ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20190421122826
Subject review

Machine Learning Techniques for Screening and Diagnosis of Diabetes: a Survey

Yun-lei SUN, Da-lin ZHANG

Abstract: Diabetes has become one of the major causes of national disease and death in most countries. By 2015, diabetes had affected more than 415 million people worldwide. According to the International Diabetes Federation report, this figure is expected to rise to more than 642 million in 2040, so early screening and diagnosis of diabetes patients have great significance in detecting and treating diabetes on time. Diabetes is a multifactorial metabolic disease, its diagnostic criteria is difficult to cover all the ethology, damage degree, pathogenesis and other factors, so there is a situation for uncertainty and imprecision under various aspects of medical diagnosis process. With the development of Data mining, researchers find that machine learning is playing an increasingly important role in diabetes research. Machine learning techniques can find the risky factors of diabetes and reasonable threshold of physiological parameters to unearth hidden knowledge from a huge amount of diabetes-related data, which has a very important significance for diagnosis and treatment of diabetes. So this paper provides a survey of machine learning techniques that has been applied to diabetes data screening and diagnosis of the disease. In this paper, conventional machine learning techniques are described in early screening and diagnosis of diabetes, moreover deep learning techniques which have a significance of biomedical effect are also described.

Keywords: Deep Learning; diabetes; feature extraction; Machine Learning

1 INTRODUCTION

According to the International Diabetes Federation (IDF) [1] statistics, there were 415 million people suffering from diabetes around the world in 2015. By 2040 this number is expected to rise to over 642 million, as a consequence, diabetes has become the main cause of national disease and death in most countries. Diabetes is a group of metabolic diseases in which a person has high blood glucose, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced [2, 3]. If diabetes patients cannot control blood sugar well, it is effortless to induce cardiovascular, nervous system, eye, foot and other systemic diseases. Patients whose conditions are severe can also suffer from diabetic ketoacidosis, with a high disability [4]. Diabetes has a very great deal of harm to the human body, causing a series of complications, affecting the patient physical and mental health, bringing a heavy burden to family and society. Generally speaking, diabetes can be segmented into three types: type 1 diabetes, type 2 diabetes and gestational diabetes. Type 1 diabetes is an autoimmune disease that occurs in childhood. In this type of diabetes, the pancreatic cells that secrete insulin have been destroyed. Type 2 diabetes is caused by insulin resistance in various organs, leading to a marked increase in insulin demand, which accounts for almost 90% of the diabetes cases [5]. Gestational diabetes tends to occur among pregnant women, as the pancreas does not make sufficient amount of insulin [5].

The standards of early screening and diagnosis of diabetes are still in the exploratory stage on account of the unclear ethology and pathogenesis of diabetes. Through the continuous understanding of diabetes, the criteria of screening and diagnosis are constantly changing. Early diagnosis of diabetes mainly depends on clinical symptoms and signs. In 1965, the World Health Organization (WHO) first published diabetes diagnostic norm based on the clinical characteristics, but this criteria did not mention the diagnosis threshold of blood sugar levels [6]. With the developing understanding of diabetes, diagnostic criteria gradually increased fasting blood glucose (FPG), oral glucose tolerance test (OGTT), glycosylated haemoglobin (HbA1c) and other physiological parameters, as shown in Tab. 1. In 1980, the fasting blood glucose level was viewed as the main diagnostic norm [7]. In 1997, the new standard of American Diabetes Association (ADA) [8] increased the OGTT parameters. The standard of ADA published in 2010 [9] increased the HbA1c parameter. According to the latest research results, the ADA diagnostic criteria in 2017 [10] proposed the new standard of suffering from diabetes with FPG $\geq$ 126 mg/dl or HbA1c $\geq$ 6.5% or OGTT $\geq$ 200 mg/dl. As we can see from the evolution of the history of diabetes diagnosis, diagnosis of diabetes increasingly relies on the epidemiological support.

Table 1 The changing of diagnostic criteria of diabetes

Time	FPG	HbA1c	OGTT
1980	≥140 mg/dl [6]		
1997	≥126 mg/dl [7]		≥200 mg/dl
2010	≥126 mg/dl [8]	≥6.5%	≥200 mg/dl
2017	≥126 mg/dl [9]	≥6.5%	≥200 mg/dl

Accurate screening and diagnosis of diabetes require more effective features and have a high demand on the judgement which can be closer to the nature of the disease [4]. Some studies found that if we consider metabolic changes in diabetes from the perspective of body metabolism, doctors can better make a diagnosis of the type of diabetes and help patients with the more appropriate diabetic treatment. Metabolomics is a new discipline that has been developed in recent years to analyse all the low molecular weight metabolites of a certain organism or cell qualitatively and quantitatively. Through the change of endogenous metabolites and intermediates in diabetes and the evolution of coping rules, the metabolic status of the body can be further understood [11, 12].

On the basis of the study of early screening and diagnostic criteria for diabetes, diagnostic standards are increased from the initial clinical symptoms and signs to FPG, OGTT, HbA1c and other physiological parameters. Simultaneously clinical and demographic signs are also included in the diagnostic reference, such as sex, age, race/ethnicity, haemoglobin disease/anaemia, body mass index (BMI), cardiovascular disease, family history/ Genetic, medication records, etc. Even sedentary lifestyle

is one of the risky factors [13]. However, there is still no way to find out the pathogenesis of diabetes from the field of biology. It is urgent to clarify the pathology and diagnostic criteria of diabetes, it has a great significance in delaying the occurrence and development of diabetes, choosing drugs, reducing the incidence of diabetic complications and extending life expectancy.

With the continuous development of artificial intelligence and data mining technology, researchers begin to consider using machine learning techniques to search for the characteristics of diabetes. Machine learning techniques can find implied pathogenic factors in virtue of analysing and using diabetic data, with a high stability and accuracy in diabetic diagnosis. Therefore, machine learning techniques which can find out the reasonable threshold of risky factors and physiological parameters provide new ideas for screening and diagnosis of diabetes.

This paper is organized as the following: Section 2 offers the application of conventional machine learning methods in the screening and diagnosis of diabetes mellitus. This section introduces the supervised learning method represented by decision tree (DT), support vector machine (SVM) , artificial neural network (ANN) and the unsupervised learning method represented by clustering and association rules. Section 3 introduces the advantages of deep learning technology in medical data processing compared to conventional machine learning methods. Various applications of several network structures will then be introduced, such as deep belief network (DBN), deep recurrent neural networks (RNN) and deep Convolutional Neural Networks (CNN), which are applied in extracting feature, diagnosing disease and laying the foundation for screening and diagnosis of diabetes. Section 4 presents discussion and future works.

