

Representation Learning for Electronic Health Records: A Survey

Peiying Chen

School of Physics and Technology, Wuhan University, Wuhan, Hubei, China

pyChen@whu.edu.cn

Abstract. With the wide application of Electronic Health Record (EHR) in hospitals in past few decades, researches that employ artificial intelligence (AI) and machine learning methods based on EHR data have been explosive. With such EHR data, one can engage in many tasks such as risk prediction, treatment recommendation, information imputation, etc. The performance of classification or prediction highly depends on the quality of data representation, i.e., representing original records into numerical vectors to facilitate further learning. However, there is little research that focuses on the representation learning techniques for EHR data at present, which makes it hard to understanding the development trend of EHR learning in a global map. In this paper, we bridge this gap by systematically investigating the related research efforts that apply the representation learning on EHR data. We analyze and conclude the techniques used in the typical representation learning approaches as well as the limitations and advantages of them. The survey would provide a comprehensive reference for further analysis and application in EHR research.

1. Introduction

With the recent increasing requirements of medical care, the application of medical data is treated as an effective approach to mitigate the widening gap between health care requirements and supplies [1]. The application of Electronic Health Record (EHR) has brought health care to a new era. Medical workers benefit from the convenience of EHR by the long-lasting record interval and data sharing between medical institutions, e.g. financial institutions leverage the secondary usage of EHR to evaluate risk of insurance [2]. More remarkably, some researches target at using such data to develop new medical tools based on data-driven method, which forms a novel trend. In fact, due to the high-dimensionality, temporal dependency, sparsity, irregularity and heterogeneity [3-5], there are still many obstacles in making full of use in medical data, in which looking for effective approaches to represent EHR data is a fundamental issue.

In EHR study, representation learning is a necessary preparation step of any learning-based framework. In the past, researchers usually work with domain experts to label and represent the data manually. The defect of such method is apparent that it is time consuming and the representation is hard to be meaningful and context-aware. Alternatively, learning for representation provides a novel way that automatically derives the representation from the original EHR data. The representation could be either simple or highly abstract, depending on different algorithms. Furthermore, because of the diversity of the learning process, there are opportunities to discover novel representation algorithms. Although representation learning has been applied in a lot of data-related research fields, the study that focuses on the application to EHR representation, especially from a technical perspective, is still insufficient. In



fact, there are still many challenges in deriving effective representation of health data, where current investigation is not enough for supporting such research on the EHR.

In this paper, we attempt to conduct a survey to study the representation learning methods on EHR in a comprehensive way, particularly from technical perspective. We demonstrate and discuss several popular representation learning methods that widely used in EHR research, such as word embedding, machine learning, deep learning and hybrid learning. We analyze how the data characteristics and the representation targets impact the choosing of representation learning methods. We introduce some typical works that apply representation learning methods in EHR study and evaluate their performance. Ultimately, we review the application of representation learning in medical domain, analyze the advantages of each methods and discuss the opportunities, challenges and some potential further works based on present study from our perspective.

2. Representation Learning

The performance of a data-driven algorithm is highly based on the representation of data. Representation learning is method of extract useful information from raw data that can promote the accuracy and efficiency of classifiers or predictors [6]. There are many approaches of Representation Learning, such as Word Embedding, Word2Vec, Deep Learning etc. In this part, we will demonstrate several widely used Representation Learning approaches in detail.

2.1. Word Embedding

Embedding is a term derived from mathematic, the meaning of embedding is to represent one mathematical structure by another form. In NLP (Natural Language Processing), the Word Embedding is the collective name of the methods that map the phrases and words into vectors that composed of real numerical value. There are different approaches of word embedding, such as Simple Embedding (One-hot) and Word2Vec. In this section, we will introduce the widely used embedding methods.

