



# The Role and Mechanism of the Diabetes Control Based on the Association Rule Apriori Algorithm

Shuqi Yang<sup>(✉)</sup>

Xishuangbanna Vocational and Technical College, Jinghong, Yunnan, China  
yangshuqi456@163.com

**Abstract.** This paper is based on the Apriori algorithm to obtain more efficient data mining methods, analysis of *Eucommia Leaf* prevention and treatment of diabetes and mechanism. The classic Apriori algorithm of association rule mining is used to analyze the medical data with the characteristics of privacy, polymorphism, incompleteness, timeliness and redundancy. Firstly, Apriori algorithm is used to find the frequent itemsets of data in the database, and then strong association rules are generated according to the frequent itemsets to find useful association relationships or patterns between itemsets in massive data. The final purpose is to analyze the application of association rules mining in clinical disease monitoring, evaluation of drug treatment effect and prevention of diabetes by *Eucommia ulmoides* leaves.

**Keywords:** Association rules mining · Apriori algorithm · Disease prevention · Data correlation

## 1 Introduction

Diabetes mellitus is a common endocrine and metabolic disease, which seriously affects patients' physical and mental health. Traditional Chinese medicine for diabetes has been a long-standing treatment. *Eucommia ulmoides* leaves have been used in the treatment of diabetes in our country. The experiments of relevant scholars have also confirmed that *Eucommia ulmoides* leaves have the effects of reducing blood fat and blood sugar. *Eucommia ulmoides* polysaccharides mainly improve the immune response ability and immune ability by improving the body's immune response ability. Related studies have shown that the aqueous extract of *Eucommia ulmoides* leaves can reduce the level of oxidative stress in diabetic rats, improve the ability of free radical scavenging in rats, inhibit the formation of lipid peroxide, and enhance the antioxidant capacity. Separation and purification is the basis of the study on the activity and structure of polysaccharides, and it is also an important step in the preparation of polysaccharides.

The generation, collection and storage capacity of scientific data and medical data has been greatly improved, which makes the amount of information in hospital database continue to expand. However, at present, the application of database in hospital is not

sufficient, and the integration and analysis of data is lack. This topic is to introduce the new technology of data mining into the analysis of the data of prevention and treatment of diabetes, and find the internal correlation in the data, so as to provide scientific decision-making basis for the diagnosis and treatment of diseases in the field of medical and health care, and also find a breakthrough for the standardized management and scientific research of clinical.

Association rule mining is the most mature, active and important research field in data mining, and hospital information system is the main source of hospital data. By using computer technology, database resources and advanced data analysis methods, this paper explores the application of data mining technology in disease prediction, disease diagnosis and other data, so as to save resources and share resources, It has become an important means of scientific research in the field of medicine.

At the same time, as a highly practical, experimental and statistical subject, medicine makes full use of a large amount of information accumulated in clinical medicine and medical research, carries out data mining on disease data, and selects the best association rule algorithm, which has important practical value for improving the quality of medical treatment. This study is based on the association rules Apriori algorithm to explore the effect and mechanism of eucommia leaf in the prevention and treatment of diabetes. This paper consists of the following parts. The first part introduces the relevant background and significance of this paper, the second part is the related work of this paper, and the third part is data analysis. The fourth part is example analysis. The fifth part is conclusion.

## 2 Related Work

We use many method for prevention disease, but the effect is very low, so the Ref [1] proposed phytoncides in the prevention and therapy of blackhead disease and their effect on the turkey immune system, the method solved some issues, but for part disease, the way was not working, so some doctor and scholar proposed if we contact the guidance effect, thus, we can prevent some disease by beta titanium alloy surface, its structure is organized by electron-beam technology [2], we know that some mechanism of  $Al(OH)_3$  is limited, so Lin S et al. used the thermal analysis way to do the prevention [3], it get the better effect. In the effect, we have also been used to alleviate the effect of quinoa, by the experiment results, the author think, the experiment is successful [4].