2 CONVENTIONAL MACHINE LEARNING TECHNIQUES

Diabetic diagnosis is based on a variety of epidemiology and genetic factors. Dangerous factors of epidemiology include smoking status, eating habits, physical activity, BMI and so on. Genetic factors are pathogenic genes which come from parents. Hence, doctors hope to consider all aspects of these factors and then predict and diagnose diabetes accurately; nevertheless researchers from the medical domain found that they could not explain the pathogenesis of diabetes. With the continuous development of Artificial Intelligence Technology, it has been found that machine learning techniques are very suitable for finding the reasonable threshold of risky factors and physiological parameters affecting diabetes. Why machine learning can achieve significant achievement in the medical domain? First of all, diabetes is a kind of chronic disease, and a lot of clinical treatment information will be generated in the process of treatment. Meanwhile, machine learning has giant advantages in handling big data problems, so the machine learning techniques can be applied to the analysis and processing of diabetes data. Secondly machine learning and medical diagnosis have the uniform objective to extract the correct and valuable information from a large number of data for making decisions. At the same time, machine learning techniques can avoid the misdiagnosis of inexperienced or tired human experts, and have a high stability and accuracy in the screening and diagnosis of diabetes [14]. Furthermore, machine learning techniques can also help patients have a clear idea of their health status as well as the situation of diabetic development, then patients can plan their own lifestyle to slow the deterioration of disease [15]. Therefore, we hope that we can use machine learning techniques to find pathogenesis of diabetes which cannot be found in the medical domain, which has great significance for treatment of diabetes patients early, the appropriate use of medicine and early rehabilitation. In this paper, the applications of conventional machine learning techniques in the early screening and diagnosis of diabetes mellitus will be introduced from two aspects: supervised learning and unsupervised learning.

2.1 Supervised Learning

The purpose of supervised learning is to establish an objective function describing the data model and to adjust the parameters of the classifier by using a set of known sample categories. The objective function is used to forecast value of output variable from a set of variables called input values of the function [16]. The training data is composed of a set of known output variable categories or variable values. There are two kinds of learning tasks in supervised learning: classification and regression. The classification model is used to predict different classes, and the regression model is used to predict numerical values. Common supervised learning techniques include DT, ANN and SVM. The following will be a brief study of different supervised learning techniques in the early screening and diagnosis of diabetes.

2.1.1 Decision Tree

DT is a tree structure, which is a form of a flowchart, and it is a classification algorithm with root nodes, intermediate nodes and leaf nodes. The class tags are assigned to the leaf nodes, the root nodes and the nonterminal nodes include different test conditions to separate the different attributes. The root node is selected based on information gain [17]. The most significant aspect is that these classification models can be easily understood by the medical practitioner. DT algorithm has the merits of clear and understandable decision-making process, which supplies a powerful method for the classification and prediction of diabetes. So some DT algorithms, for instance, Alternating Decision Trees (ADTree) with the accuracy of 83.68%, J48 with the accuracy of 91.38%, Naïve Bayesian Tree (NBTree) with the accuracy of 87.76%, Random Tree with the accuracy of 93.07%, REPTree with the accuracy of 89.22%, SimpleCart with the accuracy of 92.69% and so on, are commonly applied in medical screening and diagnosis [18]. In recent years, Sankaranrayan and Pramananda have discovered hidden knowledge from a great deal of diabetic data set via different data mining techniques such as rule-based classification and DT algorithms [19], which are significantly benificial in improving the quality of health care for diabetes patients.

Kaur and Chhabra put forward a modified J48 algorithm to forecast whether a patient is suffering from diabetes. J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules. This

paper uses the Pima Indians Diabetes Data Set and experimental results show that this improved J48 classification algorithm can reach accuracy of up to 99.87% while J48 just has the accuracy of 91.38 [20].

K. Rajesh et al. extracted features to reduce the number of features and dimension before screening and diagnosing the Pima Indian data set, so as to make the model generalization more powerful. In the selection of classification algorithm, they compare the error rate of C-RT (0.2148), CS-RT (0.2148), C4.5 (0.0938) and other classification algorithms in classification process, and finally choose the C4.5 algorithm with the lowest error rate commonly used in the medical field to classify. This algorithm can forecast whether a person is suffering from diabetes, and it can achieve the 91% classification accuracy, meaning a better performance than the other algorithms chosen [3].

Diabetic foot has a high incidence of diabetic complications, so it is important to study the risk of amputation in patients with diabetic foot. Rajesh and Sangeetha using DT algorithm C4.5 constructed two classifiers to predict the risk of future amputation of diabetes patients. DT technique focuses on attempting to determine those factors which affect the outcome of amputation in diabetic foot patients and creates a prognostic scoring program to assess the degree of inevitability involved in amputation in the diabetic foot patient. Through classification, we can obtain the following conclusions: the doppler status of flow in the affected limb and the clinical ulcer grade are the most important factors in determining the prognosis of the limb. And the results of the experiments show that complex classifier model has higher accuracy than simple one in our constructed classifier model [21].

Jelinek et al. using DT techniques searched HbA1c combination with biomarkers to screen and diagnose diabetes. The experimental result shows that if HbA1c levels are below or equal to the current cut-off of 6.5%, additional biomarkers could be used together with HbA1c to improve diagnostic accuracy in type 2 diabetes. Compared to single HbA1c cut-off standard, this method improves diagnostic accuracy as well as provides a new idea for diagnosing diabetes. Paying attention to the relationship between HbA1c and oxidative stress markers, inflammatory cytokines and so on, they drew the conclusion that both 8-hydroxy-2-deoxyguanosine (8-OhdG), an oxidative stress marker, and interleukin-6 (IL-6) improved classification accuracy [22].

2.1.2 Support Vector Machine

SVM is a supervised learning algorithm for classification. The nature of the SVM removal sample over-fitting increases the prediction accuracy [17]. It uses a kernel function to transform data from input space into a high dimensional feature space in which it searches for a separating hyperplane. To transform input space into desired feature space, it is essential to select kernel functions. There are some ordinarily used kernel function, such as linear, polynomial, puk and Radial Basic Function (RBF) kernels. Tapas Ranjan Baitharu et al. compared some kernel functions and found that Linear Kernel is more beneficial for dealing with diabetes datasets [23].

