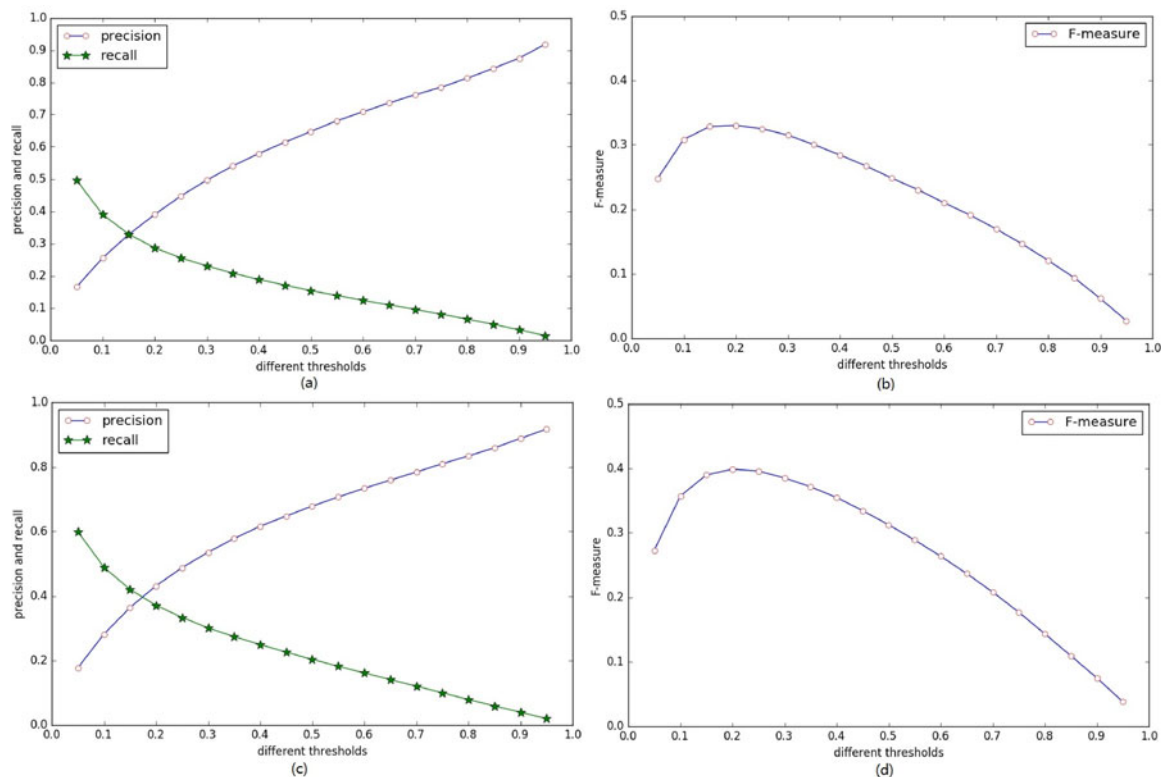


Fig. 5. Model performance of the ICD-9 coding on test sets of MIMIC-II (a), (b) and MIMIC-III (c), (d).

TABLE 6
The Best Parameters of Network Architecture
(CNN Part) on MIMIC-II

	CNN part
word embedding size	100
Maximum length	700
convolutional kernel size	1, 3, 5
convolutional filter number	64, 64, 64
convolutional layer number	1
dropout rate	0.9

(CNN part): 1, 3, 5, convolutional filter number (CNN part): 64, 64, 64, convolutional layer number (CNN part): 1, dropout rate (CNN part): 0.9, document embedding size (D2V part): 128, the number of neurons in hidden layer (D2V part): 64, dropout rate (D2V part): 0.9. Then we test the threshold parameter over different values. Figs. 5a and 5b show the effect of threshold parameter on the classification of MIMIC-II where the threshold value ranges from 0.1 to 0.95. Note that the best value is no longer 0.2, but 0.25. When the threshold is 0.25, our model outperforms the hierarchy-based SVM with F-measures of 0.335 over 0.293. Both precision and recall are higher than the hierarchy-based SVM. The experimental result on MIMIC-II is obviously poorer than MIMIC-III because the number of training data of MIMIC-II is only a half of that of MIMIC-III.