If we use the food mycotoxins to do the related experiment, we also get the better experiment results, but all setting are only experiment, so if we can set different experiment item and group, the effect is very clear [5]. By TLR4/NLRP3 signalling pathway in rats with nonalcoholic fatty liver disease, the Ref [6] also use thus techniquis to against inflammatory damage and know the effect of shenling baizhu powder is good [6]. Different methods have different effect, so Ref [7] used the different treatment of metabolic syndrome and forecast the effect successfully. For the auricular acupressure, Han R et al. think if the way they proposed can prevent and control the effect with children, then all problems will be solved [8]. Based above ways have some defects, so Ref [9] does thus a review for these ayurvedic plants. However some approach assumed that there is enough complete effect for these disease, but it only is for the experiment. Some medical institutions at home and abroad have also extracted and analyzed the medical record data of

the hospitalized and discharged patients in the hospital information system, such as the medical expenses, disease examination and so on, obtained the demand rules of hospital beds, drugs, doctors and so on, and put forward the corresponding management strategies, which have achieved remarkable results in reducing the medical cost of patients and the operating cost of the hospital.

## 2.1 Diabetes Overview

Diabetes is one of the most serious endocrine disorders, mainly manifested as glucose, fat and protein metabolism disorders. According to the latest data from the International Diabetes Federation (IDF), the prevalence of diabetes was 9.3% (463 million people) in 2019, and 10.9% (700 million) in 2045. The urban sector (10.8%) exceeded the rural sector (7.2%). 1/2 of diabetic patients (50.1%) did not know that they had diabetes. China is already the “most devastated area” of diabetes in the world. According to Ningguang studies, the total number of diabetic patients in China has reached 92 million 400 thousand. The prevalence rate of diabetes in adults is as high as 11.6%, increasing by 3000 people per day, that is, every 10 people have diabetes. According to investigation and research, Type 2 Diabetes Mellitus (T2DM) has accounted for 95% of patients with diabetes mellitus (Mellitus). Type 2 diabetes is becoming a serious threat to global health. However, our understanding of the etiology and the best treatment of this disease is not completely clear. The current drug treatment can not well control the sustainable development of hyperglycemia, and even lead to adverse reactions such as hypoglycemia, which makes patients bear further economic burden. What is more terrible is the damage of large blood vessels, microvessels, heart, brain and kidney induced by hyperglycemia and hyperlipidemia, which brings great pressure to the regional health system and economy. Therefore, it is urgent to develop the best and effective treatment method, and to further study the prevention and treatment of diabetes.

Modern medicine has begun to study and focus on *Eucommia ulmoides* leaf, which is the dry fruit of *morusalba* L. in Moraceae. It is also called mulberry camp. It is a traditional dual-purpose resource for medicine and food in China. It was first published in the newly revised materia medica of the Tang Dynasty. It is now cultivated in central and Northern China from northeast to southwest provinces and regions, northwest to Xinjiang. This study proved that *Eucommia ulmoides* leaves in Xinjiang are rich in flavonoids, polysaccharides, alkaloids, vitamins and amino acids. They are medicinal mulberry germplasm resources in Xinjiang. Due to the unique natural environment of Xinjiang, the drought and barren *Eucommia ulmoides* leaves have become special and rare resources. It has the functions of tonifying blood, calming, reducing blood sugar and delaying aging.

In 1988, the Ministry of health listed mulberry as the first group of medicinal and food homologous plants, and was praised by the medical community as “the best health fruit in the 21st century”. It is a special medicinal and edible plant used for the treatment of hypertension, hyperlipidemia and diabetes, and has been evaluated consistently by cardiovascular experts in China. The lipid-lowering, hypoglycemic and antioxidant activities of different polar extracts of *Eucommia ulmoides* leaves are gradually being widely recognized. Hassimotto and others found that flavonoids in medicinal mulberry have

strong antioxidant capacity. Fu Daxu screened the antihypertensive activity of *Eucommia ulmoides* leaves in Xinjiang and found that its effective parts are anthocyanins and other flavonoids. It is considered that its effect is mainly related to the antioxidant effect of flavonoids. In the previous research, our research group successively adopted ultraviolet spectrophotometry, high performance liquid chromatography. The contents of flavonoids, polysaccharides, alkaloids (DNJ) and amino acids in *Eucommia ulmoides* leaves were determined by high performance capillary electrophoresis and other spectral and chromatographic separation techniques, which were consistent with the results in the literature. In view of this, we believe that *Eucommia ulmoides* leaves have multiple components of anti diabetic activity. It is necessary to further study the effective ingredients and possible targets of Xinjiang *Eucommia ulmoides* leaves for preventing and treating diabetes, which will open up a new direction for the development of mulberry resources and provide a new breakthrough for finding new antidiabetic drugs.