Calisir et al. proposed an automatic diagnostic system called LDA-MWSVM for feature extraction and diagnostic classification of diabetes. The system is made up of three parts: first, use linear discriminant analysis (LDA) for feature extraction, subsequently, use Morlet Wavelet Support Vector Machine (MWSVM) to classify, finally, the correct diagnosis performance of this automatic system based on LDA–MWSVM for the diagnosis of diabetes is calculated by using sensitivity and specificity analysis, classification accuracy, and confusion matrix, respectively. We can conclude from experimental results that the method of LDA and MWSVM based on classifier is capable of screening and diagnosing diabetes [24].

Protein-protein interactions (PPI) is considered to be the key strategies to understand the molecular mechanism underlying any disease. For the purpose of searching the nature causes of diabetes, some researchers use machine learning techniques to forecast protein-protein interactions for diagnosis of diabetes. Vyas et al. built and analysed protein-protein interactions related to diabetes by using SVM, biomedical text mining and network analysis. First, this paper uses text mining method to obtain the latest news of diabetes-related proteins and extract the unknown recessive diabetes protein knowledge from literature. Then we use the LibSVM software [25] based on SVM strategy to construct the classification model for predicting PPI. The model developed in the present work with an accuracy of 78.2% can be used to identify diabetes mellitus related proteins. It is of important meaning for uncovering disease-associated mutations, identifying drug targets and biomarkers for complicated disease [26].

In order to predict glucose levels in type 1 diabetes, Georga et al. developed a model of glucose metabolism to conduct health care for diabetes patients. This model can predict a large variety of stimulating blood glucose responses and provide knowledge of abnormal blood glucose changes. This paper solves the problem of subcutaneous glucose prediction in type 1 diabetes patients by using support vector regression technique, it cannot only give a better understanding of the dependencies between the input variables and the glucose, but also the problem of predicting hypo/hyperglycaemia events in addition to predicting values of significant importance since it can offer the patient with alerting capabilities [27].

In many machine learning methods of screening and diagnosis of diabetes mellitus, SVM has higher accuracy. But SVM has the flaw of poor interpretation, which is a big drawback in medical field. Han et al. used SVM to screen diabetes and add an ensemble learning module to turn the "black box" of SVM into comprehensible and transparent rule. First, obtains SVM model with better classification accuracy through training parameters, puts the support vector into model to predict, and the prediction tag of support vector will replace the original label of support vector. Subsequently, the artificial data was entered into the random forest (RF) algorithm, and the best rule set is generated by adjusting the rule induction method and evaluated with the test set. Results show that our proposed model has high quality in terms of diagnosis with precision, which means the diagnosis ability of the model. The model can evaluate undiagnosed individuals in an understandable form and provide more comprehensive and transparent representation for end users [28].

The essential method to treat diabetes is to detect potentially causative genes, which benefit researchers to develop new drugs which not only control the disease but also can treat it. But it is the biggest problem that genes data are large in number with different time series data or dimensionality. To solve this problem, Atul Kumar et al. proposed a modified version of SVM that can reduce dimension, known as Support Vector Machine Recursive Feature Elimination (SVMRFE). The idea of SVMRFE is to build a model which can rank the importance of genes and eliminate the genes which cannot cause diabetes. Experimental results further identify the involvement of the coding genes in type 2 diabetes. This is conducive to the development of targeted drugs to fundamentally treat patients [29].

2.1.3 Artificial Neural Network

ANNs are notably suitable for predicting the result of disease diagnosis, which is a powerful tool for analysing complex clinical data. During the training process, ANNs use known data and identify the complex relationship between the input and the output. After training, ANNs can be used to predict the output value of given input data. In general, ANNs are efficient for many complex, non-linear or incomplete data.

Metabolic syndrome (Mets) refers to the body's protein, fat, carbohydrates and other metabolic disorders substances leading to a series of clinical syndromes. Mets is one of the well-known risky factors of causing chronic diseases such as cardiovascular disease, type 2 diabetes, cancer and chronic kidney disease. Whence, it has a great significance in determining the cause of diabetes by using ANN to identify metabolic syndrome. Hirose et al. applied ANN for prediction of the 6-year incidence of MetS by using clinical data [30]. Chen et al. concluded that ANN is preferable to the traditional logistic regression analysis for assessing the risk of MetS from sex, age, BMI, waist circumference, waist-to-height ratio, hip circumference, systolic and diastolic blood pressure [31]. Darko Ivanovic´ proposed the feed-forward ANN with back propagation as the training algorithm to predict MetS. The contribution of this paper is simple input vector, a more detailed search of optimal ANN architecture [32].

Overweight individuals are at higher risk for developing type 2 diabetes than general population. So as to study the correlation between overweight population and type 2 diabetes, Wang et al. established a blood glucose prediction model for overweight patients by analysing the correlation between blood glucose and biochemical parameters. First, using multiple linear regression (MLR) they select diabetes-related variables which include the levels of fasting glucose (fs-GLU), blood lipids, and hepatic and so on. The back-propagation artificial neural network (BP-ANN) model was then trained to predict blood glucose levels. The results showed that the proposed BP-ANN model could predict fasting blood glucose level in overweight patients based on six related fasting biochemical metabolism indicators (age, alanine aminotransferase, urea nitrogen, total protein, uric acid, body mass index) [33].

Input values of the ANN often involve laboratory parameters. To diagnose diabetes in patients who do not have blood tests or blood pressure measurements, Sumathy et al. proposed an ANN technique. The inputs of ANN designed are based on the symptoms, which could appear during the early stages of diabetes and based on the physical conditions. This proposed methodology brings great convenience to diabetes patients, reduces the cost of different medical tests and serves as a better tool for the diagnosis of diabetes [34].

ANNs are applied to screening and diagnosing diabetes along with diabetic complications. Rau et al. applied the ANN to prediction model of liver cancer in patients with type 2 diabetes for timely detection and diagnosis. Based on the prediction model, they came up with a web-based application to allow physicians to detect the probability of liver cancer in future 6-year period [35].

In recent years, mobile health medical equipment has been developed, ANN has been also applied to mobile devices to realize automatic diagnosis of diabetes, Giveki et al. used neural network algorithms to develop a distributed end-to-end three-tier health care system. The first layer is a sensor and wearable equipment for monitoring the vital signs of the body, the second layer is client devices such as PDAs and computers between the first and last layers of the mediator and communicator, the third layer is a high-end powerful desktop server to provide users with medical services and database operations. The ANN model is used to diagnose diabetes in the second and third layers [36].