3.7 Result Analysis

Tables 3 and 8 compare the micro-average precision, micro-average recall and micro-average F-measure of our method with the other baseline model on MIMIC-III and MIMIC-II datasets, respectively. Our method significantly outperforms the flat-SVM and hierarchy-based SVM on the two datasets. Note that micro-average F-measure is the most important metric. On MIMIC-III dataset, DeepLabeler achieves micro-average precision, micro-average recall and micro-average F-measure of 0.486, 0.351 and 0.408, respectively, significantly outperforming flat-SVM (0.635, 0.158 and 0.253, respectively) and hierarchy-based SVM (0.414, 0.280 and 0.335, respectively). On MIMIC-II dataset, DeepLabeler obtains micro-average precision, micro-average recall and micro-average F-measure of 0.475, 0.258 and 0.335, respectively, performing better than flat-SVM (0.562, 0.130 and 0.211, respectively) and hierarchy-based SVM (0.395, 0.233 and 0.293, respectively). These micro-average F-measure values have showed that DeepLabeler is much better than flat-SVM and hierarchy-based SVM. Compared the results of MIMIC-III with those of MIMIC-II, we can see that the performance of DeepLabeler on MIMIC-III is better than MIMIC-II. We believe that it is because the number of samples of MIMIC-III is bigger than that of MIMIC-II due to

TABLE 7
The Best Parameters of Network Architecture
(D2V Part) on MIMIC-II

	D2V part
document embedding size	128
the number of neurons in hidden layer	64
dropout rate	0.9

TABLE 8
Model Performance of the ICD-9 Coding on Test Sets of
MIMIC-II, Which Contains 5,031 Associated Codes

	Micro- average Precision	Micro- average Recall	Micro- average F-measure
flat-SVM	0.562	0.130	0.211
hierarchy-based SVM	0.395	0.233	0.293
D2V+CNN (threshold = 0.25)	0.475	0.258	0.335
D2V+CNN (threshold = 0.5)	0.616	0.159	0.253

the fact that the deep learning model is often more effective when more data are available.

To discover the vital elements in the success of our proposed model, we conduct a study by removing individual components in our network. Specifically, we have tested the performance of models without the CNN part or D2V part. From the results on the MIMIC-III dataset presented in Table 9, we can see that the CNN part is the most effective component in our network as the micro-average precision, micro-average recall and micro-average F-measure drop to 0.375, 0.261 and 0.308, respectively (without CNN part). D2V is also important as the micro-average precision, micro-average recall and micro-average F-measure drop to 0.440, 0.366 and 0.399, respectively (without D2V part). We conclude that compared with CNN part, the D2V part in our model is necessary for dealing with global features, which contributes to classification performance enhancement.

Our results indicate that the proposed model predicts ICD-9 codes with a much higher F-measure than the traditional methods such as flat SVM and hierarchy-based SVM. Because the semantic features of text are difficult to extract manually, using SVM only is not able to effectively address this NLP task of text semantic parsing and multi-label classification, even considering the hierarchical structure of the ICD-9 codes. Our model ignores the hierarchical information, but it naturally takes into account the dependent relationship among codes. This multi-label learning method can be considered as a multi-task learning process, and each label prediction task is not independent with each other [52]. As there may be a close link between different diseases, we should consider this correlation between different ICD-9 codes when constructing a new model. Besides the correlation between ICD-9 codes, our model considers both local features by CNN and the global features by D2V. The proposed model captured the word-level features, phrase-level features and their order especially the semantic features of the document. It is important to note that F-measure is not very high because we used relatively strict

TABLE 9
Model Performance of the ICD-9 Coding on Test Sets of
MIMIC-III, Which Contains 2,282 Discharge Summaries
and 6,984 Associated Codes