## 2.2 Association Rules in Data Mining

Mining rules is an important research topic in the field of data mining, and it is also an important part of database knowledge discovery. With the continuous accumulation of data, many industry stakeholders are more and more interested in mining relevant rules in the database. Association rules and mining algorithms are the main contents of association rule mining.

Correlation analysis is an important method of data mining. If the values of two or more data items repeat with a high probability, they have a certain correlation, and relevant rules can be established for these data items. Data association is an important knowledge that can be found in database. Reflects the independence or relevance of one event relative to other events. If there is an association between two or more attributes, the attribute value can be predicted based on other attribute values. For example, 90% of customers who buy bread buy milk. If you put these two products together and sell them in department stores, the sales will increase. In large databases, these related rules usually need to be filtered. Typically, support and trust thresholds are used to exclude unnecessary association rules. Supports the proportion of rules displayed in all cases. Confidence represents the proportion of cases represented by the rule when the preconditions are met. r. Apriori algorithm proposed by Agrawal et al. Is the most famous and important association rule discovery algorithm in the field of data mining. Therefore, the research direction of related rule discovery has shifted from single concept level related rule discovery to multi concept level related rule discovery. The effective algorithms of association rule mining and association rule mining in fuzzy association rules are further studied. The basic model of association rule mining is shown in Fig. 1.

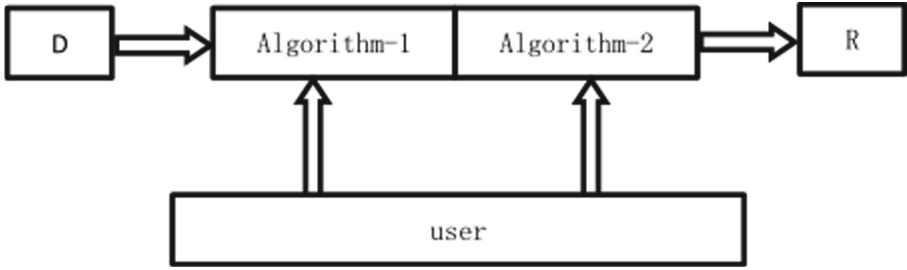


Fig. 1. Basic model of association rule mining

### 3 Data Analysis

#### 3.1 Definition of Association Rules

Association rules collect specified data items and transaction databases (all transactions are data item sets), filter the frequency relationship of data item sets in the transaction database, and find valuable correlations between data item sets through a large amount of data. Association rules mainly reflect the relevance between things. For a record reflecting the same thing, if it has both characteristic attribute a and characteristic attribute B, then characteristic attribute a and B are related. A typical application example of association rule analysis is to analyze customers' shopping habits according to a large number of customers' shopping records (for example, when buying some items, they will also buy other items), so it is also called "shopping basket analysis". There are also related phenomena in the field of medicine. If a disease may induce a variety of complications at the same time, there is a correlation between the disease and its complications.

The confidence of association rules is the percentage of both y and X contained in transaction database d, which can also be regarded as conditional probability  $P(Y/X)$ , then the percentage is the confidence (c) of association rules  $X \rightarrow Y$ , which can be calculated as follows:

$$Confidence(X \rightarrow Y) = |T : X \cup Y \subseteq T, T \in D| / |T : X \subseteq T, T \in D| \quad (1)$$

The promotion degree life of rule ( $x = y$ ) is the ratio of the probability of rule  $X \rightarrow Y$  being established to the probability of feature attribute being independent of Y. it can be calculated as follows:

$$Life(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{Support(Y)} \quad (2)$$

It can be seen from the above formula that  $Life(X \rightarrow Y) = Life(Y \rightarrow X)$ .