2.2 Unsupervised Learning

In medical domain, there are some clinical data without any corresponding labels. Supervised learning techniques are not suitable for dealing with data without any corresponding labels. However, unsupervised learning is good at discovering associations between hidden structures of data or variables without label [16]. Common unsupervised learning techniques include clustering technique and association rule technique, so the following will introduce some applications of different unsupervised learning techniques in the early screening and diagnosis of diabetes.

2.2.1 Clustering Techniques

Although there is a large amount of unlabelled data in the medical domain, clustering method can find important patterns in unorganized and massive data collections. Velu and Kaswan diagnosed diabetes by using EM algorithms, h-means+ clustering, and genetic algorithms to form clusters with similar symptoms. The experimental results show that the h-means+ clustering method has better clustering performance than other clustering methods [37].

Clustering method can also be applied to feature extraction. Sideris et al. proposed a feature extraction framework based on clustering (hierarchical clustering) by analysing disease information, which is tested by using the data set of patients with heart failure and the data set of diabetes patients. The experimental results show that this method can diagnose the severity of patient's condition and improve the prediction accuracy of the risk assessment of patients' readmission [38].

2.2.2 Association Rule Learning

Association rule is the rule whose support degree and trust degree meet the given threshold value respectively, association rules are of the form $\{X_1, ..., X_n\} \rightarrow Y$. The most

well-known association rule discovery algorithm is Apriori, which was proposed by Rakesh Agrawal in 1994. Association rule is one of the most valuable tools for unsupervised data analysis in biology and bioinformatics, which can be used for biological sequence analysis, gene expression data analysis, and frequent patterns discovery from biological data.

Association rule is of great significance in detecting the risk of diabetes in patients, which has a great significance in taking appropriate treatment at the right time. Murari Devakannan Kamalesh et al. summarized the rules, analysed and compared four methods. Experimental results indicated that the method called bottom-up summarization (BUS) can find high-risk diabetes subgroups with high accuracy [39].

Ramezankhani et al. used association rule to find the pathogenesis of type 2 diabetes mellitus. This paper uses the Apriori algorithm, first uses minimal support to identify all frequent itemsets in the database, then applies these frequent itemsets and minimum confidence to generating rules, finally extracts two rules for men and women respectively. Experimental results explain that Fasting Blood Glucose (IFG), impaired glucose tolerance (IGT) and BMI ≥30 kg/m^2, family history of diabetes, wrist circumference >16.5 cm and waist height ≥0.5 can increase the risk of diabetes for female. For men, the combination of IGT, IFG, city residence time (>40 years), central obesity, total cholesterol and high density lipoprotein ratio ≥5.3, low physical activity, chronic kidney disease and wrist circumference >18.5 cm are identified as a risk model for diabetes. Association rules apply to different combinations of risk factors rather than individual predictive factor [40].

While association rule mining to electronic medical records can discover sets of risk elements and their corresponding subpopulations of developing diabetes, however, electronic medical records usually have the characteristics of high-dimensional, which leads to generating a very large set of association rules, reducing the interpretability of entire rule set. In order to solve this problem, G.J. Simon et al. used the rule set aggregation technique to compress the original rule set into a more compact set that can be interpreted. In this paper, we propose four kinds of association rule set aggregation techniques, which can apply association rules mining to determining the combination of empathy in clinical. Finally, the performance, applicability, advantages and disadvantages of the four methods are evaluated [41].

The object of rule mining is to find a small set of integrated rules based on a training dataset. However, the accuracy of these algorithms is not high while considering the sensitivity and specificity. Ramalingaswamy Cheruku et al. proposed SMO-based rule miner called SM-RuleMiner to generate a comprehensive optimal ruleset while balancing accuracy, sensitivity and specificity [42].

Electronic health records (EHRs) consist of multiple time series of clinical variables collected for a specific patient, such as laboratory test results and medication orders. For the purpose of making full use of time series information, Batal et al. proposed the Recent Temporal Pattern (RTP) mining framework, which is used to establish classification model, accurately detect adverse reactions, and apply it to monitor the future development of patients' conditions, so as to improve the performance of the classifier [43].

3 DEEP LEARNING

Conventional machine-learning techniques were limited in dealing with natural data in their raw format. To solve this problem, researchers put forward the concept of representation learning, which is a set of methods that allow a machine to be fed with raw data and to automatically discover the representations needed for detection or classification [44]. Deep-learning methods are representation-learning methods with multiple layers of representation, obtained by combining simple but non-linear modules, each of which transforms a level of representation (starting with the original input) into a higher, more abstract representation [44]. Deep learning differs from "shallow" learning. In deep learning, we have more than one hidden layer between input and output layers. We compute the input to each unit except input layer, as the weighted sum of units from the previous layer; then we usually apply nonlinear transformation, or activation function to the input of a unit so as to acquire a new representation of the input from previous layer [45]. We have weights on links between units from layer to layer. For deep learning, we can summarize that by calculating the forward flow from input to output, and then propagating errors from output back to input, the weights can be updated to optimize some loss functions [46]. Most Deep Neural Networks (DNNs) can be classified into three major categories: supervised learning, unsupervised learning and Hybrid or semisupervised networks. In supervised learning, labelled data are used to train DNNs and learn the weights that minimize the error to forecast a target value for classification or regression, whereas in unsupervised learning, the training is performed without requiring labelled data. We can cluster, extract feature and reduce dimensionality by unsupervised learning. In semisupervised learning, DNNs combine an initial training procedure with an unsupervised learning step to extract the most relevant features and then use those features to classify exploiting a supervised learning step [47].

Because the characteristics of medical data are high-dimensional, noisy, heterogenous, sparse and so on we anticipate that deep learning will have a better performance in dealing with this type data. First, DNNs require very large data sets, while patients with chronic diseases produce a large number of medical records. Second, DNNs have the excellent ability to deal with the high-dimensional, sparse, noisy data with nonlinear relationships which are just the characteristics of medical data [47]. Third, DNNs have high generalization ability, therefore we can train a data set and then apply well-trained model to new data sets. DNNs have the suitability of biological data and the potential applications. Deep learning also can be applied in many fields of medicine, such as translational bioinformatics, medical imaging, medical informatics, public health [47]. This section covered several applications of deep learning methods, including DBNs, deep RNNs, deep CNNs and so on. The aim of applying these deep networks is to extract medical features and lay the foundation for screening and diagnosis of diabetes.