Model	Micro- average Precision	Micro- average Recall	Micro- average F-measure
Only using CNN	0.440	0.366	0.399
Only using D2V	0.375	0.261	0.308
DeepLabeler (CNN+D2V)	0.486	0.351	0.408

evaluation metrics other than looser metrics used in other multi-label or hierarchical classification. The limited number of documents and the unbalanced ratio of ICD-9 codes in MIMIC dataset are the main reasons for the poor F-measure in the ICD-9 automatic coding. To better evaluate the performance of our model, we compare the experiment results on MIMIC-II and MIMIC-III, and find the performance of our model is greatly improved (with F-measure of 0.408 versus 0.335) when training with the data of double size. We believe that DeepLabeler has potential to acquire better performance if sufficient good-quality data are available.

DeepLabeler, unlike traditional models, uses deep learning technique to extract features rather than BoW representation. BoW representation has difficulty in capturing rich semantic information in large-scale clinical free texts. The basic idea of DeepLabeler is to integrate two types of diverse features to improve the performance of automated ICD-9 coding. DeepLabeler shows nearly 21 percent improvement (0.408 versus 0.335) in F-measure over hierarchy-based SVM. DeepLabeler uses two types of features: global features and local features, which are both obtained by deep learning methods, meaning that they can be extracted from the training data without human participation. The global features are from an unsupervised learning model trained by all words in documents and their appearance order is considered, while the local features are captured from the word-level and phrase-level information without the order information. The good performance gained by DeepLabeler is due to the diversity and the accuracy of the captured features, which complement to each other.

We emphasize that our strategy of capturing semantic information of free texts in medical records is very efficient, bringing about great improvements in model performance, as shown in our experiments. To the best of our knowledge, DeepLabeler is the first method to directly incorporate two diverse semantic features implemented by deep learning in automated ICD-9 coding.

4 CONCLUSION

Hospitals and healthcare providers rely on medical coding to record medical services and related causes and conditions during a patient visit. This coding process is normally done manually and thus is very time-consuming. To reduce the time and financial cost of manual coding, it is necessary to develop an automatic ICD-9 coding system. In this problem, each ICD-9 code can be regarded as a class label and each medical record can have multiple ICD-9 codes, and so the ICD-9 automatic coding is a large-scale multi-label classification problem. In this study, we presented DeepLabeler, which combined CNN part with D2V part to extract local and global features. It obtained a better performance on MIMIC-III dataset over the traditional methods if only using discharge summaries from each patient. Our computational results show that DeepLabeler achieved about 15 percent increase in micro F-Measure over flat SVM or hierarchy-based SVM. To better evaluate the proposed model, we also applied it on MIMIC-II dataset and obtained 0.335 F-measure value. The most important idea of DeepLabeler is to integrate two types of diverse features: local features for words or phrase, global features for document information. These new features might shed light on developing efficient algorithms for other multi-label classification problems in the field

of clinical free texts, with a large amount of training instances. We analyze the deep neural network structure and find that CNN is the most effective component in our network, which is quite intuitive and makes sense. The D2V part in our model is necessary for enhancing classification performance since it extracts well-recognized global features.

There are two advantages of DeepLabeler. The first one is that it can automatically extract effective representative features. The second one is that we only need to train an end-to-end model rather than thousands of binary classifiers, and considered the dependency relationship between each labels naturally. Therefore, DeepLabeler has the potential to be applied to more interesting areas to understand and analyze medical record.

In the future, it would be interesting to further explore the limitations of DeepLabeler and improve the coding performance by incorporating other types of features and other typical deep learning architectures, such as LSTM model [53], [54], [55], attention model [27], [55], [56] which recently achieved excellent results in addressing natural language understanding tasks. We also expect that learning task can be further improved by using more powerful architecture if we have more high-quality data. In particular, we plan to apply DeepLabeler to address other NLP tasks about medical records or literatures, for example, disease inference [57] and MeSH indexing [30], [58].

