World Academy of Science, Engineering and Technology
International Journal of Bioengineering and Life Sciences
Vol:15, No:9, 2021

Incorporating Lexical-Semantic Knowledge into Convolutional Neural Network Framework for Pediatric Disease Diagnosis

Xiaocong Liu, Huazhen Wang, Ting He, Xiaozheng Li, Weihan Zhang, Jian Chen

***Abstract*—**The utilization of electronic medical record (EMR) data to establish the disease diagnosis model has become an important research content of biomedical informatics. Deep learning can automatically extract features from the massive data, which brings about breakthroughs in the study of EMR data. The challenge is that deep learning lacks semantic knowledge, which leads to impracticability in medical science. This research proposes a method of incorporating lexical-semantic knowledge from abundant entities into a convolutional neural network (CNN) framework for pediatric disease diagnosis. Firstly, medical terms are vectorized into Lexical Semantic Vectors (LSV), which are concatenated with the embedded word vectors of word2vec to enrich the feature representation. Secondly, the semantic distribution of medical terms serves as Semantic Decision Guide (SDG) for the optimization of deep learning models. The study evaluates the performance of LSV-SDG-CNN model on four kinds of Chinese EMR datasets. Additionally, CNN, LSV-CNN, and SDG-CNN are designed as baseline models for comparison. The experimental results show that LSV-SDG-CNN model outperforms baseline models on four kinds of Chinese EMR datasets. The best configuration of the model yielded an F1 score of 86.20%. The results clearly demonstrate that CNN has been effectively guided and optimized by lexical-semantic knowledge, and LSV-SDG-CNN model improves the disease classification accuracy with a clear margin.

***Keywords*—**Lexical semantics, feature representation, semantic decision, convolutional neural network, electronic medical record.

I. Introduction

WITH the development of medical informatization, the EMR system is widely applied in hospitals. Increasingly, intelligent diagnosis based on EMR data is becoming a hotspot in the field of medical informatics [1], [2]. Traditionally, doctors diagnose through the patient's chief complaint, present medical history, past medical history, relevant examination, and other information, which also constitute EMR. Apparently, EMR data are characterized by semi-structured, unstructured, heterogeneous, and fuzzy semantics. How to use EMR data to establish a disease diagnosis model is an important research content of biomedical informatics [3], [4].

Considerable intelligent diagnostic models based on EMR data have been proposed [5]-[10]. As early as the 1970s, EMR data were used to construct small expert knowledge bases manually to support clinical diagnosis, but due to the inefficiency of manual work, the study of disease diagnosis stagnated [11], [12]. In 2015, IBM Watson was able to diagnose diseases based on knowledge graph and infer appropriate treatments for patients. However, this knowledge graph fails to deal with the wide scope and complexity of medical knowledge. Recently, research on EMR data has made a breakthrough using deep learning methods. Research on University of California, San Francisco (UCSF) and Uchicago Medical systems by Google shows that deep learning is an effective intelligent diagnosis model in that it can automatically extract features from massive EMR data without traditional feature engineering [13]. However, the feature representation learned from the deep learning merely derives from the statistics of a large amount of data and diagnosis lacks professional knowledge guidance. Thus, it is far from reaching the practical level in accuracy [14].

In recent years, several research teams have tried to incorporate professional knowledge into deep learning. Fang et al. [15] presented a knowledge-enhanced ensemble method named Latent Semantic Imputation (LSI) to interpret relations in knowledge graphs as linear translation from one word to another for enhancing word embedding. Xu et al. [16] developed a methodology using symbolic knowledge in deep learning. Specially, they constructed a logical-constraint semantic loss, which captures the symbolic knowledge and adds previously-lost information to neural networks. It made the neural network achieve good performance on semi-supervised multi-class classification. Choi et al. [17] proposed a graph-based attention model that supplements electronic health records with hierarchical information inherent to medical ontologies and performed the model in several prediction tasks. However, the idea of integrating semantic knowledge at the lexical level has not been noticed. Practically, lexical semantics can fully express domain knowledge in highly specialized domains such as medicine. In the scenario of medicine, doctors utilize medical terms to precisely describe the health of patients so that EMR contains abundant entity types such as symptoms and disease. Apparently, medical terms are important semantic resources in medical data mining tasks. It is straightforward to think of making full use of these medical terms to learn a disease diagnosis model with good performance.

In this paper, we present a method to integrate lexical-semantic knowledge from abundant entities into a CNN framework for disease classification. The integration of lexical-semantic knowledge in this paper includes two aspects. Firstly,

Huazhen Wang is with the College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China (e-mail: wanghuazhen@hqu.edu.cn).

medical terms are vectorized to generate LSV. Secondly, the semantic distribution of medical terms serves as SDG to adjust the pattern recognition. Or rather, LSV is concatenated with the embedded word vectors generated by word2vec to enrich the feature representation, and SDG is used for the optimization of the deep learning model. Fig. 1 shows the differences between our model and the traditional deep learning model.

