Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-020-02239-w

ORIGINAL RESEARCH

Intelligent mining algorithm for complex medical data based on deep learning

Xiaofeng Li[1] · **Dong Li**[2] · **Yuanbei Deng**[3] · **Jinming Xing**[4]

Received: 13 December 2019 / Accepted: 17 June 2020

Abstract
In order to address the problems of low precision, long time-consuming and low recall rate in mining complex attribute medical data in medical information, an intelligent mining algorithm for complex attribute medical data based on deep learning is proposed. Discretized medical data with complex attributes and converted it into a data type suitable for deep learning research, the convolutional neural network is used to analyze the association mapping relationship between complex attribute medical data sets and extract association rules of data. According to the degree of association between complex attribute medical data sets in multi-dimensional subspace to realize the effective mining of complex attribute medical data. The results show that the proposed algorithm takes less time and can extract association rules accurately, the data priority control efficiency is higher, the data mining accuracy is better, and the data mining recall rate is much higher than other methods, which verifies the feasibility of the proposed algorithm.

Keywords Deep learning · Complex attributes · Medical data · Convolutional neural network · Association rules

1 Introduction

With the rapid development of computer information technology in the current society, the construction of hospital informatization is becoming more and more mature. The initial hospital information system has gradually upgraded to the electronic medical record information system. The extensive use of this system has produced a large number of medical data. How to mine complex medical data intelligently is become an urgent problem in the medical industry (Ting et al. 2017; Hui et al. 2018). Nowadays, many data mining technologies have been widely used in banking, commerce, industry and other fields, and achieved remarkable results. Many experts and scholars have made some progress in the research of medical data mining technology, but the field of medical information is facing difficulties such as fewer talents and poor academic skills, more difficulty in complex attributes of medical data mining, and a wide range of coverage, which greatly limits the application of mining technology in medical data (Pazhoumand 2018; Miholca et al. 2018). Therefore, it has important practical and theoretical significance for the research of data mining technology and its application in the medical field.

With the rapid development of computer technology, the application of intelligent technologies such as deep learning and data mining in the medical field has attracted the attention of relevant scholars. Zhang and Wang (2017) Conducted in-depth research on medical text entity recognition methods and achieved good research results, verifying the superiority of computer technology; Fisher et al. (2018) applied mining intelligent technology well, and solved the problem of data loss based on mining technology; Altman (2017) pointed out that machine intelligence has powerful algorithmic capabilities, can find hidden patterns in data, acquires features of things based on data association patterns, and solves complex medical dataset calculation problems; Ward et al. (2018) established a Python-based software platform to facilitate data-driven method analysis, using data mining and statistical learning methods, by using data mining

✉ Xiaofeng Li
mberse@126.com

[1] Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

[2] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

[3] College of Mathematics and Econometrics, Hunan University, Changsha 410082, China

[4] School of Physical Education, Northeast Normal University, Changchun 130024, China

Published online: 26 June 2020

and statistical learning methods, the data set is analyzed and a prediction model is established, which gives full play to the superior performance of data mining technology; Chon et al. (2018) pointed out that the existing sequential frequent itemsets mining methods have certain limitations. A fast frequent itemset mining method based on GPU is proposed. By making full use of the computing power of the GPU, a very fast performance is achieved, which is suitable for large-scale data.

On the basis of the above existing research content, this paper is based on deep learning to analyze and study medical data with complex attributes in order to obtain more accurate data mining results. This paper proposes an intelligent mining algorithm for complex attribute medical data based on deep learning. It performs discrete processing on complex attribute medical data to obtain applicable data types, and then classifies medical data with different attributes according to medical data association rules, and finally completes complex attribute medical data mining. The results show that the performance of the proposed algorithm is superior, which has helped the development of the medical field. The main contributions of this paper are as follows:

1. Discretized medical data with complex attributes, and the processed data is more suitable for subsequent research on convolutional neural networks;
2. The association rules of medical data with complex attributes are extracted, and the associations between data sets are analyzed, which lays the foundation for data mining research;
3. Before data mining, data classification was performed first, which effectively improved mining accuracy;
4. The performance of the proposed algorithm is verified by multiple groups of comparative experiments. The index content is rich and the results have certain reliability.

2 Related work

Many foreign scholars have studied medical data mining. Miholca and Onica (2017) proposed the method of extracting classification rules from medical diagnostic data by using artificial neural network to train, prune and extract rules. Jia et al. (2017) proposed Apriori algorithm to solve data mining problems, and improved the algorithm by adding constraints to potential medical data. Mining in order to complete disease diagnosis quickly and effectively, and the performance of the algorithm is verified by experiments, the overall results are satisfactory; Krishnamoorthy et al. (2017) proposed a method of mining fuzzy association rules, which can be applied to various data sets, and has a certain degree of universality; Motlagh et al. (2017) advocated using genetic algorithm to analyze data mining, and proposed a method suitable for data mining. The genetic algorithm research method of hiding quantity of sensitive rules in the original database is proved to be more applicable by experimental analysis.

At present, there are many experts and scholars in China who put forward the algorithm of medical data mining. For example, Jia et al. (2017) put forward the method of medical data mining aiming at Apriori algorithm, collect relevant mobile medical data through digital terminals, and analyze the influencing factors of data acquisition and transmission, so as to ensure the normal operation of mobile medical system; Lv et al. (2017) studied the data mining method based on genetic algorithm, obtained the target information effectively in the huge medical database, improved the genetic algorithm with K-means algorithm, and then obtained the results of medical data mining. Ye et al. (2010) proposed the attribute reduction method of conditional information entropy based on SQL language for a large number of complex data in the clinical medical record database. Miscellaneous data is simplified and cleaned, and a rough set theory model based on information theory is established to mine medical data effectively.

The above domestic and foreign research has achieved good results, but the mining effect of medical data with complex attributes is not good. There are some problems such as long mining time, low precision and low recall rate. In view of this situation, this paper proposes an intelligent mining algorithm for complex medical data based on deep learning, different from other algorithms, this paper has effectively classified medical data by extracting the association degree between medical datasets with complex attributes. Based on the classification results, the final data mining is achieved by related processing. The results show that the proposed algorithm takes a short time, the accuracy of association rule extraction is high, the data priority control efficiency is high, and the average recall rate can be as high as about 75%, which has good data mining results.