2.1.1. One-hot Encoding. One-hot encoding is one of the simplest methods of Word Embedding. One-hot encoding uses a vector composed by one high bit (1) and several low bits (0) to represent a feature. For example, if we are discussing the effect of a treatment to some certain illnesses, we can assign the effect such as improved, worsen and ineffective as three different states. The three states of the feature correspond to three vectors ([100],[010],[001]). For a paragraph include many features, the final one-hot vector is generated through combine each of single feature vector together.

As a fundamental method of Word Embedding, One-hot coding is comprehensive and concise. However, this method represents the features of words only by discrete number 1 or 0, ignoring the logic of contextual semantics within the word sequences. In addition, when the number of features is large, one-hot coding usually confront the problem of dimension explosion.

2.1.2. Word2Vec. Word2vec is a 3-layer neural networks proposed by Mikolov et al [7]. The input layer is inputted one-hot code and connect to a linear hidden layer, where the one-hot code will generate by a weight matrix. Finally, the final vector is produced by a Softmax transformation. Generally, there are two typical implementations for the Word2vec, the Continuous Bag-of-Words Model (CBOW) and the Skip-gram Model, we will introduce the two in the following parts.

1) Continuous Bag-of-Words (CBOW)

The CBOW is one of the popular embedding models for implementation of the Word2Vec, which is first proposed by Mikolov et al [7]. In the CBOW model, context is represented by multiple words for a given target words.[8] The input layer consists of one-hot encoded input contexts $\{x_1, \dots, x_C\}$, where the window size is C and the vocabulary size is V . The hidden layer is a N -dimensional vector. The final output layer is the output word y that is also encoded by one-hot. The input vector encoded by one-hot is connected to the hidden layer by a $V \times N$ -dimensional weight matrix W ; the hidden layer is connected to the output layer by a weight matrix W' of $N \times V$. The entire CBOW model can be shown in Figure 1.

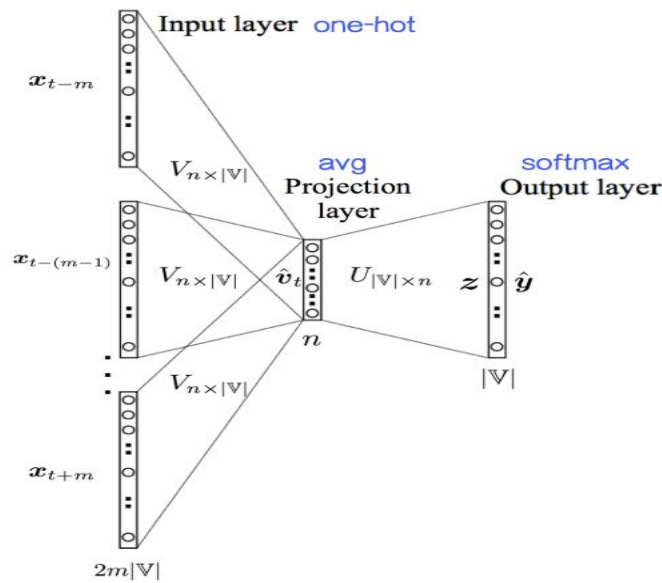


Figure 1. Training Structure of CBOW [9]

2) Skip-gram

Different from the CBOW, Skip-gram uses a given current word to predict the words in context. The input layer of the Skip-gram receives one-hot encoded N -dimensional word 0-1 vectors, which is connected to hidden layer by a $V \times N$ weight matrix W . Then, the hidden layer is connected to the output layer by a $N \times V$ weight matrix W' . Eventually, the word vectors are outputted and represent the feature of the target word. The overall Skip-gram model is shown in Figure 2.