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61622213, No. 61420106009, 111 Project (No. B18059) and the Fundamental Research Funds for the Central Universities-Freedom Explore Program of Central South University under Grant No. 2018zzts563.

REFERENCES

- [1] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genetics*, vol. 13, no. 6, 2012, Art. no. 395.
- [2] J. C. Ferrão, F. Janela, M. D. Oliveira, and H. M. Martins, "Using structured EHR data and SVM to support ICD-9-CM coding," in *Proc. IEEE Int. Conf. Healthcare Informat.*, 2013, pp. 511–516.
- [3] S. Wang, X. Li, L. Yao, Q. Z. Sheng, G. Long, et al., "Learning multiple diagnosis codes for ICU patients with local disease correlation mining," *ACM Trans. Knowl. Discovery Data*, vol. 11, no. 3, 2017, Art. no. 31.
- [4] Y. Yan, G. Fung, J. G. Dy, and R. Rosales, "Medical coding classification by leveraging inter-code relationships," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 193–202.
- [5] A. Perotte, R. Pivovarov, K. Natarajan, N. Weiskopf, F. Wood, and N. Elhadad, "Diagnosis code assignment: Models and evaluation metrics," *J. Amer. Med. Informat. Assoc.*, vol. 21, no. 2, pp. 231–237, 2013.
- [6] S. V. Pakhomov, J. D. Buntrock, and C. G. Chute, "Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques," *J. Amer. Med. Informat. Assoc.*, vol. 13, no. 5, pp. 516–525, 2006.
- [7] J. Medori and C. Fairon, "Machine learning and features selection for semi-automatic ICD-9-CM encoding," in *Proc. NAAACL HLT 2nd Louhi Workshop Text Data Mining Health Documents*, 2010, pp. 84–89.
- [8] P. Ruch, J. Gobeill, I. Tbahriti, and A. Geissbühler, "From episodes of care to diagnosis codes: Automatic text categorization for medico-economic encoding," in *Proc. Amer. Med. Informat. Assoc. Annu. Symp.*, 2008, Art. no. 636.
- [9] M. Erraguntla, B. Gopal, S. Ramachandran, and R. Mayer, "Inference of missing ICD 9 codes using text mining and nearest neighbor techniques," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, 2012, pp. 1060–1069.