3 Intelligent mining algorithms for medical data with complex attributes

3.1 Discretization of mixed attribute medical data

In the process of medical data mining with mixed attributes, firstly, medical data is processed, expressed and transformed. That is, medical data with mixed attributes is pretreated, and the data format after pretreatment is transformed into a form acceptable to the deep learning mining algorithm (Qian et al. 2018; Sun et al. 2018). For the deep learning mining algorithm, Discretized data forms are the most acceptable form. Discretization of numerical data is beneficial to speed up the progress of mining and improve the effect of data mining. Therefore, it is necessary to convert mixed attribute medical data into discrete form (Yanhong 2018; Wang et al. 2017).

In the existing data preprocessing methods, it is easy to ignore the discretization of the data, which leads to poor preprocessing results. This paper processes the mixed attribute medical data and converts it into discrete data, as follows:

1. Set decision table A containing attribute data of mixed attribute medical data and customize the minimum attribute importance threshold ϖ;
2. Observe the continuous attributes in A, where, $s \in S$, Calculate the importance of elements in a continuous attribute set S, if the calculation result is greater than the threshold ϖ, it is marked as an important attribute(Yi et al. 2017; Lan et al. 2020), otherwise it is deleted;
3. Set all the important attributes, sort them according to the numerical value, divide the equivalence classes according to the decision attributes, and construct the numerical string $\left(s_1, s_2\right)$ of the continuous attribute s, if the following formula is satisfied:

   $$\max\left(Q\left(s_1\right), Q\left(s_2\right)\right) \geq \min\left(U\left(s_1\right), U\left(s_2\right)\right) \tag{1}$$

   The data result set *Endset* can be expressed as:

   $$Endset = \max\left(Q\left(s_1\right), Q\left(s_2\right)\right) + \min\left(U\left(s_1\right), U\left(s_2\right)\right) \tag{2}$$

   If $\max\left(Q\left(s_1\right), Q\left(s_2\right)\right) \min\left(U\left(s_1\right), U\left(s_2\right)\right)$, then the data result set *Endset* can be expressed as:

   $$Endset = \frac{\max\left(Q\left(s_1\right), Q\left(s_2\right)\right) + \min\left(U\left(s_1\right), U\left(s_2\right)\right)}{2} \tag{3}$$

   where $Q\left(s_i\right)$ is the attribute value. $U\left(s_i\right)$ is the attribute boundary value.
4. According to the obtained data result set *Endset*, the discretization of the decision table A can be completed.

By combining the above steps, the discretization processing of mixed attribute medical data can be realized, which lays the foundation for efficient data mining.

3.2 Medical data association rules extraction

The difficulty of extracting association rules from medical data makes it difficult to connect input and output tuples. Therefore, it is necessary to extract association rules between medical data. Association rule is interesting relationships between different variables (Nguyen et al. 2019; Song et al. 2017). Convolutional neural network is one of the representative algorithms of deep learning. In the paper, convolutional neural network is used to extract association rules from medical data.

The relationship between input and output tuples of medical data is complex, and it is difficult to find, mainly because of the difficulty of extracting rules. The main reasons are as follows:

1. Assuming that there are n input layer nodes in the data network, there will be 2^n different input modes. No matter what the value of n is, there will be more complex association rules.
2. The input tuple determines the activation value of the hidden layer nodes, but the activation value can be a continuous value. Under this condition, it is difficult to determine the association rules between the output layer node values and the hidden layer node activation values (Oskouei et al. 2017; Hua et al. 2018).

In order to improve the efficiency of medical data mining with complex attributes, the convolution neural network is used to extract the association rules of medical data, analyze the association mapping (association mapping is a one-to-one, one-to-many, and many-to-many relationship between objects, different objects can form a kind of mapping) between medical data, and obtain all the association rules of medical data sets with complex attributes by association mapping. At the same time, the state of medical data mining with complex attributes is judged by probability estimation method. The mining factor and relative error are introduced to improve the precision of medical data mining with complex attributes (Roy et al. 2017; Sheng et al. 2016).

According to the above complex attributes, the structure diagram of medical data mining is analyzed concretely. Choose $G = (V, E)$ to represent the topology map, where V represents the hierarchical structural elements constituting the medical information network; E represents the edges connecting the hierarchical structural elements. In $V = (V_1, V_2, \ldots, V_n)$, V_i represents complex attribute medical data set, $V_i = (x_{1i}, x_{2i}, \ldots, x_{ni})$, where x_{ji} represents an effective medical data set of complex attribute medical data set.

Suppose that the association degree(association degree is the similarity between data sets) between V_i and V_k of complex attribute medical data sets is represented by association attribute group $(\alpha_{ik}, \beta_{ik}, \theta_{ik})$, as its name implies, the association attribute group is the set of association values between the medical datasets V_i and V_k of complex attributes. where α_{ik} represents the size association between complex attribute medical data sets, β_{ik} represents the semantic association between complex attribute medical data sets, and θ_{ik} represents the category association between complex attribute medical data sets.

The association mapping relationship between complex attribute medical data sets is:

1. The association attribute group $(\alpha_{ik}, \beta_{ik}, \theta_{ik})$ between the complex attribute medical data set V_i and the data set V_k can represent the association degree of random generated data for these two complex attribute medical data sets (Takumi et al. 2018; Nancy et al. 2017).
2. The association coefficient matrix is used to represent the association attribute group, and this matrix can represent the average association degree of all medical data in the two complex attribute medical data sets (Palaniappan and Awang 2018; Wang and Han 2018), which is expressed as show in Eq. (4).

$$K_1\begin{pmatrix}\alpha_{ik}\\ \beta_{ik}\\ \theta_{ik}\end{pmatrix} = \begin{pmatrix}a_{i1} & \cdots & a_{1k}\\ \vdots & \cdots & \vdots\\ a_{1k} & \cdots & a_{i1}\end{pmatrix}\begin{pmatrix}\beta_{i1} & \cdots & \beta_{1k}\\ \vdots & \cdots & \vdots\\ \beta_{1k} & \cdots & \beta_{i1}\end{pmatrix}\begin{pmatrix}\theta_{i1} & \cdots & \theta_{1k}\\ \vdots & \cdots & \vdots\\ \theta_{1k} & \cdots & \theta_{i1}\end{pmatrix} \tag{4}$$

   where K_1 is the correlation coefficient.
3. There are some differences among medical data sets with complex attributes besides correlations. Using the following Eq. (5), the difference coefficient matrix of medical data sets with complex attributes is given.

$$K_2\begin{pmatrix}\frac{1}{\alpha_{ik}}\\ \frac{1}{\beta_{ik}}\\ \frac{1}{\theta_{ik}}\end{pmatrix} = \begin{pmatrix}\frac{1}{a_{i1}} & \cdots & \frac{1}{a_{1k}}\\ \vdots & \cdots & \vdots\\ \frac{1}{a_{1k}} & \cdots & \frac{1}{a_{i1}}\end{pmatrix}\begin{pmatrix}\frac{1}{\beta_{i1}} & \cdots & \frac{1}{\beta_{1k}}\\ \vdots & \cdots & \vdots\\ \frac{1}{\beta_{1k}} & \cdots & \frac{1}{\beta_{i1}}\end{pmatrix}\begin{pmatrix}\frac{1}{\theta_{i1}} & \cdots & \frac{1}{\theta_{1k}}\\ \vdots & \cdots & \vdots\\ \frac{1}{\theta_{1k}} & \cdots & \frac{1}{\theta_{i1}}\end{pmatrix} \tag{5}$$

   where K_2 is the coefficient of difference.