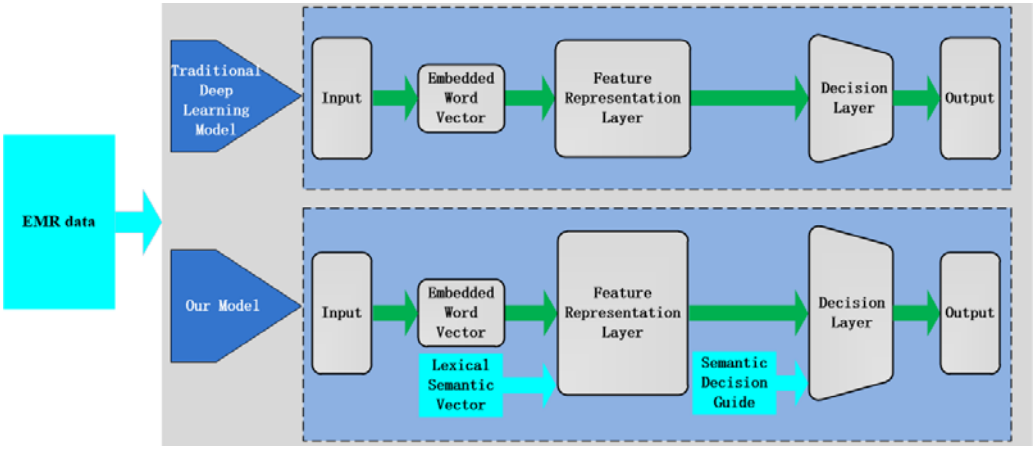


Fig. 1 Differences between our model and the traditional deep learning model. The strategy of LSV and SDG with lexical-semantic knowledge were utilized in the traditional deep learning model to improve model performance

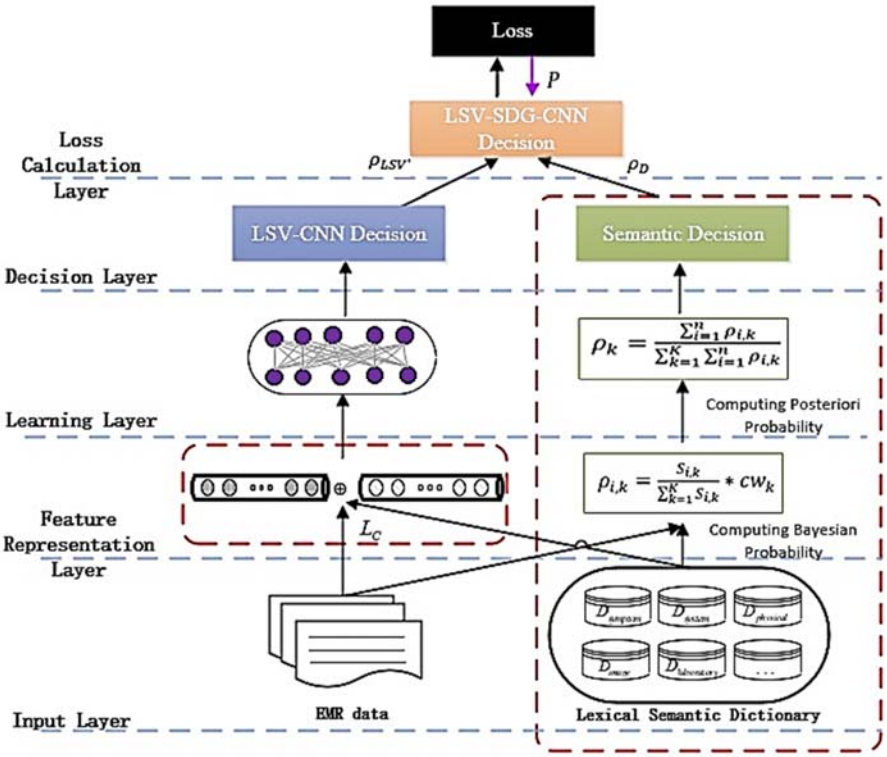


Fig. 2 Algorithm flow of the LSV-SDG-CNN model. In the feature representation layer, the strategy of LSV is utilized to concatenate with the embedded word vectors to enrich the feature representation. In the decision layer, semantic decision based on the strategy of SDG is used to correct the deep learning model

II. Methods

A. The Proposed Model LSV-SDG-CNN

Fig. 2 shows an overview of the proposed model LSV-SDG-CNN. Compared with the traditional deep learning models, the LSV-SDG-CNN model focuses on using lexical semantics to enrich feature representation and guide deep learning models' optimization.