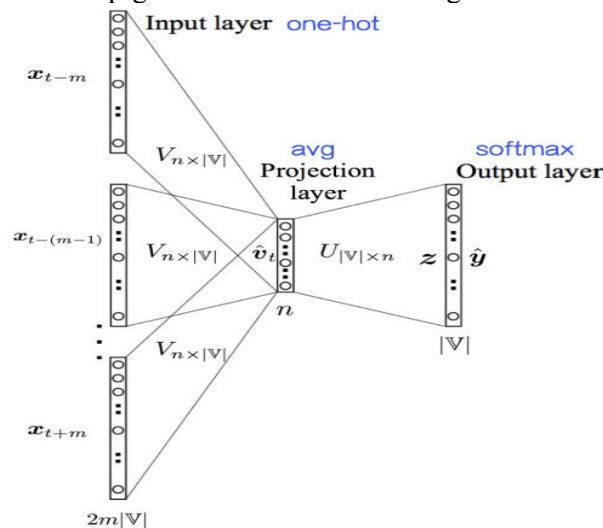


Figure 2. Training Structure of Skip-gram [9]

2.2. Machine Learning/Deep Learning

1) Machine Learning

The concept of Machine Learning was raised by Arthur Samuel in 1959[3]. With the developing of hardware, machine learning gradually becomes one of the most active area in data science. In Electric Health Record (EHR) analysis, the EHRs data are used as data samples that can be utilize by certain learning models to join the model training.

2) Deep Learning

Deep learning was originated from the research of neuroscience, which works as a non-linear model that can transform the input representation into high-level abstract form. The vast majority architectures

of deep learning algorithms are based on the framework of artificial neural network (ANN), which is composed of numerous layers of interconnected nodes. The nodes between the input layer and the output layer is called hidden layer, which are assigned a set of weight factors that trained in the learning model. It is noticeable that there are increasing research works that apply machine learning or deep learning related techniques to represent EHR data. Follows list two typical learning-based representation learning techniques.

2.2.1. Automated Encoder (AE). Autoencoder is a specific type of neural network that can zip the data. The architecture of Autoencoder is an hourglass-shaped neural network. There are the maximum number of nodes in the input layer and the output layer, and minimum number in the middle of hidden layer called the bottleneck. The data samples will be encoded layer-by-layer with the decrease of nodes in each layer. When reaching the bottleneck, the representation is highly abstract and dimension reduced. Then, the representation will be decoded through the layer-by-layer increasing nodes to generate a representation with the same dimension as input one. (See Figure 3)

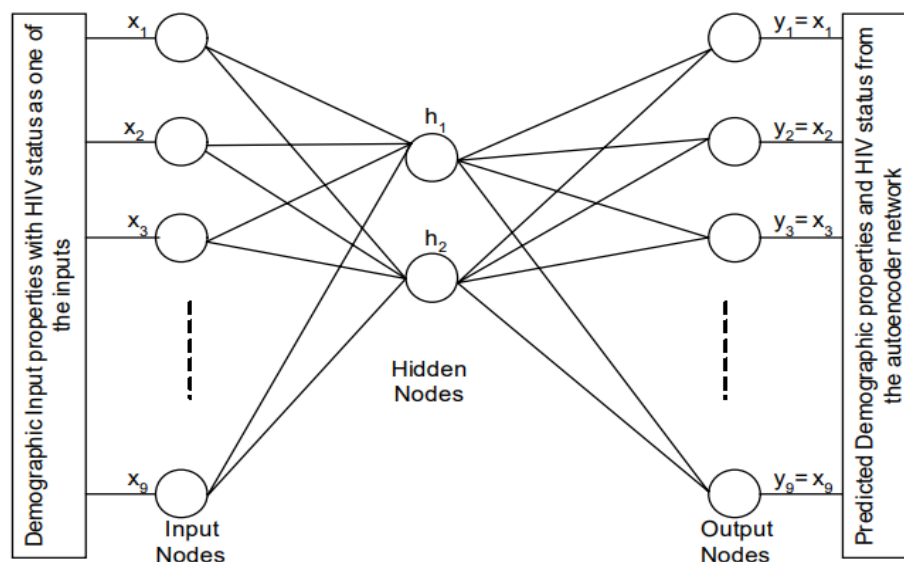


Figure 3. The overall structure of AE [10]

2.2.2. Restricted Boltzmann Machine (RBM). Restricted Boltzmann Machine was firstly proposed by Hinton and Sejnowski in 1986. It is a generative stochastic neural network composed of a visible layer and a hidden layer. The nodes in the same layer are not connected to each other. The nodes of RBM are binary-valued, and connected to a $m \times n$ weight matrix W . RBM is able to deal with some complex problem and reduce the risk of overfitting. It is widely used in dimension reduction, classification and feature learning.