Through the correlation coefficient matrix and difference coefficient matrix of the complex attribute data set, the correlation mapping between the complex attribute data set V_i and the data set V_k is as shown in Eq. (6).

$$V_i \to \frac{K_1}{K_2}V_2, \ldots, \to \frac{K_1}{K_2}V_k \tag{6}$$

After getting the association mapping between the complex attribute medical data set V_i and the data set V_k the association rules of the complex attribute medical data set are obtained by using the correlation matrix, so as to find the complex attribute data set V_i and V_k from the medical data set. The comprehensive complex attribute data set $f(V_i, V_k)$ can be expressed as Eq. (7).

$$f(V_i, V_k) = \begin{Bmatrix}V_1 \cdots V_i\\ \vdots \quad \cdots \quad \vdots\\ V_i \cdots V_1\end{Bmatrix}\begin{Bmatrix}\alpha_{ik}\\ \beta_{ik}\\ \theta_{ik}\end{Bmatrix} + \begin{Bmatrix}V_1 \cdots V_i\\ \vdots \quad \cdots \quad \vdots\\ V_i \cdots V_1\end{Bmatrix}\begin{Bmatrix}\frac{1}{\alpha_{ik}}\\ \frac{1}{\beta_{ik}}\\ \frac{1}{\theta_{ik}}\end{Bmatrix} \tag{7}$$

After finding out the V_i and V_k of medical data sets with complex attributes from medical data sets, they can be classified according to the association mapping between the two medical data sets.

The frequency of medical data mining with complex attributes is calculated by probability estimation method, and its expression is given by Eq. (8):

$$P(V_i) = \frac{\sum_{i=1}^{k} V_i^2}{\sum_{i=1}^{k} V_i^2} f(V_i, V_k) \tag{8}$$

where $P(V_i)$ is the frequency function of medical data mining with complex attributes.

In order to improve the precision of medical data mining with complex attributes, two parameters, mining factors and relative error are introduced in this paper to calculate the data mining frequency. The medical data mining frequency of complex attributes can be expressed as shown in Eq. (9).

$$P(V_i) = \lambda^{-1}\zeta \frac{\sum_{i=1}^{k} V_i^2}{\sum_{i=1}^{k} V_i^2} f(V_i, V_k) \tag{9}$$

where λ is the factor of medical data mining with complex attributes, and its value is (0, 1). Getting the appropriate λ value can maximize the frequency of medical data mining with complex attributes. ζ is the relative error between the expected and actual probability of medical data mining with complex attributes.

3.3 Intelligent mining of medical data with complex attributes

Assuming that the sample of complex attributes medical data set is distributed in a multi-dimensional subspace, the greater the association degree between two random complex attributes medical data set samples in a subspace, the stronger the correlation of data set samples; the smaller the association degree, the weaker the correlation of data set samples (Shameer et al. 2018; Afzali and Mohammadi 2018). When classifying medical data sets with complex attributes in the same space, the association degree of root data sets is needed to determine the mining rules for medical data (Bratić et al. 2018; Ramasamy and Nirmala 2017). In the multi-dimensional subspace, the analysis is performed from two aspects of the associated subspace and the unassociated subspace to complete the classification of the complex attribute medical data set.

Suppose that the dimension of subspace is represented by d. Firstly, medical data sets with different complex attributes distributed in association subspace are mined, and the matrix representation of subspace is given by using the following Eq. (10):

$$M = \begin{Bmatrix} M^1 \\ \vdots \\ M^d \end{Bmatrix} \quad d \le n \tag{10}$$

Assuming that two complex attribute medical datasets V_i and V_k are in unassociated subspace M^i and M^d respectively, and the Euclidean distance between the two spaces is expressed by $D(i,k)$, the Euclidean distance between the two complex attribute medical datasets is expressed by $d(i,k)$. Using the following formula, the mining expressions of two complex attributes medical data sets in unassociated subspace are shown in Eq. (11).

$$W_1(M^i, M^k) = \frac{\sigma}{2} \begin{Bmatrix} M^1 \\ \vdots \\ M^d \end{Bmatrix} P(V_i)P(V_k) \times \log_2 \sqrt{D(i,k)^2 + d(i,k)^2} \tag{11}$$

where σ represents the mining factors in unassociated subspace, $P(V_i)$ and $P(V_k)$ represent the mining frequencies of V_i and V_k of complex attribute medical data sets respectively.

To mine medical data sets with different complex attributes in associated subspace, it is necessary to classify the medical data sets with different complex attributes according to the association degree of self-examination (Zhang and Li 2017; Ghamisi et al. 2016). K_1 and K_2 are calculated according to Eqs. (1) and (2). On this basis, the correlation factors of V_i and V_k of medical data sets with complex attributes in the associated subspace are obtained in Eq. (12).

$$g(i,k) = \sqrt{\frac{K_1}{K_2} \begin{Bmatrix} x_{1i} \cdots x_{ki} \\ \vdots \quad \cdots \quad \vdots \\ x_{ki} \cdots x_{1i} \end{Bmatrix} d(i,k) - \ln 2 \left(\frac{K_1}{K_2} \begin{Bmatrix} x_{1i} \cdots x_{ki} \\ \vdots \quad \cdots \quad \vdots \\ x_{ki} \cdots x_{1i} \end{Bmatrix} \right)} \tag{12}$$

Based on Eqs. (11) and (12), the mining Equation of medical data sets with complex attributes in the associated subspace can be obtained, which is expressed as shown in Eq. (13).