As shown in Fig. 2, in the input layer, both EMR data and the lexical-semantic dictionary are fed to our LSV-SDG-CNN model. In the feature representation layer, each word in EMR data is transformed into an embedded word vector and a LSV by the lexical-semantic dictionary, respectively. These two

kinds of vectors are combined as a mixed vector L_c to propagate into the learning layer of the deep learning model which is specified as CNN in our model. Then, LSV-CNN decision ρ_{LSV} can be obtained in the decision layer. At the same time, a semantic decision ρ_D is calculated by the Bayesian probability $\rho_{i,k}$ of each semantic term in input EMR data according to the lexical-semantic dictionary. Thus, the semantic decision ρ_D is incorporated with LSV-CNN decision ρ_{LSV} to synthesis the output P of our LSV-SDG-CNN model, which is denoted as:

$$P = \frac{\rho_D + \rho_{LSV}}{2} \quad (1)$$

B. The Construction of Lexical-Semantic Dictionary in Pediatrics

Lexical-semantic dictionary, an innovative component of the LSV-SDG-CNN model, can be used for the construction of LSV, semantic decision, and word segmentation. The main principle of constructing a lexical-semantic dictionary is that the dictionary should contain two aspects of semantic terms: words with lexical meaning in domain knowledge and with grammatical meaning [18].

In this paper, we construct a Chinese lexical-semantic dictionary in Pediatrics based on the characteristics of pediatric EMR. A typical EMR includes a set of descriptions, such as the patient's basic information, chief complaint, present history, previous history, family history, and examinations, and then is accompanied by an initial diagnosis. Through the exploration of EMR data, we found that there are some descriptions of symptoms, human organs, examinations, and acronyms commonly used in medicine, which have medical lexical meaning. Besides, there are some numerals, measures, conjunctions, negative words, and degree modifiers, which have grammatical meanings.

According to the characteristic of pediatric EMR data, we extract semantic terms from the following groups: (1) clinical symptom $D_{symptom}$; (2) eight systems of the human body D_{system}; (3) physical examination $D_{physicial}$; (4) image examination D_{image}; (5) laboratory examination $D_{laboratoy}$; (6) acronyms $D_{acronym}$; (7) degree modifiers $D_{modifier}$; (8) negative adverbs D_{not}; (9) conjunctions $D_{conjunctions}$; (10) numeral $D_{numeral}$; and (11) measure $D_{measure}$. Specifically, semantic terms from the first six groups have lexical meaning in domain knowledge, while semantic terms from the last five groups have grammatical meaning. And the detailed information is presented in Table I.

After designing the classes of the lexical-semantic dictionary, we extract pediatric medical terminology with lexical meaning from textbook in China: Pediatrics (seventh edition) [19]. Besides, since those terms with grammatical meaning mentioned above are extremely important to the semantic expression of EMR data, we extract these terms from the modern Chinese grammar library. As a result, a lexical semantic dictionary is built with a size of 4194.

TABLE I
DESCRIPTION OF SEMANTIC TERM

No.	Name of groups	Meaning	Example
1	clinical symptom	abnormal changes in the body after illness	cough
2	eight systems of the human body	eight systems of the human body	digestive system
3	physical examination	detection and measurement of human morphological structure and functional development level	the body's temperature
4	image examination	examination in radiology department or imaging department	computed tomography
5	laboratory examination	physical or chemical examination in the laboratory to determine the characteristics of the substance being examined	blood routine examination
6	acronyms	words that are solidified into a freely usable linguistic unit	red blood cell
7	degree modifiers	degree modifiers for diseases, symptoms, etc.	cough in the morning
8	negative words	adverbs for negating later words	not
9	conjunctions	words used to connect words, phrases or sentences	and
10	numeral	words used to denote quantity	one
11	measure	words used to denote units of quantity	degree

C. Feature Representation Based on Word2vec and LSV

The EMR data need to be converted into data that the computer can recognize, that is, each word needs to be converted into a corresponding vector. At present, there are two mainstream methods for word vector representation, namely one-hot and word2vec. One-hot represents each word as an n-dimensional vector, where n is the vocabulary size. The attribute value corresponding to the word is 1, and the other attribute values are 0. However, it will often cause the curse of dimensionality. Word2vec focuses on finding the mapping relationship between words and vectors by deep learning methods so that the words with similar meanings are clustered in the vector space [20]. Word2vec is powerful, but it does not utilize readily available knowledge or guidance by subject matter experts.