3.1 Deep Belief Network

DBN can be considered as a composition of Restricted Boltzman Machines (RBMs) where each sub-network's hidden layer is connected to the visible layer of the next RBM. DBN has great advantages in disease feature extraction and data mining of HER [48].

Most patients of type 1 diabetes have symptoms of hypoglycaemia. In recent years, research has got the result that heart rate (HR) and correct QT interval (QTc) of the electrocardiogram (ECG) signal are found as the most common physiological parameters to be effected from hypoglycaemic reaction. San et al. using DBN detect blood glucose levels and diagnose whether diabetes patients have symptoms of hypoglycaemia. The results indicate that DBN can effectively detect the occurrence of hypoglycaemia [49].

Mehrabi et al. search the common time pattern in the diagnostic matrix through the DBN. In this paper, the longitudinal records of each patient are expressed as diagnostic matrices. Deep learning algorithm is used to discover conventional patterns between patients, providing new ideas for discovering new potential correlations and generating new hypotheses [50].

Diabetic complications can be caused by long-term uncontrolled blood glucose and one of the most common complications is diabetic retinopathy [51]. In order to carry out its early diagnosis, Arunkumar proposed a DBN method to extract features by extracting variables related to diabetic retinopathy, which also contributes to the development of automated screening systems [52].

3.2 Deep Recurrent Neural Network

RNN is a class of neural network whose connections of units form a directed cycle. This nature grants its ability to work with temporal data, so it is good at working with tasks on time series data, or sequential data.

Edward Choi et al. developed a Doctor AI to predict clinical events via RNNs based on EHRs. They put records as input to forecast diagnosis and medication categories for a subsequent visit. The strategy of this model is to develop a generic way to represent the patient's temporal healthcare experience so as to predict all diagnoses, drug classes, and visit times [8].

Jagannatha and Yu et al. proposed a predictive model of sequence tag structures based on RNNs in clinical texts. The focus of this work is to label clinical events (drugs, indications, and adverse drug events) and event-related attributes (drug dose, pathway, etc.) in unstructured clinical records [53].

In order to solve the problem of long-term dependency between tags, Jagannatha et al. used Long Short-Term Memory (LSTM) as our RNN model and model the CRF pairwise potentials using neural networks. The experimental results show that structured prediction models are good directions to improve the accurate phrase extraction of clinical entities [54]. Pham et al. introduced DeepCare, a deep dynamic neural network that can read medical records and predict future medical outcomes. Built on LSTM, DeepCare introduces time parameterizations to handle irregular timing by moderating the forgetting and consolidation of illness memory. We demonstrate the efficacy of DeepCare for disease progression modelling and readmission prediction in diabetes [55].

In addition, Lipton et al. used circulating neural networks to model clinical time series with missing values [14]. Che and Purushotham et al. proposed that the GRU-D circulatory neural network model also can be used to treat multiple clinical time series with missing values [56].

3.3 Deep Convolutional Neural Network

CNNs have had the greatest impact within the field of medical domain. Its architecture can be defined as an interleaved set of feedforward layers implementing convolutional filters followed by reduction, rectification or pooling layers. Each layer in the network originates a high-level abstract feature [57].

Nguyen P et al. present Deepr, a new end-to-end deep learning system that learns to extract features from medical records and predicts future risk automatically. Deepr is a multi-layered architecture based on CNNs, it is able to uncover the underlying space of diseases and interventions, show the relationships between them, so as to detect care patterns and disease [58].

Torre et al. applied deep CNNs to automatically classify retinal images. They divided the retina images into five levels using supervised deep learning. Results show that it is possible to achieve a quadratic weighted kappa classification score over 0.75, which is not far from human expert reported scores of 0.80 [59].

Pratt et al. use the deep CNN to diagnose diabetic retinopathy and classify its severity to determine the fundus color. Deep CNNs can automatically classify micro-aneurysms, exudate and haemorrhages on the retina [60]. On the data set of 80,000 images used our proposed CNN achieves a sensitivity of 95% and an accuracy of 75% on 5,000 validation images [61].

3.4 Other Deep Networks

Autoencoders can be successfully applied in the feature extraction and dimensionality reduction. In 2016, Miotto R et al. came up with a new unsupervised deep feature learning method, which is a three-layer stack of denoising Autoencoders. It was used to capture hierarchical regularities and dependencies in the aggregated EHRs of approximately 700,000 patients from the Mount Sinai data warehouse. This method can obtain a general-purpose patient representation from EHR data, making the clinical predictive model more convenient. The experimental results are significantly better than those achieved using representations on the basis of raw EHR data and alternative feature learning strategies. In the prediction of severe diabetes, the performance of "deep patient" is far ahead. Its innovation lies in the discovery that the application of deep learning to EHRs can be demonstrated by patients, thus helping us improve clinical predictions and provide a deep learning framework for strengthening clinical decision-making [59, 60].

Nie L et al. put forward a novel deep learning scheme to automatically infer the possible diseases of the given questions in community-based health services [64].

Kamble et al. used Restricted Boltzmann Machine to detect whether patient has diabetes. If patients have diabetes, the DT technique will be used in this paper to detect whether the patient has type 1 or type 2 diabetes [65].

4 CONCLUSIONS AND FUTURE WORK

We choose to apply the machine learning method to extracting diabetes feature so as to get a generic model for diagnosing and predicting diabetes. However, because the conventional machine learning model relies on the applied data set and selected parameters [66], the generality of the

disease prediction model is reduced, and only can be applied to specific populations. It is well-known that different risky assessment models rely on different population data sets, therefore the lack of repeatability and external validation is the biggest challenge of conventional machine learning in diagnosing and predicting diabetes [15]. At the same time, it is tough or unsatisfying to deal with the high-dimension, heterogeneity and sparse data sets. In addition, the performance of conventional machine learning method depends on data pre-processing extremely, the ability to handle raw data sets is restricted.

In the wake of deep learning, some problems which are difficult to be solved by conventional machine learning methods begin to have new ideas and new methods. Due to the lack of commonality in conventional machine learning, researchers have developed a universal disease diagnosis model that does not rely on the data sets, which is of great significance to model versatility. Simultaneously because of the characteristics of deep learning, it is very good at discovering intricate structures in high-dimension, noise, and heterogeneity and sparseness data [67].