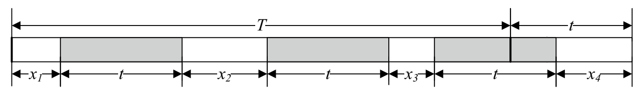


Fig. 1 Structure model of priority list control

$$W(V^i, V^k) = (P(V_i) - P(V_k))g(i,k)d(i,k) \times \frac{e^{g(i,k)}}{\begin{Bmatrix} x_{1i} \cdots x_{ki} \\ \vdots \quad \cdots \quad \vdots \\ x_{ki} \cdots x_{1i} \end{Bmatrix}} \tag{13}$$

Assuming that the fixed threshold of association degree between complex attribute medical data sets is expressed by $T(V)$ in space M^i, when the correlation factor $g(i,k)$ between complex attribute medical data sets is greater than $T(V)$, the two complex attribute medical data sets have strong correlation. Using Eq. (14) to represent the classification result $f_{M^i}(V_i)$ of the medical dataset with complex attributes as

$$f_{M^i}(V_i) = \begin{Bmatrix} V_1 \cdots V_k \\ \vdots \quad\quad \vdots \\ V_k \cdots V_1 \end{Bmatrix} W(V^i, V^k) - \ln \left(\sqrt{\sum_{i=1}^{k} (P(V_i) - P(V_k)} \right) \tag{14}$$

When the correlation factor $g(i,k)$ between complex attribute medical data sets is less than $T(V)$, the two complex attribute medical data sets have weak correlation. By using Eq. (15) to represent the classification result $f_{M^i}(V_i)$ of the medical dataset with complex attributes as.

$$f_{M^i}(V_i) = \frac{\begin{Bmatrix} V_1 \cdots V_k \\ \vdots \quad\quad \vdots \\ V_k \cdots V_1 \end{Bmatrix} W(V^i, V^k)}{(\pi e^2 - 1)} + \frac{e}{2} \sum_{i=1}^{k} (P(V_i) - P(V_k)) \tag{15}$$

where e is a weak association factor.

According to Eqs. (14) and (15), the classification results of data sets can be obtained. According to the classification results, the priority control of medical data with complex attributes can be processed, and then the intelligent mining results of medical data can be obtained. The intelligent mining algorithm for complex attribute medical data is:

Input: Medical datasets with complex attributes and classification results;

Ouput: Intelligent mining results of medical datasets with complex attributes.

Initialized complex attribute medical data and performed mining analysis, as follows:

1. Design a structural model of priority control for complex attribute medical data mining. The model consists of a cluster head, several cooperative cluster heads, and cluster members. The structural model of priority list control is shown in Fig. 1.
2. Assuming that the test data set of medical data with complex attributes is p_i and the medical data mining task n_j is executed, the corresponding priority attribute $DR(p_i, n_j)$ can be expressed as shown in Eq. (16).

$$DR(p_i, n_j) = rwd_{ik} \times PET(p_i, n_j) \tag{16}$$

   where, $PET(p_i, n_j)$ denotes the time window function needed to execute the medical data mining task n_j, and rwd_{ik} denotes the dynamic trend function of the test data set.
3. Using the priority attributes of complex medical data in the above formula, the input control parameters of medical data can be obtained, and the mining tasks can be balanced by using the time difference of transmission between mining tasks.
4. Assuming Y is the cyclic stack control vector of d dimension list, in order to realize the compactness measurement of clustering nodes and reduce the delay of medical data mining, time series analysis is carried out in mining time windows and scale vectors. Set time slot node, then the average mutual information mapping result $F\{x_n\}_{n-1}^{N}$ of time series $\{x_n\}_{n-1}^{N}$ in the subspace can be expressed as Eq. (17).

$$F\{x_n\}_{n-1}^{N} = Y[x(0), x(1), ..., x(N-1)]^d \tag{17}$$

   The above-mentioned mutual information is a parameter used to measure the correlation between two data sets. The average mutual information is the correlation of the data set obtained after eliminating the uncertainty from the overall point of view. By mapping in the subspace, the average mutual information mapping result can be obtained.
5. The medical data set with complex attributes has autocorrelation. According to the average mutual information mapping result obtained by Eq. (17), the priority list control model is obtained, which is described as shown in Eq. (18).

$$Y(n) = DR(p_i, n_j) \max\left[F\{x_n\}_{n-1}^{N}\right]^d \tag{18}$$

6. According to the priority list control model, the priority of the data set is sorted, and intelligent mining of medical data is realized according to the sorted result.
7. End

Based on the above analysis steps, the intelligent mining results of medical data can be obtained. The proposed algorithm flow is shown in Fig. 2.

4 Experimental and results

4.1 Experimental environment and data set

The experimental data are all from Kaplan Meier Plotter database (https://kmplot.com/analysis/). Kaplan Meier Plotter database is a common database of mRNA profiling chips

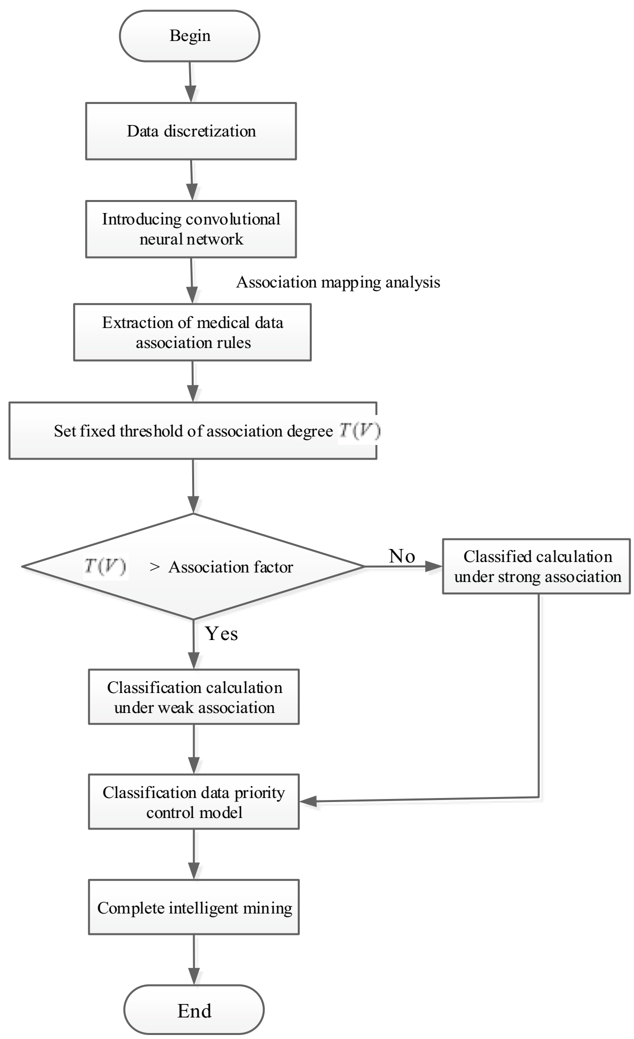


Fig. 2 Intelligent mining algorithm process of medical data with complex attributes

containing five types of cancer and has certain comprehensiveness. In this database, one million cancer medical data sets are selected, and the extracted data sets are simulated and tested by MATLAB7.0 software. The operating system is Windows 10, which is equipped with Visual C++ 6.0 compiler environment. The experimental analysis is carried out in this environment. The experimental data set can be divided into five categories, which are specifically described in Table 1.