In this paper, the strategy of LSV is proposed to convert semantic knowledge into semantic vectors to enrich feature representation. The concrete scheme is introduced in Fig. 3. Each word in each EMR is expressed as a mixed vector. First of all, an embedded word vector is generated by word2vec, which is notated $L_w = (x_1, x_2, \cdots, x_m)$ where m is the dimension of vector embedding. Additionally, a semantic vector derived from the lexical-semantic groups is constructed, which is notated as:

$$L_s = \left(x_{symptom}, x_{system}, x_{physicial}, x_{image}, x_{laboratory}, x_{acronym}, x_{modifier}, x_{not}, x_{conjunction}, x_{numeral}, x_{measure}\right)$$

each element of which comes from the constructed lexical-semantic dictionary. The semantic vector represents each word as an 11-dimensional vector. When a word belongs to a group, the corresponding attribute value of the group is 1, and the attribute value of other groups is 0. Thus, the embedded word vector L_w and the semantic vector L_s are combined to form a new vector denoted as:

$$L_c = [L_w : L_s] \quad (2)$$

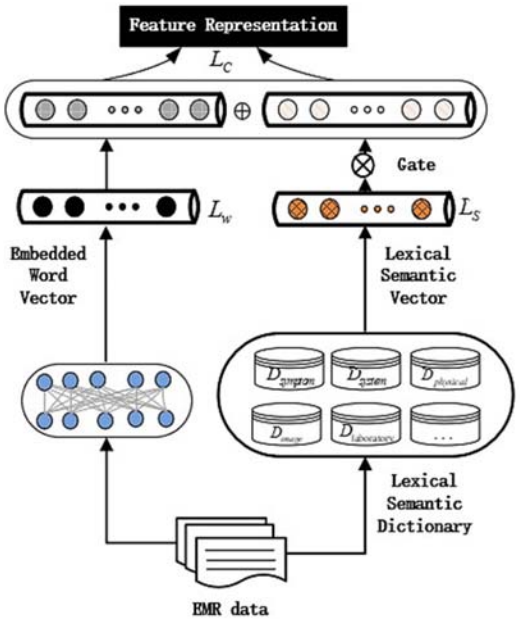


Fig. 3 Feature representation based on word2vec and LSV. Each word in each EMR is expressed as a mixed vector which consists of an embedded word vector generated by word2vec and a semantic vector derived from the lexical-semantic groups

Since words belonging to different groups have different importance for different tasks, the pre-trained vector L_c is finetuned for each task. Overall, L_c is constructed as a feature representation based on statistical learning and lexical features simultaneously.

D. Optimization of CNN Model Based on SDG

In this paper, a strategy of SDG is proposed for the optimization of deep learning models with semantic decisions. The core process of SDG can be described as follows: (1) The Bayesian probability $\rho_{i,k}$ of each term in the lexical-semantic dictionary is calculated using (3); (2) The posterior probability ρ_k that the training sample belongs to the k-th class is computed by (4), which is designated as semantic decision ρ_D for the input data; (3) ρ_D serves as SDG and is incorporated with the CNN decision ρ_{CNN} to construct the loss function, which is used to iteratively optimize the model. The Bayesian probability $\rho_{i,k}$ that the i-th semantic term in a training sample belongs to the k-th class is defined as:

$$\rho_{i,k} = \frac{S_{i,k}}{\sum_{k=1}^{K} S_{i,k}} * cw_k \quad (3)$$

where $S_{i,k}$ denotes the number of occurrences of the i-th semantic term in class k samples, K is the number of disease categories, and cw_k means category weights of class k samples. According to $\rho_{i,k}$, the posterior probability ρ_k that the training sample belongs to the k-th class is defined as:

$$\rho_{i,k} = \frac{\sum_{i=1}^{n} \rho_{i,k}}{\sum_{k=1}^{k=K} \sum_{i=1}^{n} \rho_{i,k}} \quad (4)$$

where n is the number of semantic terms.

The semantic decision ρ_D and SDG decision P_{SDG} are defined by the following form:

$$\rho_D = (\rho_1, \rho_2, \cdots, \rho_K) \quad (5)$$

$$P_{SDG} = \frac{\rho_D + \rho_{CNN}}{2} \quad (6)$$

Specifically, Fig. 4 takes the English sentence "Parts of the story are ok, but the acting was awful" as an example to illustrate the algorithm principle of SDG. First, according to the lexical-semantic dictionary, semantic terms "ok", "but" and "awful" are extracted from this sentence. Based on Bayesian theory, the posterior probabilities of these three semantic terms are then calculated. Furthermore, according to the categories "positive" and "negative", semantic decision ρ_D for this sentence is calculated as 0.4 and 0.6, which is utilized to correct CNN model. As shown in Fig. 4, the strategy of SDG can utilize semantic knowledge to correct the CNN decision ρ_{CNN} and improve the accuracy of diagnostic models.

III. Experiments and Results

A. Datasets

There are four kinds of Chinese EMR from the pediatric department which have been collected. We denote them as 7-classification, 8-classification, 32-classification, and 63-classification. The details of the experimental datasets can be observed in Table II. For each dataset, data were randomly divided into train, validation, and test set in 8:1:1 ratio for five times, yielding five trained models, and we report the average performance.