3. Representation Learning for HER

As the information stored in EHR explosively increasing, many applications based on data science have been developed by using the HER data. EHR including data of diagnoses, radiological images, prescription, laboratory result and more. The applications usually utilize the data to implement predict or classify work. [11] However, the prediction and classification algorithms can't directly use the original data from EHR which is imputed in natural language by medical worker. Therefore, extract and process the crucial information from EHR is an important preparatory work. Representation Learning is a model that can leverage a big amount of unlabeled data to learn good intermediate representations. In EHR researching field, Representation Learning is widely used to process the original EHR data. In this section, we will expound the application of representation learning in EHR researching.

3.1. Word Embedding

Word embedding as one of the most used representation techniques, has been widely adopted in EHR learning and analysis. A few works directly use the one-hot encoding as the data representation. For instance, [12] applies two statistical relational learning (SRL) algorithms to the task of predicting primary myocardial infarction. The learning is conducted on the simple one-hot code of category data, such as date, sex, age, etc. The experiment results show that the prediction outcomes produced by single SRL algorithm outperform their propositional analogs. [13] employs one-hot and word vector to model the diagnosis event and predict heart failure using a neural network. They used one-hot encoding to represent each diagnostic event sequence and employ a long short-term memory networks (LSTM) for heart failure prediction with the modeled diagnosis events as the input. Word2vec is widely used to represent data in EHR researching, [14] uses skip-gram model to pre-train the clinic information from two large clinic database with a 10 tokens window size. [15] uses word2vec to train the word embeddings with a empirical 100 dimension and use the representation to expand abbreviation. They finally achieved 82.27% accuracy, which is close to expert human performance. [16] employs word2vec tool using skip-gram model and tf-idf model to convert the clinical record into numerical vector. they find that word2vec perform relatively accurate for SSAE (Stacked Sparse Auto-Encoder) classifier than SRBM (Sparse Restricted Boltzmann Machine), for which tf-idf have a better performance. In [17], researchers propose a framework to extract the information according to EHR data. They use word2vec to create the distributed representation for the concepts of the gender and family relationships that associated with the family medical history. [18] train the word2vec model with continuous bag-of-words (CBOW) method to learn medical feature embedding from HER data. And input the data into a convolutional neural network prediction model. Their model shows noticeable accuracy in heart failure (86.30%) and diabetes (98.44%) prediction. [19] applies word2vec to pretrain the embedding matrix in their model Deepr. For the original input, they employ time stamps between medical events to improve the accuracy of prediction of unplanned readmission after discharge. In their result, Deepr shows superior performance than BoW+LR baseline. [20] use skip-gram embedding to initialized embedding layer values in their models with unlabeled EHR corpus and use recurrent neural network (RNN) to improve exact phase labeling. Their results suggest that their structured prediction model is a good direction for improving the exact phrase extraction for medical entities.