The general model introduced in this paper uses deep learning to extract medical features and develop medical diagnosis, which lays a foundation for the feature extraction, diagnosis and treatment of diabetes. In several applications of deep learning networks, DBNs can be configured to avoid overfitting and can be applied to various types of biomedical data, and it is suitable to extract diabetic feature. RNNs are good at dealing with time series data or sequential data. CNNs are mainly applied to image recognition, for example, they classify retinal images to diagnose diabetic retinopathy. Autoencoders which have the ability to learn flexible and rich representations of data can be successfully used for feature extraction and dimensionality reduction, and they have some excellent applications in diabetes diagnosis. Tab. 2 shows the comparison of different algorithms.

Table 2 The comparison of different algorithms

Algorithm	DataSet	Accuracy
ADTree	DiScRi	83.68% [18]
J48	DiScRi	91.38% [18]
NBTree	DiScRi	87.76% [18]
RandomTree	DiScRi	93.07% [18]
REPTree	DiScRi	89.22% [18]
SimpleCart	DiScRi	92.69% [18]
Improved J48 classification algorithm	The Pima Indians Diabetes Data Set	99.87% [20]
C4.5 algorithm	The Pima Indians Diabetes Database	91% [3]
C-RT	The Pima Indians Diabetes Database	78.5% [3]
CS-RT	The Pima Indians Diabetes Database	78.5% [3]
LDA-MWSVM	The Pima Indians Diabetes Database	89.74% [24]
SVM strategy	PPI	78.2% [26]
SVM+RF	PIMA	89.02% [28]
SVMRFE	GEO, DGAP	83.9% [29]
BP-ANN	346 overweight Chinese people patients ages 18–81 years	99.87% [33]
RTP	NEURO	77.34% [43]
CNNs	5,000 validation images	75% [61]

Although deep learning behaves wonderfully in medical domain, it still has many problems to be solved. Firstly, we cannot explain the entire deep learning model despite some recent work on visualizing high level features by using the weight filters. Secondly, in the training process of deep learning network (especially in the case of small datasets), a common problem is over fitting. In this case, the network is able to memorize the training examples normally, but cannot be generalized to new samples that it has not already observed. Then if we apply deep learning tools, we are supposed to pre-process the dataset as input for deep learning network. It is a great challenge to find an effective classification model by correctly preprocessing the data and finding the optimal hyperparameter set. The last aspect that we would like to underline is that numerous deep learning networks can be easily fooled [57].

In the future perspectives of machine learning methods for screening and diagnosis diabetes, we should combine the power of deep learning to learn about flexible, rich data representations with the transparency and interpretability of traditional machine learning techniques [47]. A combination of these approaches uses supervised, unsupervised, and reinforce learning to understand the process of diabetic diagnosis and develop personalized medical care. And in the future development process it is better to add the domain knowledge of diabetes to improve the performance of the model. Through the combination of several aspects, we hope to extract new feature to clear diabetes pathology. This review was to plant a seed of interest for screening and diagnosis of diabetes with machine learning method. Through this paper, we see a lot of value in increased collaboration between biologists and the computational biology community, simultaneously open up a discussion about the many opportunities that biological data not just diabetes data offer for application of this approach, including the hurdles to overcome and many possible directions moving forward [47].

Acknowledgements

This work was supported in part by the Fundamental Research Funds for the Central Universities (Grant No. 18CX02019A).

5 REFERENCES

[1] Bloomgarden, Z. (2016). Questioning glucose measurements used in the International Diabetes Federation (IDF) Atlas. *Journal of Diabetes, 8*(6): 746-747. https://doi.org/10.1111/1753-0407.12453

[2] Zhao Ming, Wang Xiaoxia, & Zhu Xiaowei. (2014). Understanding diabetes from the diagnosis of diabetes mellitus. *Journal of Diagnostics Concepts & Practice*, 2, 226-228.

[3] Rajesh, K. & Sangeetha, V. (2012). Application of Data Mining Methods and Techniques for Diabetes Diagnosis. *International Journal of Engineering and Innovative Technology (IJEIT), 2*(3). 224-229.

[4] Thangadurai, D. K. & Nandhini, N. (2016). Comparison of datamining algorithms for prediction and diagnosis of diabetesmellitus. *International Journal of Scientific & Engineering Research, 7*(5), 2229-5518.

[5] Iyer, A., Jeyalatha, S., & Sumbaly, R. (2015). Diagnosis of diabetes using classification mining techniques. *International Journal of Data Mining & Knowledge Management Process, 5*(1), 1-14. https://doi.org/10.5121/ijdkp.2015.5101

[6] McLintock, J. S. (April 1966). Diabetes Mellitus**.** Technical Report of a W. H. O. Expert Committee Series No. 310(1965) 44 pp., 9½" × 6Ū"5/-, World Health Organisation, Geneva**,** The Annals of Occupational Hygiene**,** 9(2), p. 91.