TABLE II
Distribution of the Datasets

Application	Name of diseases	#Samples
Seven-classification	allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma, (no critical), diarrhea, cough variant asthma	49,333
Eight-classification	acute upper respiratory tract infection, allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma (no critical), diarrhea, cough variant asthma	93,428
32-classification	See appendix	133,861
63-classification	See appendix	145,712

B. Experimental Setup

"Jieba" Chinese text segmentation system with a precise pattern has been adopted for EMR segmentation. Additionally, the lexical-semantic dictionary was applied in the process of word segmentation to make unregistered medical words be correctly segmented.

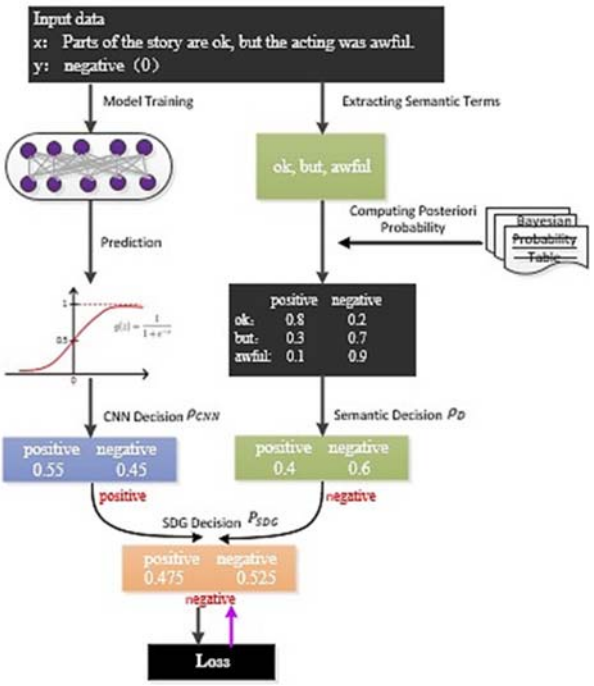


Fig. 4 Algorithm principle of SDG. An example to illustrate the algorithm principle of SDG, that is, how to use semantic decision based on the lexical-semantic dictionary to guide CNN's optimization

After word segmentation, the EMR data need to be further transformed into word vectors that the computer can recognize and process. According to our LSV strategy, two kinds of vectors were combined to form the input of the basic CNN model: one was the embedded word vector by word2vec and the other was the LSV generated based on the lexical-semantic dictionary. Continuous Bag of Words (CBOW) method in word2vec was applied for feature representation. Different dimensions of embedded word vectors should be used for classification problems with different scales [21]. Therefore, the length of 50, 80 and 100 has been explored to be the dimension of $L_w = (x_1, x_2, \cdots, x_m)$ for different sizes EMR datasets. In the 7-classification application, words were embedded in the 50-dimensional vector space. Taking cough and fever as examples, they were converted into vectors, expressed as $[-3.982, -0.670, -1.754, ..., 3.048]_{50}$ and $[-4.487, -5.976, -5.417, ..., 1.216]_{50}$.

To obtain the best schema of the CNN model, we used a grid search method to find an optimized set up of parameters, including convolution kernel size, dropout rate, activation function, and mini-batch size, etc. The optimal parameter setup is shown in Table III.

Admittedly, each of two strategies, LSV or SDG can be incorporated with CNN, which forms LSV-CNN model and SDG-CNN model, respectively. Thus, this paper takes LSV-CNN model, SDG-CNN model, and CNN as baselines. We established CNN, LSV-CNN, SDG-CNN and LSV-SDG-CNN on four kinds of EMR datasets for model performance comparison. Precision, accuracy and F1 score are used to evaluate the performances of these models. And we use cross validation to reduce the likelihood of overfitting.

TABLE III
PARAMETERS OF CNN MODEL

Parameter name	Parameter values
number of layers	1
convolution kernels size	7
number of channels	128
dropout rate	0.5
activation function	Relu
mini-batch size	64
optimizer	AdaMax
loss function	categorical cross entropy

C. Results

Tables IV-VII report the performances of CNN, LSV-CNN, SDG-CNN, and LSV-SDG-CNN on four kinds of datasets. And the loss during the training process in the 7-classification dataset are shown in Fig. 5.

TABLE IV
MODEL PERFORMANCE ON 7-CLASSIFICATION DATASET

Model	precision(%)	accuracy(%)	F1-score(%)
CNN	83.94	83.72	83.78
LSV-CNN	84.23	84.09	84.13
SDG-CNN	85.92	85.74	85.78
LSV-SDG-CNN	86.40**	86.15**	86.20**

Significantly outperforms CNN at the: ** 0.01 and * 0.05 level, ANOVA.

TABLE V
MODEL PERFORMANCE ON 8-CLASSIFICATION DATASET

Model	precision(%)	accuracy(%)	F1-score(%)
CNN	82.35	82.55	82.27
LSV-CNN	82.63	82.83	82.60
SDG-CNN	83.82	83.95	83.72
LSV-SDG-CNN	84.14**	84.27**	84.06**

Significantly outperforms CNN at the: ** 0.01 and * 0.05 level, ANOVA.