4.2 Experimental index

The experimental comparative indicators are as follows:

1. Data mining frequency and time-consuming analysis
2. Association rule extraction accuracy
3. Data priority control efficiency
4. Data mining accuracy
5. Data mining recall rate

4.3 Results

4.3.1 Data mining frequency and time-consuming analysis

Data mining frequency $P(V_i)$ is an important parameter to complete association rule extraction. In order to verify the mining effect of the proposed algorithm, the data mining frequency $P(V_i)$ obtained by the formula (9) is used as an experimental index. The data mining frequency of the proposed algorithm is analyzed and the mining time is compared and analyzed.

In the process of mapping and analyzing association rules of medical data sets, in order to test the time-consuming degree of different algorithms for medical data mining with complex attributes, the mining factor is introduced to analyze and the frequency of data mining is solved. According to the method of text calculation, under the condition of different mining factors (mining factor is constant), the broken line graph of data mining frequency is obtained, as shown in Fig. 3.

From the analysis of Fig. 3, we can see that the frequency of data mining is at a high level under different mining factors, and the state of data mining is stable. From this, we can conclude that the mining efficiency of this algorithm is higher and the time consuming is lower when mining complex attribute medical data, which can save a lot of data mining time.

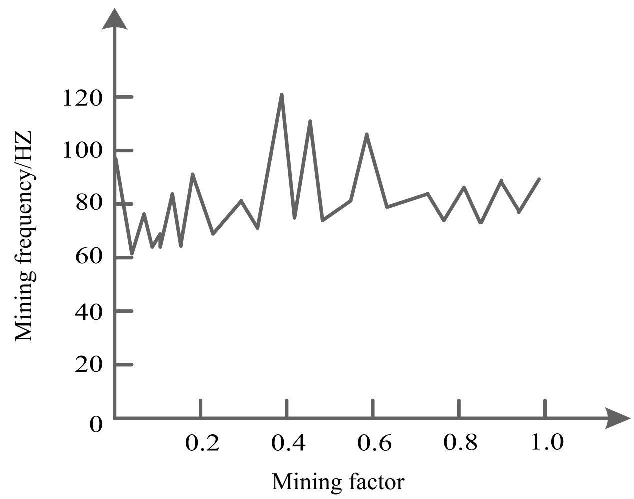


Fig. 3 Medical data mining frequency breakdown

In order to further verify the data mining efficiency of the proposed algorithm, the algorithm and literature Fisher et al. (2018), literature Altman (2017), literature Krishnamoorthy et al. (2017), literature Motlagh et al. (2017) and literature Jia et al. (2017) are compared for medical data mining time. The test results are shown in Fig. 4.

Analysis of Fig. 4 shows that the trends and amplitudes of the algorithms in literature Fisher et al. (2018) and literature Krishnamoorthy et al. (2017) are relatively consistent. As the amount of data increases, the mining time

Table 1 Experimental dataset

Dataset name	Quantity/ten thousand
Mammary cancer	20
Ovarian cancer	20
lung cancer	20
gastric cancer	20
Liver cancer	20

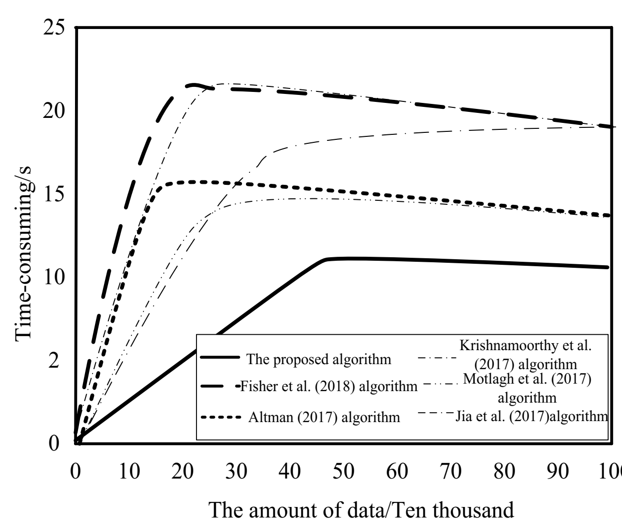


Fig. 4 Comparison of data mining time-consuming

is stable at about 20 s, and the literature Jia et al. (2017) also approaches 20 s, but the overall algorithm is lower than that in literature Fisher et al. (2018) and literature Krishnamoorthy et al. (2017). The data mining time of the algorithm in literature Altman (2017) and literature Motlagh et al. (2017) eventually approaches 15 s, which is lower than the other three algorithms, but much higher than the proposed algorithm. The maximum data mining time of the proposed algorithm is 10 s. The medical data mining time of this algorithm is obviously lower than that of literature Fisher et al. (2018) and literature Altman (2017), because the mining factor is introduced into the proposed algorithm, which makes the medical data. The efficiency of data mining has been greatly improved and the time-consuming of data mining has been shortened.

4.3.2 Comparison of association rule extraction accuracy

According to Sect. 3.2, in the process of mining complex attribute medical data, it is necessary to extract the association rules of the data to facilitate further research. The accuracy of association rule extraction will affect the mining effect of medical data. In order to verify the mining effect of the proposed algorithm, the accuracy of association rule extraction is used as an index, and the proposed algorithm is compared with the algorithms in literature Fisher et al. (2018), literature Altman (2017), literature Krishnamoorthy et al. (2017), literature Motlagh et al. (2017) and literature Jia et al. (2017). The results of the association rule extraction accuracy test are shown in Fig. 5.

According to Fig. 5, with the increase of experimental data, the accuracy of association rules extraction is increasing, and the rate of improvement is larger.