3.2. Machine Learning/ Deep Learning

Some of the researchers employ machine learning/deep learning technique to study the representation of EHR. For example, [20] combine RNN with a pre-training word2vec model to improve the accuracy of phase extraction and obtain a satisfying result. [21] applies an RNN initialized by skip-gram in their model doctor AI. They conclude that their RNN-based model can learn efficient patient representation from a large amount of longitudinal patient record and shows high performance in transfer learning. In their result, their model pre-trained in large dataset improve 10% performance than which directly trained in smaller dataset. In [22], researchers investigated deep neural network (DNN) for Named Entity Recognition (NER) which is applied in word embedding in Chinese clinical text. Their results show that DNN-base word embedding achieved the highest performance at the minimal feature setting. [23] explore an RNN framework initialized by skip-gram word embedding method. The RNN is trained in sentence level and document level. Their results show that their RNN model has achieved a noticeable leap in performance compare to the traditional CRF method. [24] employ a CNN for extracting phenotype for each patient and fuse them with temporal domain, the proposed model was validated on real world EHR data a quantitatively and qualitatively. [25] applies autoencoder approach in an imputation model to solve the missing information in EHR. In their works, the weights and bias of autoencoder is trained only by present features and the missingness is labeled in the process. The trained vector is used as a resource for the further imputation work. Their results show that autoencoder has strong accuracy in imputation work and contributes to disease progression predictor. In [26-27], a stack of denoising autoencoders are applied in their model to learn the representation of patients and achieve the top performing in severe diabetes, schizophrenia and various cancer. [28] applies denoising

autoencoder (Ds) for phenotype extraction in EHR. The Ds is trained by unlabeled clinical data and successfully learn the structure of the high-dimensional EHR data. In the result, their model performs good dimensionality reduction capability that able to promote the discovery of new subtype of disease. Restricted Boltzmann machine (RBM) as a popular feature extracting approach is widely used in EHR word embedding. [29] use RBM to extract feature from EHR data by assign a weight for different type of transcripts measurement such as height, weight, blood pressure and temperature. [30] applies RBM to embed medical object in a low-dimensional vector space. In their novel model, embedding coefficients are restricted to be nonnegative and the learning is trained by clinical structures originated from the disease taxonomy, the procedure hierarchy and the temporal progression of care and disease. Their result shows that the representation derived from their model achieved meaningful clinical feature grouping and able to facilitate short-term risk stratification. The F-scores of their model are 0.21 for moderate-risk and 0.36 for high-risk which is much higher than those obtained by clinicians.

3.3. Hybrid Learning

Hybrid learning is a short of effective learning model that usually applied in representation learning. There are many remarkable works attempt the method and obtain reliable result. [31] proposes a robust hybrid approach based on O-Learning. in [32], researcher annotate the clinical text by an open source annotation tool called Knowtator, and propose an Unstructured Information Management Architecture (UMIA) framework to extract the information from medical entity. In their result, the system level F1-measure reaching a high level for exact matches and precision (>85%) for the medication information such as dosage and frequency, but not as good for durations. [33] proposed a Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System (MKL with ANFIS) based deep learning method for heart disease diagnosis. For the representation learning part, they apply MKL method to separate the feature variable between healthy and heart disease patient. Their result shows that MKL with ANFIS method has the highest specificity of 94% and second highest sensitivity of 95.9% among various existing approach. [34] use Word2Vec and GloVe to pretraining the data and applies an RNN model to learning the representation of the de-identification data in EHR. In their result, recall on their data set discharge summaries of over 90%, which is a considerable achievement. [35] uses CBOW model combine with an unsupervised learning model to derive the representation of EHR. The representation is used to evaluating the similarity between patient. Their result achieves Rand index of 0.9887, which is superior to others that the second-best model performs 0.6769 of Rand index. [36] applies GloVe with a Graph-based attention model to derive the representation of patient to solve the problem of constrained volume of data. The result shows their model perform significant improvement on low-frequency diseases and small datasets with a 10% outperformance than basic RNN on their dataset MMIC-III. [37] applies One-hot code to pre-notate the representation. Then employs a fully connected network layer to reduce feature dimension, and use a CNN to derive the final representation. The representation is used to identify and rank the similarity among patients. Their result shows that their CNN-based frameworks outperform the widely used distance metric learning methods.

4. Discussion

We conclude the typical investigated representation approaches in the EHR learning researches as shown in Table 1. Overall, the Word2Vec representation approaches shows high performance in well-organized EHR data. This can be attributed to its high dependency of the relation between neighboring context. When processing EHR data with much missing information, researchers tend to use other methods to carry out the representation.