- [10] M. Dermouche, J. Velcin, R. Flicoteaux, S. Chevret, and N. Taright, "Supervised topic models for diagnosis code assignment to discharge summaries," in *Proc. 17th Int. Conf. Intell. Text Process. Comput. Linguistics*, 2016, pp. 485–497.
- [11] A. Perotte, F. Wood, N. Elhadad, and N. Bartlett, "Hierarchically supervised latent dirichlet allocation," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 2609–2617.
- [12] P. Chen, A. Barrera, and C. Rhodes, "Semantic analysis of free text and its application on automatically assigning ICD-9-CM codes to patient records," in *Proc. 9th IEEE Int. Conf. Cognitive Informat.*, 2010, pp. 68–74.
- [13] L. Pereira, R. Rijo, C. Silva, and M. Agostinho, "ICD9-based text mining approach to children epilepsy classification," *Procedia Technol.*, vol. 9, pp. 1351–1360, 2013.
- [14] E. Pons, L. M. Braun, M. M. Hunink, and J. A. Kors, "Natural language processing in radiology: A systematic review," *Radiol.*, vol. 279, no. 2, pp. 329–343, 2016.
- [15] Y. Chen, H. Lu, and L. Li, "Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity," *PLoS One*, vol. 12, no. 3, 2017, Art. no. e0173410.
- [16] I. Goldstein, A. Arzumtshyan, and Ö. Uzuner, "Three approaches to automatic assignment of ICD-9-CM codes to radiology reports," in *Proc. Amer. Med. Informat. Assoc. Annu. Symp.*, 2007, Art. no. 279.
- [17] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [19] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [21] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3036–3044.
- [22] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.
- [23] A. E. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 13, 2016, Art. no. 160035.
- [24] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 2014 Conf. Empirical Methods Natural Language Process.*, 2014, pp. 1746–1751.
- [25] M. Zeng, M. Li, Z. Fei, Y. Yu, Y. Pan, and J. Wang, "Automatic ICD-9 coding via deep transfer learning," *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.04.081.
- [26] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2267–2273.
- [27] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol.*, 2016, pp. 1480–1489.
- [28] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol.*, 2013, pp. 746–751.
- [29] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
- [30] S. Peng, R. You, H. Wang, C. Zhai, H. Mamitsuka, and S. Zhu, "DeepMesh: Deep semantic representation for improving large-scale mesh indexing," *Bioinf.*, vol. 32, no. 12, pp. i70–i79, 2016.
- [31] J. H. Lau and T. Baldwin, "An empirical evaluation of doc2vec with practical insights into document embedding generation," in *Proc. 1st Workshop Representation Learning NLP*, 2016, pp. 78–86.
- [32] N. Maslova and V. Potapov, "Neural network doc2vec in automated sentiment analysis for short informal texts," in *Proc. Int. Conf. Speech Comput.*, 2017, pp. 546–554.
- [33] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 806–813.
- [34] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, and J. Wang, "Applications of deep learning to MRI images: A survey," *Big Data Mining Analytics*, vol. 1, no. 1, pp. 1–18, 2018.
- [35] E. Grefenstette, P. Blunsom, N. de Freitas, and K. M. Hermann, "A deep architecture for semantic parsing," in *Proc. ACL 2014 Workshop Semantic Parsing*, 2014, pp. 22–27.
- [36] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 373–374.
- [37] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 160–167.
- [38] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuska, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.
- [39] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [40] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1338–1351, Oct. 2006.
- [41] J. Nam, J. Kim, E. L. Mencia, I. Gurevych, and J. Fürnkranz, "Large-scale multi-label text classification revisiting neural networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2014, pp. 437–452.
- [42] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.
- [43] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data Mining and Knowledge Discovery Handbook*. Berlin, Germany: Springer, 2009, pp. 667–685.
- [44] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 7, pp. 1079–1089, Jul. 2011.
- [45] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Devin, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Operating Syst. Design Implementation*, 2016, pp. 265–283.
- [46] A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutopoulos, "Evaluation measures for hierarchical classification: A unified view and novel approaches," *Data Mining Knowl. Discovery*, vol. 29, no. 3, pp. 820–865, 2015.
- [47] R. Kavuluru, A. Rios, and Y. Lu, "An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records," *Artif. Intell. Med.*, vol. 65, no. 2, pp. 155–166, 2015.
- [48] G. Tsatsaronis, G. Balikas, P. Malakasiotis, I. Partalas, M. Zschunke, M. R. Alvers, D. Weissenborn, A. Krithara, S. Petridis, D. Polychronopoulos, et al., "An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition," *BMC Bioinf.*, vol. 16, no. 1, 2015, Art. no. 138.
- [49] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in *Proc. 50th Annu. Meet. Assoc. Comput. Linguistics: Short Papers-Volume 2*, 2012, pp. 90–94.
- [50] H. Xu, M. Dong, D. Zhu, A. Kotov, A. I. Carcone, and S. Naar-King, "Text classification with topic-based word embedding and convolutional neural networks," in *Proc. 7th ACM Int. Conf. Bioinf. Comput. Biol. Health Informat.*, 2016, pp. 88–97.
- [51] M. Li, X. Wu, Y. Pan, and J. Wang, "hF-measure: A new measurement for evaluating clusters in protein–protein interaction networks," *Proteomics*, vol. 13, no. 2, pp. 291–300, 2013.
- [52] Y. Huang, W. Wang, L. Wang, and T. Tan, "Multi-task deep neural network for multi-label learning," in *Proc. 20th IEEE Int. Conf. Image Process.*, 2013, pp. 2897–2900.
- [53] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2873–2879.
- [54] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. 2014 Conf. Empirical Methods Natural Language Process.*, 2014, pp. 1724–1734.