#### 3.2 Apriori Algorithm

Apriori algorithm is a fast mining algorithm proposed by R. Agrawal in 1994. It is also the most famous and influential frequent pattern mining algorithm. The most important property of frequent pattern mining algorithm is the property of Apriori algorithm. In

practical research, in order to avoid the high computational cost of exponential search space, we use the nature of apriori algorithm to realize frequent pattern mining algorithm to prune the search space.

The basic idea of Apriori algorithm is to generate candidate itemsets of specific scale, and then scan the database and count them to determine whether these candidate itemsets are frequent itemsets. The specific implementation process is to first scan all transactions in the database, calculate the occurrence times of each item, generate 1-candidate set C, and then determine 1-frequent set l according to the preset minimum support, and then  $L \times L$  performs connection operation to generate 2-candidate set C, scans all transactions in the database again, calculates the occurrence times of each element in C, and determines 2-frequent set l according to the preset minimum support. This process is repeated until the k-frequent set L is generated, and it is impossible to generate the  $(K + 1)$  itemset satisfying the minimum support. For the j-candidate set C ( $J = 3, \dots, K$ ), if the  $(J - 1)$  subset of an element is not the  $(J - 1)$  frequent set, it will be deleted. Accordingly, we give its operation architecture, as shown in Fig. 2.

Apriori algorithm is simple and easy to implement. When the data set is small and the frequent patterns are short, most people choose Apriori algorithm. In the Apriori algorithm, the minimum support and minimum confidence are the key thresholds, which are also the constraints of the algorithm. However, only these two conditions are not enough when processing medical data. The algorithm generates candidates and tests the support of each candidate. When the number of candidates is large, interesting medical rules cannot be found quickly; In addition, Apriori algorithm needs to scan the data set the number of times equal to the length of the longest frequent pattern. If the data set is large and the frequent pattern is long, the execution efficiency is restricted. Therefore, restrictions should be used in the procedure if interesting medical rules are to be found quickly.

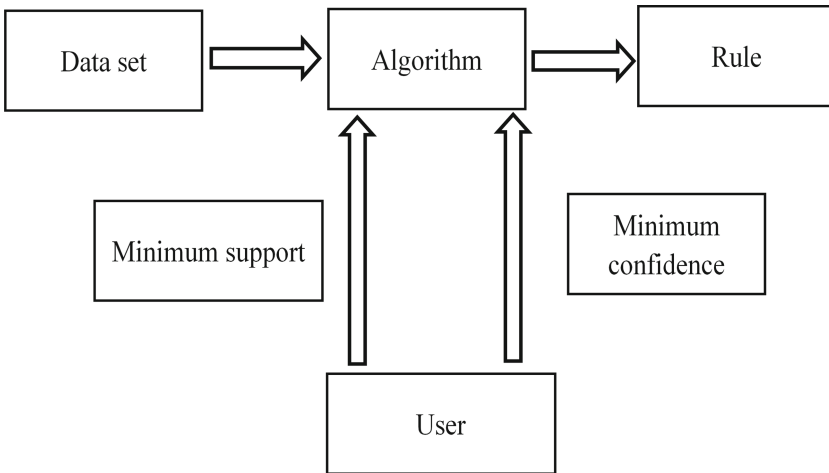


Fig. 2. Basic model of association rule mining

### 3.3 Application of Association Rules Mining in Prevention and Treatment of Diabetes Mellitus with *Eucommia Ulmoides* Leaves

Association rules mining was used to study the relationship between diabetes mellitus and *Eucommia ulmoides* leaf prevention. The results show that association rule mining can not only find the information correlation hidden in the data, but also quantify the strength of various correlations.