The RNN based models perform well in EHR data that contains adequate time stamps and they are not very strict to the organization of data. However, the training speed is relatively low for RNN. In some cases, RNN shows slight outperformance compared with other methods but considerably falls behind in efficiency.

Table 1. Conclusion of typical representation in EHR learning research

Author	Task	Representation Techniques	Reference
Choi E et al.	Transfer learning for patient record	RNN	[21]
Weiss J C, Natarajan S, Peissig P L, et al.	Predicting primary myocardial infarction from electronic health records	SRL	[12]
Jin B, Che C, Liu Z, et al.	predict heart failure	One-hot+LSTM	[13]
J. A. Fries	Temporal information extraction	Word Embedding+RN N	[14]
Liu Y, Ge T, Mathews K S, et al.	Clinical abbreviation expansion	Word2Vec	[15]
Jacobson O, Dalianis H.	Healthcare-associated infections prediction	Skip-gram	[16]
Shah S, Luo X.	Extracting Modifiable Risk Factors	Word2Vec	[17]
Che Z, Cheng Y, Sun Z, et al.	Diseases risk prediction	CBOW	[18]
Nguyen P, Tran T, Wickramasinghe N, et al.	Prediction of unplanned readmission after discharge	Word2Vec+CN N	[19]
Jagannatha A N, Yu H.	Improving the accuracy of phase extraction	Word2Vec+RN N	[20]
Wu Y, Jiang M, Lei J, et al.	Word embedding in Chinese clinical texts	DNN	[22]
Jagannatha A N, Yu H.	Medical event detection	RNN + skip-gram	[23]
Beaulieu-Jones B K, Moore J H.	Missing data imputation	Autoencoder	[25]
Miotto R, Li L, Kidd B A, et al.	Predicting the future disease	Autoencoder	[26][27]
Beaulieu-Jones B K, Greene C S.	Phenotype extraction	Autoencoder	[28]
Gupta P, Sivalingam U, Pölsterl S, et al.	Identifying patients with diabetes	RBM	[29]
Tran T, Nguyen T D, Phung D, et al.	Learning vector representation of medical objects	RBM	[30]
Cheng Y, Wang F, Zhang P, et al.	Extracting phenotype for risk prediction	CNN	[24]
Liu Y, Wang Y, Kosorok M R, et al.	Estimating personalized dynamic treatment regimens	O-Learning	[31]
Meystre S M, Thibault J, Shen S, et al.	Prescription extraction	Knowtator+UM IA	[32]
Manogaran G, Varatharajan R, Priyan M K.	Heart disease diagnosis	MKL + ANFIS	[33]
F. DERNONCOURT, J. Y. Lee, O. Uzuner, and P. Szolovits	Learning the representation of the de-identification data	Word2Vec+Glo Ve+RNN	[34]
Zhu Z, Yin C, Qian B, et al..	Measuring patient similarities	CBOW+ Unsupervised Learning	[35]
Choi E, Bahadori M T, Song L, et al.	Representation learning for constrained dataset	GloVe+ Graph-based attention model	[36]
Suo Q, Ma F, Yuan Y, et al.	Identify and rank the similarity among patients	One-hot+CNN	[37]

The CNN related representation learning methods are quite easy in model training and show high performance in high dimensional data. But for time sequence data, CNN shows lower performance than RNN.

The RBM related representation learning is concise, explicit and extendable. RBM can estimate representation on unseen data by a matrix operator in very short time, which makes the RBM be able to process complex and big scale data. However, the RBM can't effectively extract rich domain relations or prior knowledge [38]. For the presence of diagnosis codes aggregated over time in the medical data, RBM shows its limitation.

The hybrid learning representation can combine the advantages of different approaches. Usually, Word Embedding would be employed as a pretraining model. Machine learning and Deep learning are applied to generate more abstract and precise representation.