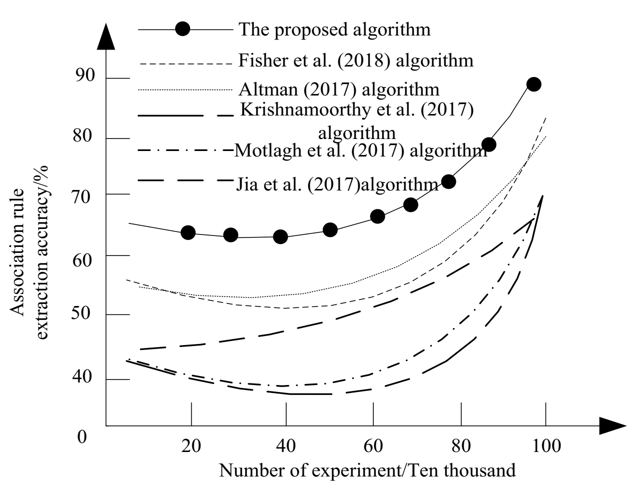


Fig. 5 Comparison of association rule extraction accuracy

Among them, the accuracy of association rules extraction obtained by data mining algorithm in this paper is higher. When the experimental data amount is one million, the accuracy of association rules extraction is as high as 90%. The extraction accuracy of the association rules in literature Fisher et al. (2018) and literature Altman (2017) is also high, up to 80%. Followed by the literature Jia et al. (2017) algorithm, the accuracy rate is in the range of 45–70%, although the accuracy of literature Krishnamoorthy et al. (2017) and literature Motlagh et al. (2017) can reach up to 70%. However, it can be clearly seen that the accuracy of these two algorithms is lower than the other algorithms, and the results are not good. According to the above analysis, it can be seen that the accuracy of the proposed algorithm extraction is high, which can well match the output value of medical data. This is because the proposed algorithm uses the convolutional neural network to extract the association rules of medical data, and uses the association coefficient matrix and the difference coefficient matrix to obtain the association mapping of the data set. The association mapping relationship is used to obtain all the association rules of the complex attribute medical data set, thereby improving the accuracy of association rule extraction, and the medical data mining effect is better.

4.3.3 Comparison of data priority control efficiency

The accuracy of association rules extraction will affect the accuracy of medical data mining. At the same time, the priority control efficiency of medical data will affect the efficiency of data mining. Therefore, before comparing the frequency and time-consuming of data mining, the data priority control efficiency of different algorithms is compared. Priority control processing of medical data with complex attributes is an important step to promote intelligent mining of medical data. The efficiency of data priority control of this algorithm is compared with that of literature Fisher et al. (2018), literature Altman (2017), literature Miholca and Onica (2017), literature Jia et al. (2017) and literature Ye et al. (2010). The test results are shown in Fig. 6.

According to Fig. 6, in the initial test stage, except for the algorithm in literature Altman (2017), the efficiency of priority control of medical data in other algorithms has shown an upward trend. As the number of experiments increases, the efficiency of data priority control tends to be stable and the fluctuation range is small. However, the data priority control efficiency of the proposed algorithm eventually tends to be about 92%. The data priority control efficiency of the algorithm in literature

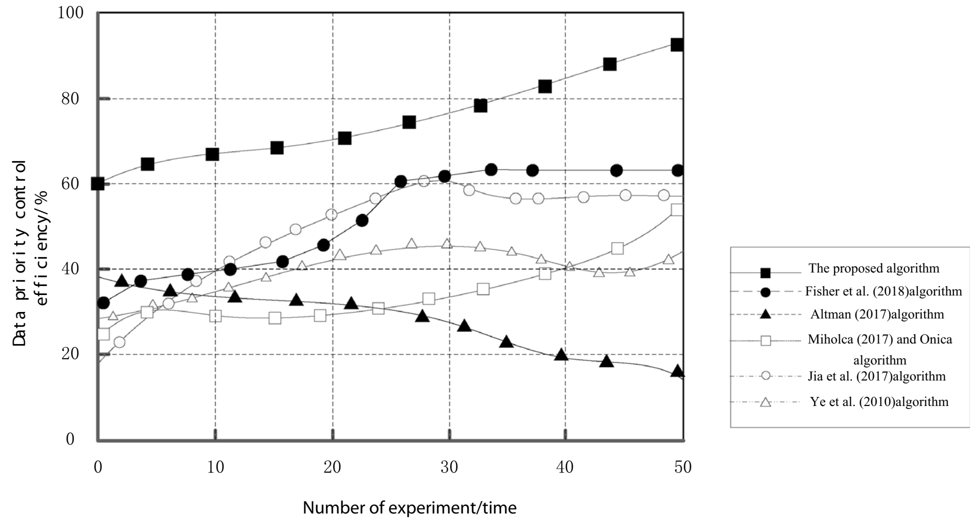


Fig. 6 Comparison of data priority control efficiency

Fisher et al. (2018) finally tends to be about 62%. The data priority control efficiency of the algorithms in literature Miholca and Onica (2017) and literature Jia et al. (2017) finally tends to be about 59%. The literature Jia et al. (2017) algorithm finally tends to be about 45%. The gap is more obvious, before the data priority control, the proposed algorithm classifies the medical data set, and the effect of the priority control of the classified data is obvious. It can be seen that when using the proposed algorithm for data mining, the data processing effect is better, which helps to improve the efficiency of medical data mining.

4.3.4 Comparison of accuracy comparison of data mining

High-precision mining of medical data is an important purpose of the proposed algorithm. The above experiments verify the differences of different algorithms in the accuracy of active value calculation, the efficiency of data priority control and the time-consuming of data mining are verified. On this basis, the accuracy of different algorithms in medical data mining is compared and analyzed. Taking one million medical sampling data as the object to test, the test interval is set to 10 s, and the test duration is set to 6 h. The mining accuracy of the proposed algorithm is compared with the algorithms in literature Miholca and Onica (2017), literature Jia et al. (2017), literature Krishnamoorthy et al. (2017), literature Motlagh et al. (2017) and literature Ye et al. (2010). The test results are shown in Table 2.

As can be seen from the analysis Table 2, the proposed algorithm establishes a priority control model and implements intelligent data mining according to the priority of the data set, which greatly improves the accuracy of medical data mining. The overall accuracy is over 90.0% and the highest is 98.6%.