TABLE VI
MODEL PERFORMANCE ON 32-CLASSIFICATION DATASET

Model	precision(%)	accuracy(%)	F1-score(%)
CNN	73.09	73.54	72.50
LSV-CNN	73.52	73.92	72.98
SDG-CNN	74.46	74.76	73.93
LSV-SDG-CNN	74.72**	75.04**	74.24**

Significantly outperforms CNN at the: ** 0.01 and * 0.05 level, ANOVA.

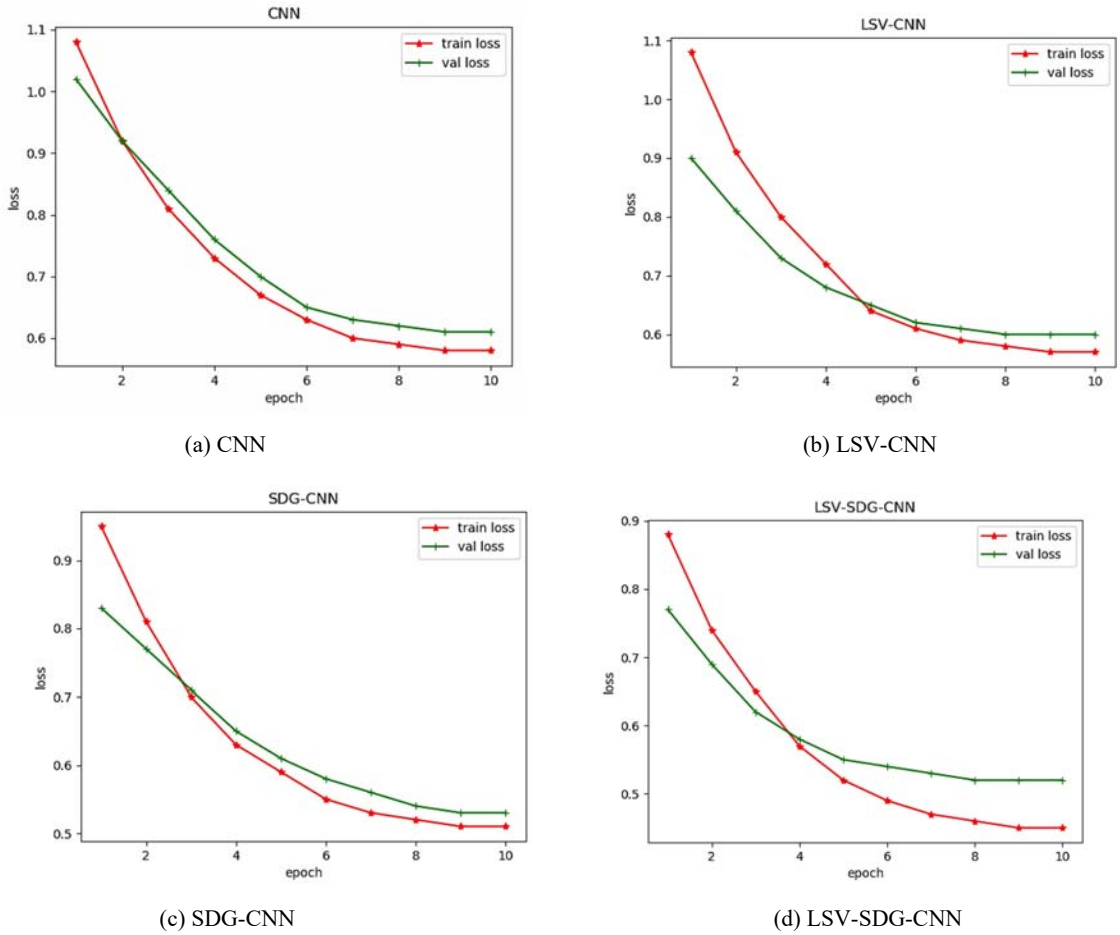
TABLE VII
MODEL PERFORMANCE ON 64-CLASSIFICATION DATASET

Model	precision(%)	accuracy(%)	F1-score(%)
CNN	70.59	71.20	69.61
LSV-CNN	71.28	71.86	70.56
SDG-CNN	72.47	72.71	71.59
LSV-SDG-CNN	72.75**	73.03**	71.89**

Significantly outperforms CNN at the: ** 0.01 and * 0.05 level, ANOVA.

To investigate the effect of LSV and SDG strategies, we calculated the statistical significance of verification accuracy on different models using analysis of variance (ANOVA) of factorial design. The experimental results indicate that LSV and SDG strategies are feasible and effective to improve the accuracy of traditional deep learning, with statistical significance.