https://doi.org/10.1093/annhyg/9.2.91
[7] Listed, N. (1980). WHO Expert Committee on Diabetes Mellitus: second report. *World Health Organization Technical Report*, 646, 1-1.
[8] Choi, E., Bahadori, M. T., Schuetz, A., et al. (2015). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. Computer Science.
[9] Association, A. D. (2010). Diagnosis and Classification of Diabetes Mellitus. *American Family Physician, 58*(6), S62-S69. https://doi.org/10.2337/dc11-S062
[10] Association, A. D. (2017). 2. Classification and Diagnosis of Diabetes. *Diabetes Care, 40*(Suppl 1), S11. https://doi.org/10.2337/dc17-S005
[11] Cristina, M., Eric, F., Idil, E., et al. (2013). Biomarkers for Type 2 Diabetes and Impaired Fasting Glucose Using a Nontargeted Metabolomics Approach. *Diabetes, 62*(12), 4270. https://doi.org/10.2337/db13-0570
[12] Bain, J. R. & Muehlbauer, M. J. (2013). Metabolomics reveals unexpected responses to oral glucose. *Diabetes, 62*(62), 2651-2653. https://doi.org/10.2337/db13-0605
[13] Sohn, M., Talbert, J., Blumenschein, K., et al. (2015). Atypical antipsychotic initiation and the risk of type II diabetes in children and adolescents. *Pharmacoepidemiology & Drug Safety, 24*(6), 583. https://doi.org/10.1002/pds.3768
[14] Lipton, Z. C., Kale, D. C., & Wetzel, R. (2016). Directly Modeling Missing Data in Sequences with RNNs: Improved Classification of Clinical Time Series. *Proceedings of the 1st Machine Learning for Healthcare Conference, PMLR 56*, 253-270.
[15] Shankaracharya. (2017). Diabetes risk prediction using machine learning: prospect and challenges. *J Bioinfo Proteomics Rev, 3*(2), 1- 2. https://doi.org/10.15436/2381-0793.17.1317
[16] Kavakiotis, I., Tsave, O., Salifoglou, A., et al. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational & Structural Biotechnology Journal, 15*, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
[17] V. V. V. & C. A. (2015). Prediction and Diagnosis of Diabetes Mellitus - A Machine Learning Approach. *IEEE Recent Advances in Intelligent Computational Systems*, 12, 10-12.
[18] Kelarev, A. V., Stranieri, A., Yearwood, J. L., et al. (2012). Empirical Study of Decision Trees and Ensemble Classifiers for Monitoring of Diabetes Patients in Pervasive Healthcare. *Proceedings of the International Conference on Network-Based Information Systems, F*. https://doi.org/10.1109/NBiS.2012.20
[19] Sankaranarayanan, S. & Perumal, T. P. (2014). A Predictive Approach for Diabetes Mellitus Disease through Data Mining Technologies. *Proceedings of the Computing and Communication Technologies, F*. https://doi.org/10.1109/WCCCT.2014.65
[20] Kaur, G. & Chhabra, A. (2014). Improved J48 Classification Algorithm for the Prediction of Diabetes. *International Journal of Computer Applications, 98*(22), 13-17. https://doi.org/10.5120/17314-7433
[21] Kasbekar, P. U., Goel, P., & Jadhav, S. P. (2017). A Decision Tree Analysis of Diabetic Foot Amputation Risk in Indian Patients. *Frontiers in Endocrinology*, 8. https://doi.org/10.3389/fendo.2017.00025
[22] Jelinek, H. F., Stranieri, A., Yatsko, A., et al. (2016). Data analytics identify glycated haemoglobin co-markers for type 2 diabetes mellitus diagnosis. *Comput Biol Med*, 75, 90-97. https://doi.org/10.1016/j.compbiomed.2016.05.005
[23] Baitharu, T. R., Pani, S. K., & Dhal, S. K. (2015). Comparison of Kernel Selection for Support Vector Machines Using Diabetes Dataset. *Journal of Computer Sciences and Applications, 3*(6), 181-184.
[24] Ali Ir D. & Antekin, E. (2011). *An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier*. Pergamon Press, Inc. https://doi.org/10.1016/j.eswa.2011.01.017
[25] Chang, C. C. & Lin, C. J. (2011). *LIBSVM: A library for support vector machines*. ACM. https://doi.org/10.1145/1961189.1961199
[26] Vyas, R., Bapat, S., Jain, E., et al. (2016). Building and analysis of protein-protein interactions related to diabetes mellitus using support vector machine, biomedical text mining and network analysis. *Computational Biology & Chemistry*, 65, 37-44. https://doi.org/10.1016/j.compbiolchem.2016.09.011
[27] Georga, E. I., Protopappas, V. C., Ardigo, D., et al. (2013). Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine & Biology Society, 17*(1), 71-81. https://doi.org/10.1109/TITB.2012.2219876
[28] Han, L., Luo, S., Yu, J., et al. (2015). Rule Extraction from Support Vector Machines Using Ensemble Learning Approach: An Application for Diagnosis of Diabetes. *IEEE Journal of Biomedical & Health Informatics, 19*(2), 728. https://doi.org/10.1109/JBHI.2014.2325615
[29] Kumar, A., Sharmila, D. J. S., & Singh, S. (2017). SVMRFE based approach for prediction of most discriminatory gene target for type II diabetes. *Genomics Data*, 12, 28-37. https://doi.org/10.1016/j.gdata.2017.02.008
[30] Hirose, H., Takayama, T., Hozawa, S., et al. (2011). Prediction of metabolic syndrome using artificial neural network system based on clinical data including insulin resistance index and serum adiponectin. *Computers in Biology & Medicine, 41*(11), 1051-1056. https://doi.org/10.1016/j.compbiomed.2011.09.005
[31] Chen, H., Xiong, S., & Ren, X. (2014). Evaluating the Risk of Metabolic Syndrome Based on an Artificial Intelligence Model. *Abstract and Applied Analysis*, (2014-5-5), 2014(2), 353-370. https://doi.org/10.1155/2014/207268
[32] Ramezankhani, A., Pournik, O., Shahrabi, J., et al. (2015). An Application of Association Rule Mining to Extract Risk Pattern for Type 2 Diabetes Using Tehran Lipid and Glucose Study Database. *International Journal of Endocrinology & Metabolism, 13*(2), e25389. https://doi.org/10.5812/ijem.25389
[33] Wang, J., Wang, F., Liu, Y., et al. (2016). Multiple Linear Regression and Artificial Neural Network to Predict Blood Glucose in Overweight Patients. *Experimental & Clinical Endocrinology & Diabetes, 124*(1), 34-38. https://doi.org/10.1055/s-0035-1565175
[34] Sumathy, M., Mythili, P., Kumar, P., et al. (2010). Diagnosis of Diabetes Mellitus based on Risk Factors. *Blood*. https://doi.org/10.5120/1473-1989
[35] Rau, H. H., Hsu, C. Y., Lin, Y. A., et al. (2016). Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. *Computer Methods & Programs in Biomedicine*, 125, 58. https://doi.org/10.1016/j.cmpb.2015.11.009
[36] Giveki, D., Salimi, H., Bahmanyar, G. R., et al. (2012). Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search. *Computer Science*, abs/1201.2173.
[37] Velu, C. M. & Kashwan, K. R. (2013). Visual data mining techniques for classification of diabetic patients. *Proceedings of the Advance Computing Conference, F*. https://doi.org/10.1109/IAdCC.2013.6514375
[38] Sideris, C., Pourhomayoun, M., Kalantarian, H., et al. (2016). A flexible data-driven comorbidity feature extraction framework. *Computers in Biology & Medicine*, 73, 165-172.