- [55] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," in *Proc. 2015 Conf. Empirical Methods Natural Language Process.*, 2015, pp. 379–389.
- [56] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "Gram: Graph-based attention model for healthcare representation learning," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 787–795.
- [57] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T.-S. Chua, "Disease inference from health-related questions via sparse deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2107–2119, Aug. 2015.
- [58] K. Liu, S. Peng, J. Wu, C. Zhai, H. Mamitsuka, and S. Zhu, "Meshlabeler: Improving the accuracy of large-scale mesh indexing by integrating diverse evidence," *Bioinf.*, vol. 31, no. 12, pp. i339–i347, 2015.



Min Li received the PhD degree in computer science from Central South University, China, in 2008. She is the awardee of the NSFC Excellent Young Scholars Program of China, in 2016. She is currently a professor with the School of Information Science and Engineering, Central South University, Changsha, Hunan, P.R. China. Her main research interests include bioinformatics and systems biology.



Zhihui Fei received the BSc degrees in Hubei Normal University, China, in 2015. He is currently working toward the postgraduate degree in bioinformatics with Central South University. His currently research interests include bioinformatics, medical data mining, and deep learning.



Min Zeng received the BS degree from Lanzhou University, in 2013, and the MS degree from Central South University, in 2016. He is currently working toward the PhD degree in the School of Information Science and Engineering, Central South University, China. His research interests include bioinformatics, machine learning, and deep learning.



FangXiang Wu (M'06-SM'11) received the BSc and MSc degrees in applied mathematics from the Dalian University of Technology, Dalian, China, in 1990 and 1993, respectively, and the first PhD degree in control theory and its applications from Northwestern Polytechnical University, Xi'an, China, in 1998, and the second PhD degree in biomedical engineering from the University of Saskatchewan (U of S), Saskatoon, Canada, in 2004. During 2004–2005, he worked as a post-doctoral fellow in the Laval University Medical

Research Center (CHUL), Quebec City, Canada. He is currently a professor of the Division of Biomedical Engineering and the Department of Mechanical Engineering with the University of Saskatchewan. His current research interests include computational and systems biology, genomic and proteomic data analysis, biological system identification and parameter estimation, and applications of control theory to biological systems. He has published more than 260 technical papers in refereed journals and conference proceedings. He is serving as the editorial board member of three international journals, the guest editor of several international journals, and as the program committee chair or member of several international conferences. He has also reviewed papers for many international journals. He is a senior member of the IEEE.



Yaohang Li received the MS and PhD degrees in computer science from Florida State University, Tallahassee, Florida, in 2000 and 2003, respectively. He is an associate professor with the Department of Computer Science, Old Dominion University, Norfolk, Virginia. His research interests include computational biology, Monte Carlo methods, and scientific computing. After graduation, he worked at the Oak Ridge National Laboratory as a research associate for a short period. Before joining ODU, he was an associate professor with the Computer Science Department, North Carolina A&T State University.



Yi Pan received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, in 1991. He is a regents' professor of computer science and an Interim associate dean and chair of biology with the Georgia State University. He joined Georgia State University, in 2000 and was promoted to full professor in 2004, named a Distinguished University professor in 2013 and designated a regents' professor (the

highest recognition given to a faculty member with the University System of Georgia) in 2015. He served as the chair of the Computer Science Department from 2005–2013. He is also a visiting Changjiang Chair professor with Central South University, China. His profile has been featured as a distinguished alumnus in both the Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. His research interests include parallel and cloud computing, wireless networks, and bioinformatics. He has published more than 330 papers including more than 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 6,500 times. He has served as an editor-in-chief or editorial board member for 15 journals including seven IEEE Transactions. He is the recipient of many awards including the IEEE Transactions Best Paper Award, four other international conference or journal Best Paper Awards, four IBM Faculty Awards, two JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at more than 50 international conferences around the world.



Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the vice dean and a professor in the School of Information Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network.

He has published more than 150 papers in various International journals and refereed conferences. He is a senior member of the IEEE.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.