5. Conclusion

Although representation learning approaches have been successfully used in many EHR researches, the representation itself is usually treated as a black box. In other words, the interpretability and the real usage of such representation are still lack of focuses. Also, because of the high-dimensionality, temporal dependency, sparsity, irregularity and heterogeneity of EHR data, the representation approaches on the EHR data are extremely diverse. Therefore, it is challenging to seek a comprehensive way to derive such representation approaches. In this paper, we address the challenge by presenting typical representation learning approaches applied in EHR data related researches and analyzing their performance in technical perspective. Additionally, we conclude their advantages and limitations by the comparison between techniques and targets. The work can be a useful reference of further EHR data related researches.

Reference

- [1] Ludwick D A, Doucette J. 2009. *International journal of medical informatics*, Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. **78(1)**: 22-31.
- [2] Fleming N S, Culler S D, McCorkle R, et al. 2011, *Health Affairs*. The financial and nonfinancial costs of implementing electronic health records in primary care practices. **30(3)**: 481-489.
- [3] Hripcsak G, Albers DJ. 2013; *Am Med Inform Assoc*. Next-generation phenotyping of electronic health records. **20**:117–21.
- [4] Jensen PB, Jensen LJ, Brunak S. 2012; *Nat Rev Genet*, Mining electronic health records: towards better research applications and clinical care. **13**:395–405.
- [5] Luo J, Wu M, Gopukumar D, et al. 2016. *Biomed Inform Insights*. Big data application in biomedical research and health care: a literature review.**8**:1–10.
- [6] Bengio Y, Courville A, Vincent P. 2013, *IEEE transactions on pattern analysis and machine intelligence*, Representation learning: A review and new perspectives. **35(8)**: 1798-1828.
- [7] Mikolov T, Chen K, Corrado G, et al. 2013. *arXiv preprint*. Efficient estimation of word representations in vector space. arXiv:1301.3781
- [8] Goldberg Y, Levy O. 2014. *arXiv preprint*. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv:1402.3722
- [9] Rong X. 2014. *arXiv preprint*. word2vec parameter learning explained. arXiv:1411.2738.
- [10] Betechnuoh B L, Marwala T, Tettey T. 2006, *Current Science*, Autoencoder networks for HIV classification. **(00113891)**, 91(11).
- [11] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, 2010, *AMIA Joint Summits on Translational Science proceedings*. "Secondary Use of EHR: Data Quality Issues and Informatics Opportunities.", vol. pp. 1–5.
- [12] Weiss J C, Natarajan S, Peissig P L, et al. 2012, *AI Magazine*, Machine learning for personalized medicine: Predicting primary myocardial infarction from electronic health records. **33(4)**: 33-33.
- [13] Jin B, Che C, Liu Z, et al. 2018, *IEEE Access*, Predicting the risk of heart failure with ehr sequential data modeling. **6**: 9256-9261.