https://doi.org/10.1016/j.compbiomed.2016.04.014
[39] Kamalesh, M. D., Prasanna, K. H., Bharathi, B., et al. (2016). *Predicting the Risk of Diabetes Mellitus to Subpopulations Using Association Rule Mining*. Springer India. https://doi.org/10.1007/978-81-322-2671-0_6
[40] Kazempour-Ardebili, S., Ramezankhani, A., Eslami, A., et al. (2017). Metabolic mediators of the impact of general and central adiposity measures on cardiovascular disease and mortality risks in older adults: Tehran Lipid and Glucose Study. *Geriatrics & Gerontology International*. https://doi.org/10.1111/ggi.13015
[41] Simon, G. J., Caraballo, P. J., Therneau, T. M., et al. (2015). Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus. *IEEE Transactions on Knowledge & Data Engineering, 27*(1), 130-141. https://doi.org/10.1109/TKDE.2013.76
[42] Cheruku, R., Edla, D. R., & Kuppili, V. (2017). SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. *Comput Biol Med*, 81, 79-92. https://doi.org/10.1016/j.compbiomed.2016.12.009
[43] Batal, I., Fradkin, D., Harrison, J., et al. (2012). Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, F*. https://doi.org/10.1145/2339530.2339578
[44] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-44. https://doi.org/10.1038/nature14539
[45] Zhang, D., Sui, J., & Gong, Y. (2017). Large scale software test data generation based on collectiveconstraint and weighted combination method. *Tehnicki Vjesnik, 24*(4), 1041-1050. https://doi.org/10.17559/TV-20170319045945
[46] Li, Y. (2017). Deep Reinforcement Learning: An Overview. eprint arXiv:1701.07274.
[47] Mamoshina, P., Vieira, A., Putin, E., et al. (2016). Applications of Deep Learning in Biomedicine. *Mol Pharm, 13*(5), 1445-1454. https://doi.org/10.1021/acs.molpharmaceut.5b00982
[48] Ravi, D., Wong, C., Deligianni, F., et al. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics, 21*(1), 4-21. https://doi.org/10.1109/JBHI.2016.2636665
[49] San, P. P., Ling, S. H., Nguyen, H. T., et al. (2016). Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. *Proceedings of the International Conference of the IEEE Engineering in Medicine & Biology Society, F*. https://doi.org/10.1109/EMBC.2016.7591483
[50] Mehrabi, S., Sohn, S., Li, D., et al. (2015). Temporal Pattern and Association Discovery of Diagnosis Codes Using Deep Learning. *Proceedings of the International Conference on Healthcare Informatics, F*. https://doi.org/10.1109/ICHI.2015.58
[51] Sun, Yunlei & Zhang, Dalin. (2019). Diagnosis and Analysis of Diabetic Retinopathy based on Electronic Health Records, IEEE, 1-3. https://doi.org/10.1109/ACCESS.2019.2918625
[52] Arunkumar, R. & Karthigaikumar, P. (2015). Multi-retinal disease classification by reduced deep learning features. *Neural Computing & Applications*, 1-6.
[53] Jagannatha, A. N. & Yu, H. (2016). Bidirectional RNN for Medical Event Detection in Electronic Health Records. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, F*. https://doi.org/10.18653/v1/N16-1056
[54] Jagannatha, A. N. & Yu, H. (2016). Structured prediction models for RNN based sequence labeling in clinical text. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, F*. https://doi.org/10.18653/v1/D16-1082
[55] Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016) DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. In: Bailey J., Khan L., Washio T., Dobbie G., Huang J., Wang R. (eds) *Advances in Knowledge Discovery and Data Mining. PAKDD 2016. Lecture Notes in Computer Science, vol 9652*. Springer, Cham. https://doi.org/10.1007/978-3-319-31750-2_3
[56] Che, Z., Purushotham, S., Cho, K., et al. (2018). Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*, 8, Article number: 6085. https://doi.org/10.1038/s41598-018-24271-9
[57] Ravi, D., Wong, C., Deligianni, F., et al. (2016). Deep Learning for Health Informatics. *IEEE Journal of Biomedical & Health Informatics*, PP(99), 1-1. https://doi.org/10.1109/JBHI.2016.2636665
[58] Nguyen, P., Tran, T., Wickramasinghe, N., et al. (2017). $\mathtt {Deepr}$: A Convolutional Net for Medical Records. *IEEE Journal of Biomedical & Health Informatics, 21*(1), 22-30. https://doi.org/ 10.1109/JBHI.2016.2633963
[59] Torre, J. D. L., Valls, A., & Puig, D. (2016). Diabetic Retinopathy Detection Through Image Analysis Using Deep Convolutional Neural Networks. *Frontiers in Artificial Intelligence and Applications*, 288, 58-63.
[60] Sun, Yunlei. (2019). The Neural Network of One-Dimensional Convolution-An Example of the Diagnosis of Diabetic Retinopathy, IEEE. https://doi.org/10.1109/ACCESS.2019.2916922
[61] Pratt, H., Coenen, F., Broadbent, D. M., et al. (2016). Convolutional Neural Networks for Diabetic Retinopathy. *Procedia Computer Science*, 90, 200-205. https://doi.org/10.1016/j.procs.2016.07.014
[62] Miotto, R., Li, L., & Dudley, J. T. (2016). *Deep Learning to Predict Patient Future Diseases from the Electronic Health Records*. Springer International Publishing. https://doi.org/10.1007/978-3-319-30671-1_66
[63] Miotto, R., Li, L., Kidd, B. A., et al. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci Rep*, 6, 26094. https://doi.org/10.1038/srep26094
[64] Nie, L., Wang, M., Zhang, L., et al. (2015). Disease Inference from Health-Related Questions via Sparse Deep Learning. *IEEE Transactions on Knowledge & Data Engineering, 27*(8), 1-1. https://doi.org/10.1109/TKDE.2015.2399298
[65] Kamble, M. T. P. & Patil, D. S. T. (2016). Diabetes Detection using Deep Learning Approach. *International Journal for Innovative Research in Science & Technology, 2*(12), 2349-6010.
[66] Zhang, D., Jin, D., Gong, Y., et al. (2015). Research of alarm correlations based on static defectdetection. *Tehnicki vjesnik, 22*(2), 311-318. https://doi.org/10.17559/TV-20150317102804
[67] Zhang, D. (2017). High-speed Train Control System Big Data Analysis Based on Fuzzy RDFModel and Uncertain Reasoning. *International Journal of Computers, Communications & Control, 12*(4). https://doi.org/10.15837/ijccc.2017.4.2914

Contact information:

Yun-lei SUN, Lecturer (Corresponding author)
College of Computer & Communication Engineering,
China University of Petroleum (East China), Qingdao, 266580, China
E-mail: sunyunlei@upc.edu.cn

Da-lin ZHANG, Associate Professor
National Research Center of Railway Safety Assessment,
Beijing Jiaotong University, Beijing, 100044, China
E-mail: dalin@bjtu.edu.cn