It can be seen from Tables IV-VII that, (1) Both LSV-CNN and SDG-CNN outperform the basic CNN model, and LSV-SDG-CNN models outperform other models. Therefore, we can draw the conclusion that the integration of a large amount of lexical-semantic knowledge into the deep learning model optimizes the deep learning model and the algorithm proposed in this paper can be applied to other deep learning models. (2) Comparing LSV-CNN models with SDG-CNN models, we can find that SDG-CNN models outperform LSV-CNN models by 1-2% F1 score. It can be concluded that, compared with the strategy of LSV, the strategy of SDG has a greater contribution to the optimization of the CNN model in this experiment. (3) With respect to different datasets, these four kinds of models on the seven-classification dataset outperform models on the other datasets by 1-14% accuracy, precision and F1 score, which may be attributed to more balanced data and fewer disease types. (4) The LSV-SDG-CNN model performs best on a seven-classification dataset, with a precision up to 86.40%, an accuracy up to 86.15%, and an F1 score up to 86.20%.



(a) CNN (b) LSV-CNN

(c) SDG-CNN (d) LSV-SDG-CNN

Fig. 5 Loss of validation on seven-classification dataset

IV. Conclusion

In this paper, we demonstrated the feasibility of performing Pediatric disease diagnosis integrating lexical-semantic knowledge from entities into the deep learning model. Numbers of prior studies have been conducted on disease diagnosis with the aid of entities or medical terms [22], [23]. However, all these studies just focused on extracting entities from raw medical texts and splicing the entities vectors with medical text vectors. Comparing with previous approaches, we illustrated an implementation that utilizes lexical semantics fully and finely to guide deep learning. We provide a paradigm for the construction of medical lexical semantic dictionary with two aspects and 11 groups. Based on the lexical semantic dictionary, medical terms in EMR data are vectorized to generate LSV and then concatenated with EMR vectors to enrich feature representation. Additionally, the semantic distribution of medical terms serves as SDG to adjust the pattern recognition. We saw improvements when including LSV as features and optimizing CNN with SDG. Remarkably, we were able to see further improvement when employing both LSV and SDG strategies. LSV-SDG-CNN models introduced more predictive power to the configurations. To the best of our knowledge, this is the first study that incorporates lexical-semantic knowledge into CNN model for pediatric disease diagnosis.

There are still some limitations to our study. Although LSV enriches the feature representation by using the semantics of entities and entity analogies, the results in prediction tasks are not outstanding enough. It ignores the semantic relationship

between entities. More investigations are also called for to validate the utility of our model.

APPENDIX

TABLE VIII
THE DISTRIBUTION OF PEDIATRIC EMR DATASETS

#diseases	Name of diseases	#samples
7	allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma (no critical), diarrhea, cough variant asthma	49,333
8	acute upper respiratory tract infection, allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma (no critical), diarrhea, cough variant asthma	93,428
32	acute upper respiratory tract infection, allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma(no critical), diarrhea, cough variant asthma, acute asthmatic bronchitis, abdominal pain, enterovirus infection, fever, acute nasopharyngitis, cough, herpangina, acute tonsillitis, health examination, infantile enteritis, growth hormone deficiency, acute suppurative tonsillitis, acute sinusitis, gastroenteritis, acute gastroenteritis, urinary tract infection, asthmatic bronchitis, epilepsy, pneumonia, constipation, indigestion, acute lower respiratory tract infection, mycoplasma infection	133,861
63	acute upper respiratory tract infection, allergic rhinitis, bronchitis, acute bronchitis, respiratory disease, bronchial asthma(no critical), diarrhea, cough variant asthma, acute asthmatic bronchitis, abdominal pain, enterovirus infection, fever, acute nasopharyngitis, cough, herpangina, acute tonsillitis, health examination, infantile enteritis, growth hormone deficiency, acute suppurative tonsillitis, acute sinusitis, gastroenteritis, acute gastroenteritis, urinary tract infection, asthmatic bronchitis, epilepsy, pneumonia, constipation, indigestion, acute lower respiratory tract infection, mycoplasma infection, nausea and vomiting, idiopathic thrombocytopenia purpura(ITP), acute lymphoblastic leukemia, infantile diarrhea, gastritis, allergic purpura, gastrointestinal dysfunction, neonatal hyperbilirubinemia, hematuria, tic disorders, digestive system disease, upper respiratory tract hypersensitivity reaction, enuresis, neonatal jaundice, enteritis, mucocutaneous lymph node syndrome, renal allergic purpura, ulcerative stomatitis, routine examination of children's health, herpangina, chronic sinusitis, upper respiratory disease, stomatitis, right inguinal hernia, hyperthyroidism, anemia, helicobacter pylori infection, acute pharyngitis, left inguinal hernia, headache, acute laryngitis	145,712

ACKNOWLEDGMENT

Research works in this paper are supported by the National Key Technology R&D Program of China (No.2018YFB1402500), the Social Science Planning Foundation of Fujian Province (FJ2020B0033), the Scientific Research Funds of Huaqiao University (16BS304) and Huaqiao University's Academic Project Supported by the Fundamental Research Funds for the Central Universities (TZYB-202005).