- [14] J. A. Fries, 2016, in *Proc. 10th Int. Workshop Semantic Eval.* “Brundlefly at SemEval-2016 Task 12: Recurrent neural networks vs. joint inference for clinical temporal information extraction,” pp. 1274–1279.
- [15] Liu Y, Ge T, Mathews K S, et al. 2016, *arXiv preprint*, Exploiting task-oriented resources to learn word embeddings for clinical abbreviation expansion. arXiv:1804.04225, 2018.
- [16] Jacobson O, Dalianis H. 2016, *Proceedings of the 15th workshop on biomedical natural language processing*. Applying deep learning on electronic health records in Swedish to predict healthcare-associated infections. **2016**: 191-195.
- [17] Shah S, Luo X. 2018. *IEEE International Conference on Bioinformatics & Bioengineering*. Extracting Modifiable Risk Factors from Narrative Preventive Healthcare Guidelines for EHR Integration.
- [18] Che Z, Cheng Y, Sun Z, et al, 2017 *arXiv preprint*. exploiting convolutional neural network for risk prediction with medical feature embedding. arXiv:1701.07474.
- [19] Nguyen P, Tran T, Wickramasinghe N, et al. 2017. *IEEE journal of biomedical and health informatics*, $\{\text{DeepR}\}$: a convolutional net for medical records. **21(1)**: 22-30.
- [20] Jagannatha A N, Yu H. 2016, *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing. NIH Public Access*, Structured prediction models for RNN based sequence labeling in clinical text **2016**: 856.
- [21] Choi E, Bahadori M T, Schuetz A, et al. 2016. *Machine Learning for Healthcare Conference*. Doctor ai: Predicting clinical events via recurrent neural networks **2016**: 301-318.
- [22] Wu Y, Jiang M, Lei J, et al. 2015, *Studies in health technology and informatics*. Named entity recognition in Chinese clinical text using deep neural network. **216**: 624.
- [23] Jagannatha A N, Yu H. *Proceedings of the conference. 2016, Association for Computational Linguistics. North American Chapter. Meeting. NIH Public Access*, Bidirectional RNN for medical event detection in electronic health records **2016**: 473.
- [24] Cheng Y, Wang F, Zhang P, et al. 2016. *Proceedings of the 2016 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*. Risk prediction with electronic health records: A deep learning approach **2016**: 432-440.
- [25] Beaulieu-Jones B K, Moore J H. Missing data imputation in the electronic health record using deeply learned autoencoders Pacific Symposium on Biocomputing 2017. 2017: 207-218.
- [26] Miotto R, Li L, Kidd B A, et al. 2016, *Scientific reports*, Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. 6: 26094.
- [27] Miotto R, Li L, Dudley J T. 2016. *European Conference on Information Retrieval*. Deep learning to predict patient future diseases from the electronic health records **Springer, Cham, 2016**: 768-774.
- [28] Beaulieu-Jones B K, Greene C S. 2016, *Journal of biomedical informatics*, Semi-supervised learning of the electronic health record for phenotype stratification. **64**: 168-178.
- [29] Gupta P, Sivalingam U, Pölsterl S, et al. 2015. *Technical report, Technical University of Munich, Germany*, Identifying patients with diabetes using discriminative restricted boltzmann machines. **2015**
- [30] Tran T, Nguyen T D, Phung D, et al. 2015, *Journal of biomedical informatics*, Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). **54**: 96-105.
- [31] Liu Y, Wang Y, Kosorok M R, et al. 2016. *arXiv preprint* Robust hybrid learning for estimating personalized dynamic treatment regimens. arXiv:1611.02314,.
- [32] Meystre S M, Thibault J, Shen S, et al. 2010, *Journal of the American Medical Informatics Association*, Textractor: a hybrid system for medications and reason for their prescription extraction from clinical text documents. 17(5): 559-562.
- [33] Manogaran G, Varatharajan R, Priyan M K. 2018, *Multimedia tools and applications*. Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. **77(4)**: 4379-4399.

- [34] F. Dernoncourt, J. Y. Lee, O. Uzuner, and P. Szolovits, 2017. *American Medical Informat. Assoc.*, “De-identification of patient notes with recurrent neural networks,” J. **vol. 24, no. 3**, pp. 596–606
- [35] Zhu Z, Yin C, Qian B, et al. 2016. *2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE*. Measuring patient similarities via a deep architecture with medical concept embedding **2016**: 749-758.
- [36] Choi E, Bahadori M T, Song L, et al. 2017. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. GRAM: graph-based attention model for healthcare representation learning **ACM, 2017**: 787-795.
- [37] Suo Q, Ma F, Yuan Y, et al. 2018, *IEEE transactions on nanobioscience*, Deep patient similarity learning for personalized healthcare. **17(3)**: 219-227.
- [38] Nguygen T D. 2015. *Deakin Univeristy*. Structured representation learning from complex data. **2015**.