The medical data mining accuracy of the algorithm in literature Miholca and Onica (2017) is up to 75.3%, the medical data mining accuracy in the algorithm in literature Jia et al. (2017) is up to 65.2%, the algorithm in literature Krishnamoorthy et al. (2017) is not more than 79.2%, and the algorithm in literature Motlagh et al. (2017) is not more than 76.2%, The literature Ye et al. (2010) algorithm is not more than 66.3%. The mining accuracy of these five algorithms is obviously lower than the proposed algorithm, which verifies the superior performance of the proposed

Table 2 Comparison of accuracy of medical data mining

Sampled data/ten thousand	The proposed algorithm/%	Literature Miholca and Onica (2017) algorithm /%	Literature Jia et al. (2017) algorithm /%	Literature Krishnamoorthy et al. (2017) algorithm /%	Literature Motlagh et al. (2017) algorithm /%	Literature Ye et al. (2010) algorithm /%
20	90.2	74.1	58.3	69.6	76.2	56.3
40	94.8	69.5	60.3	66.5	71.0	52.1
60	98.6	68.6	58.6	55.2	70.2	61.2
80	97.1	75.3	65.2	79.2	65.2	63.4
100	91.0	71.2	59.2	63.2	63.0	66.3

algorithm and can achieve high-precision medical data mining of complex attributes.

4.3.5 Comparison of recall rate comparison of data mining

Recall rate is one of the evaluation indicators of the database. When mining a database, the recall rate of the mining results also reflects the effect of data mining. Therefore, this paper uses data mining recall rate as an evaluation index to verify the algorithm. Comparing the proposed algorithm with literature Fisher et al. (2018) and literature Altman (2017), the test results are shown in Fig. 7.

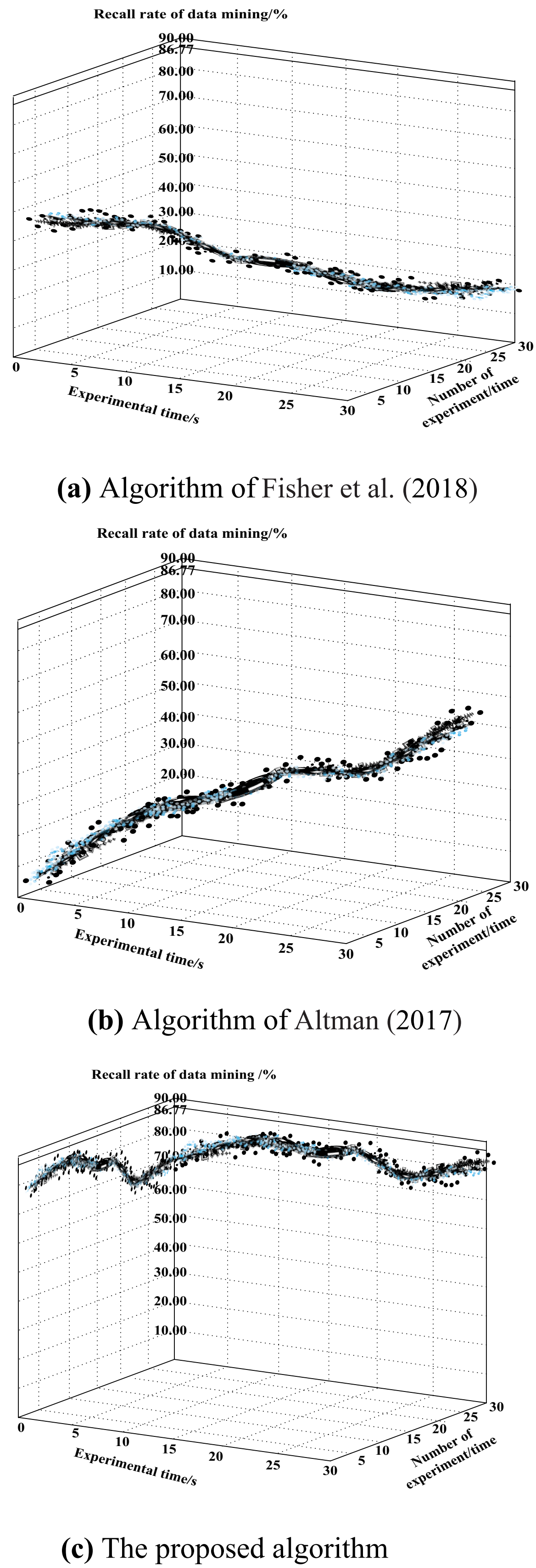
Analysis of Fig. 7 shows that under the condition of certain experimental time and number of experiments, the recall rate of medical data mining with complex attributes obtained by the proposed algorithm is relatively high, and the average recall rate can be as high as about 75%. The data mining recall rate of the algorithm in literature Fisher et al. (2018) is lower than 30% as a whole, and it shows a downward trend. Although the data mining recall rate of the algorithm in literature Altman (2017) shows an upward trend, the highest data mining recall rate is not more than 50%.

According to the results of data analysis, we can see that the proposed algorithm has absolute advantages and strong feasibility. The reason is that before the data mining, the convolutional neural network in deep learning was used to extract the association rules of medical data, which helps to improve the recall rate of data mining.

In summary, this paper chooses several indicators to compare and analyze the traditional algorithm with the proposed algorithm, and it is concluded that the proposed algorithm is superior to the traditional algorithm in data mining efficiency, accuracy and recall rate. It shows that the proposed algorithm can be applied to the mining of complex medical data with practical application.

5 Conclusions

In order to solve the problems of low precision, long time-consuming and low recall rate in traditional data mining methods, this paper proposes an intelligent mining algorithm for complex attribute medical data based on deep learning. It transforms complex attributes medical data, extracts association rules of medical data by convolutional neural network in deep learning, improves the precision of mining complex attributes medical data according to the association mapping relationship between complex attributes medical data sets, and classifies complex attributes medical data sets in multi-dimensional subspace, so as to realize the medical number of complex attributes. The results show that the proposed algorithm has low data



(a) Algorithm of Fisher et al. (2018)

(b) Algorithm of Altman (2017)

(c) The proposed algorithm

Fig. 7 Comparison of recall rate of medical data mining

mining time-consuming, high efficiency and high mining accuracy. It shows that the proposed algorithm can realize the accurate mining of medical data, and can be widely used in the medical field to provide theoretical basis for solving practical problems.

However, there are still some deficiencies in this study, and the research on uncertain data sets needs further analysis. In the next research, we will further analyze and improve the content of deep learning, fully exploit the advantages of neural networks, and apply it to medical data research to obtain more effective results.