REFERENCES

[1] A. Boonstra and M. Broekhuis, "Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions," BMC Health Services Research, vol. 10, no. 1, pp. 231–241, 2010.
[2] W. Mackinnon and M. Wasserman, "Integrated electronic medical record systems: Critical success factors for implementation," in Proc. of the Hawaii International Conference on System Sciences, 2009, pp. 1–10.
[3] Y. Li, B. Qian, X. Zhang, et al., "Graph neural network-based diagnosis prediction," Big Data, vol. 8, no. 5, pp. 379-390, 2020
[4] Y. Li, R Shishir, Solares, et al., "BEHRT: Transformer for electronic health records," Entific Reports, vol. 10, no. 1, pp. 7155-7167, 2020.
[5] J. Gao, X. Wang, Y. Wang, et al., "CAMP: Co-attention memory networks for diagnosis prediction in healthcare," in Proc. of the 19th IEEE International Conference on Data Mining, New York, 2019, pp. 1036–1041.
[6] X. S. Hang, F. Tang, H. H. Dodge, et al., "MetaPred: Meta-learning for clinical risk prediction with limited patient electronic health records," in Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2019, pp. 2487–2495.
[7] W. Wang, H. Xu, Z. Gan, et al., "Graph-driven generative models for heterogeneous multi-task learning" in Proc. of the 35th AAAI Conference on Artificial Intelligence. Menlo Park, 2020, pp. 979–988.
[8] J. Jiang, H. Wang, J. Xie, et al., "Medical knowledge embedding based on recursive neural network for multi-disease diagnosis," Artificial Intelligence in Medicine, vol. 103, no. 1, pp. 101772–101787, 2020.
[9] L. Wang, H. Wang, Y. Song, et al., "MCPL-Based FT-LSTM: medical representation learning-based clinical prediction model for time series events," IEEE Access, vol. 7, no. 1, pp. 70253–70264, 2019.
[10] H. Liang, B. Y. Tsui, H. Ni, et al., "Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence," Nature medicine, vol. 25, no. 3, pp. 433–443, 2019.
[11] A. Bordes, J. Weston, R. Collobert, et al., "Learning structured embeddings of knowledge bases," in Proc. of the 25th AAAI Conference on Artificial Intelligence, Menlo Park, vol. 25, no. 1, 2011.
[12] A. Bordes, N. Usunier, A. Garcia-Duran, et al., "Translating embeddings for modeling multi-relational data," in Proc. of the neural information processing systems, Cambridge, 2013, pp. 2787–2795.
[13] A. Rajkomar, E. Oren, C. Kai, et al., "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, vol. 1, no. 1, pp. 18, 2018.
[14] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, et al., "Opportunities and obstacles for deep learning in biology and medicine," Journal of the Royal Society Interface, vol. 15, no. 141, pp. 20170387, 2018.
[15] L. Fang, Y. Luo, K. Feng, et al., "Knowledge-enhanced ensemble learning for word embeddings," in Proc. of the World Wide Web Conference, New York, 2019, pp. 427–437.
[16] J. Xu, Z. Zhang, T. Friedman, et al., "A semantic loss function for deep learning with symbolic knowledge," in Proc. of the 35th International Conference on Machine Learning, Cambridge, 2018, pp. 5502–5511.
[17] E. Choi, M. T. Bahadori, S. Le, et al., "GRAM: Graph-based attention model for healthcare representation learning," in Proc of the 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2017, pp. 787–795.
[18] N. Rojas, L. B. Crane, E. Yeager, et al., "Introduction to Linguistics," Modern Language Journal, vol. 66, no. 4, pp. 445,1998.
[19] X. Shen, "Pediatrics," 7th ed. Beijing: People's Health Publishing House, 2013.
[20] V. Jayawardana, D. Lakmal, N. D. Silva, et al., "Deriving a representative vector for ontology classes with instance word vector embeddings," in Proc. of the Seventh International Conference on Innovative Computing Technology, 2017, pp. 79-84.
[21] T. Mikolov, K. Chen, G. Corrado, et al., "Efficient estimation of word representations in vector space," in Proc of 7th International Conference on Learning Representations, Stroudsburg, 2013, pp. 1–12.
[22] N. Liu, P. Lu, W. Zhang, et al., "Knowledge-aware deep dual networks for text-based mortality prediction," in Proc. of 35th International Conference on Data Engineering, 2019, pp. 1406–1417.
[23] H. Wang, Y. Li, S. A. Khan, et al., "Prediction of breast cancer distant recurrence using natural language processing and knowledge-guided convolutional neural network," Artificial Intelligence in Medicine, vol. 110, p. 101977, 2020.