Acknowledgements This work was supported by National Natural Science Foundation of China under Grant number 10471144, Ministry of Education Science and Technology Development Center under Grant number 2018A01002, China Postdoctoral Science Foundation under Grant number 2017M610852, Humanities and Social Sciences Foundation of the Ministry of Education under grant number 15YJC890004 and Jilin Provincial Social Science Foundation under Grant number 2016A5.

References

Afzali GA, Mohammadi S (2018) Privacy preserving big data mining: association rule hiding using fuzzy logic approach. IET Inf Secur 12(1):15–24

Altman RB (2017) Artificial intelligence (AI) systems for interpreting complex medical data sets. Clin Pharmacol Ther 101(5):585–586

Bratić B, Kurbalija V, Ivanović M et al (2018) Machine learning for predicting cognitive diseases: methods, data sources and risk factors. J Med Syst 42(12):243–245

Chon K-W, Hwang S-H, Kim M-S (2018) GMiner: A Fast GPU-based frequent itemset mining method for large-scale data. Inf Sci 439:19–38

Fisher PS, James J, Baek J et al (2018) Mining intelligent solution to compensate missing data context of medical IoT devices. Pers Ubiquit Comput 22(1):219–224

Ghamisi P, Chen Y, Zhu XX (2016) A self-improving convolution neural network for the classification of hyperspectral data. IEEE Geosci Remote Sens Lett 13(10):1–5

Hua CP, Fu-Yi Y, Ding-Dar L et al (2018) Data mining the comorbid associations between dementia and various kinds of illnesses using a medicine database. Comput Electr Eng 70(4):12–20

Hui Z, Jing H, Huang G et al (2018) Dynamic optimisation based fuzzy association rule mining method. Int J Mach Learn Cybern 23(2):1–12

Jia K, Hanjing Li, Yuan Y (2017) Application of data mining in mobile health system based on apriori algorithm. J Beijing Univ Technol 43(3):394–401

Krishnamoorthy S, Sadasivam G, Rajalakshmi MS et al (2017) Privacy preserving fuzzy association rule mining in data clusters using particle swarm optimization. Int J Intell Inf Technol 13(2):1–20

Lan K, Wang D, Fong S et al (2020) A survey of data mining and deep learning in bioinformatics. J Med Syst 42:139. https://doi.org/10.1007/s10916-018-1003-9

Lv F, Yang HP et al (2017) Application of data mining technology of genetic algorithm in medical large data. Electron Technol Softw Eng 15(5):203–205

Miholca DL, Onicaş A (2017) Detecting depression from fMRI using relational association rules and artificial neural networks. IEEE Int Conf Intell Comput Commun Process 5(56):89–102

Miholca DL, Czibula G, Czibula IG (2018) A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks. Inf Sci 5(12):441–456

Motlagh FN, Sajedi H (2017) MOSAR: a multi-objective strategy for hiding sensitive association rules using genetic algorithm. Appl Artif Intell 30(9):823–843

Nancy JY, Khanna NH, Kannan A (2017) A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals. Expert Syst Appl 78:283–300

Nguyen G, Dlugolinsky S, Bobák M et al (2019) Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev 5(20):1–48

Oskouei RJ, Kor NM, Maleki SA (2017) Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges. Am J Cancer Res 7(3):610–627

Palaniappan S, Awang R (2018) Intelligent heart disease prediction system using data mining techniques. Int J Nanomed 13(8):121–124

Pazhoumand DH (2018) Fuzzy association rule mining for recognising daily activities using kinetic sensors and a single power meter. J Ambient Intell Humaniz Comput 9:1497–1515

Qian M, Fei H, Qi H (2018) Deep learning for intelligent wireless networks: a comprehensive survey. IEEE Commun Surv Tutor 20(4):2595–2621

Ramasamy S, Nirmala K (2017) Disease prediction in data mining using association rule mining and keyword based clustering algorithms. Int J Comput Appl 27(2):1–8

Roy CA, Cheng T, Phan L et al (2017) Supporting precision medicine by data mining across multi-disciplines: an integrative approach for generating comprehensive linkages between single nucleotide variants (SNVs) and drug-binding sites. Bioinformatics 33(11):1621–1629

Shameer K, Perez-Rodriguez MM, Bachar R et al (2018) Pharmacological risk factors associated with hospital readmission rates in a psychiatric cohort identified using prescriptome data mining. BMC Med Inform Decis Mak 18:79. https://doi.org/10.1186/s12911-018-0653-3

Sheng Y, Zhao Z, Liao Y (2016) Coarse-grained distributed in-depth learning for personalized data mining. Netw New Media Technol 5(6):1–6

Song J, Hua X, Yan F (2017) Fast association rule mining algorithm for network attack data. J Discrete Math Sci Cryptogr 20(6):1465–1469

Sun J, Yan C, Wen J (2018) Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. IEEE Trans Instrum Meas 67(1):185–195

Takumi K, Katsutoshi T, Hideyuki H et al (2018) A data mining-based prognostic algorithm for NAFLD-related hepatoma patients: a Nationwide Study by the Japan Study Group of NAFLD. Sci Rep 8(1):10434–10435

Ting CK, Wang TC, Liaw RT et al (2017) Genetic algorithm with a structure-based representation for genetic-fuzzy data mining. Soft Comput 21(11):2871–2882

Wang C, Han D (2018) Image classification of adaptive control deep learning and knowledge mining. J Shenyang Univ Technol 40(3):97–102

Wang Z, Li L, Glicksberg BS et al (2017) Predicting age by mining electronic medical records with deep learning characterizes differences between chronological and physiological age. J Biomed Inform 76(C):59–68

Ward L, Dunn A, Faghaninia A et al (2018) Matminer: an open source toolkit for materials data mining. Comput Mater Sci 152:60–69

Yanhong L (2018) Fast mining method for low-dimensional redundant clustering data under attribute constraints. Sci Technol Eng 18(9):107–111

Ye M, Wu C, Hu X (2010) Research and application of medical data mining based on rough set. Comput Eng Appl 46(21):232–234
Yi L, Jianyi K, Xingdong W et al (2017) Fuzzy classification of product defect data association rules based on attribute differences. J Wuhan Univ Sci Technol 40(1):49–54
Zhang Yu, Li E (2017) New advances in data mining and integration technology of electronic health records. China Digit Med 12(9):41–44
Zhang F, Wang M (2017) Medical text entities recognition method base on deep learning. Comput Technol Autom 36(1):123–127