The pathogenesis of diabetes is complex, including oxidative stress, abnormal glucose metabolism and genetic factors. The interaction of different factors aggravates the development of the disease. The DPPH free radical scavenging rate of *Eucommia* leaves was 65.15%, and that of VC was 91.45%. Among them, the high-dose group of *Eucommia ulmoides* leaves polysaccharide had significant difference, as shown in Fig. 3. Zhong Shujuan et al. Showed that DPPH IC50 and abtsic50 of *Eucommia ulmoides* leaves were 0.35 mg/L and 0.55 mg/L respectively. In vitro antioxidant activity test of different parts of *Eucommia ulmoides* leaves, the antioxidant activity of *Eucommia ulmoides* leaves was the strongest, and the content of total flavonoids in *Eucommia ulmoides* leaves and male flowers was higher. *Eucommia ulmoides* leaves contain a lot of protein, which can be used as a source of dietary protein, and it also has more carbohydrate and fat content, providing a certain capacity.

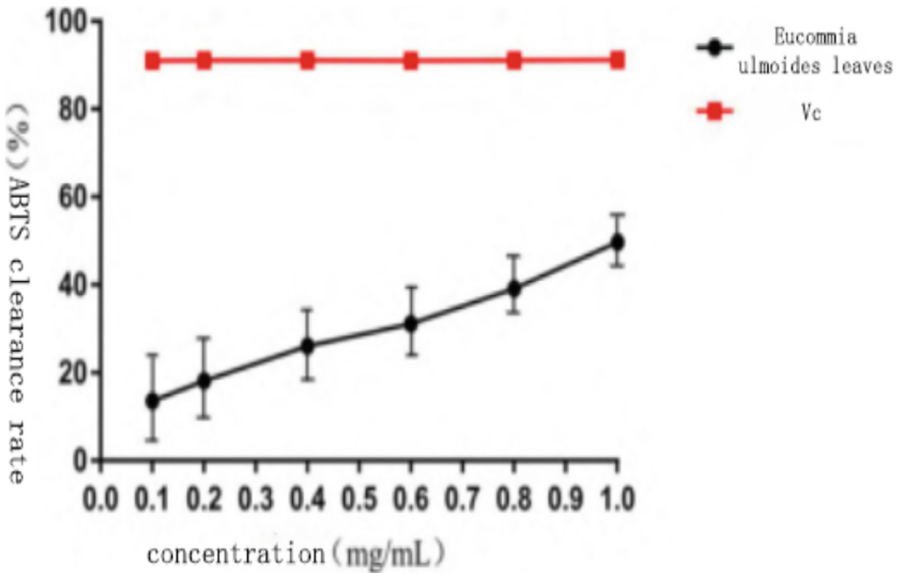


Fig. 3. Scavenging rate of *Eucommia ulmoides* leaves on ABTS free radical

Based on the Apriori algorithm of association rules, this paper discusses the effect of *Eucommia ulmoides* leaves on the prevention and treatment of diabetes. The main pharmacological effects of *Eucommia ulmoides* leaves polysaccharides are hypoglycemic, anti-aging, and hypotensive. *Eucommia ulmoides* leaves contain mineral elements, protein, vitamins, fat and other nutrients, which supplement the protein needed by the human

body and can be used as a certain source of dietary protein. The content of crude protein and various amino acids in *Eucommia ulmoides* leaves are relatively high. It contains a variety of trace elements, vitamins and other nutrients. It has the characteristics of great health function and high nutritional value. The results showed that the levels of FBG and BUN in irbesartan group, low-dose group, medium dose group and high-dose group were lower than those in model group. Among them, the levels of FBG and BUN in high-dose group were 9.21 mmol/L and 6.85 mmol/L, respectively. Xing Dongjie et al. showed that the FBG level was 14.35 mmol/L after high dose of flavone from *Eucommia ulmoides* leaves intervened in diabetes for 28 days. The water extract of *Eucommia ulmoides* leaves could play a hypoglycemic role by increasing glycolysis, and the flavonoids from *Eucommia ulmoides* leaves could stimulate the secretion of insulin from pancreas and play a hypoglycemic role. Qian zengkun and other studies showed that *Alisma orientalis* has the effect of reducing blood glucose, and its mechanism may be related to the regulation of IGF expression. The FBG of high-dose *Alisma orientalis* polysaccharide group was 19.92 mmol/L. *Eucommia ulmoides* leaves have lignans, polysaccharides, phenylpropanoids, cyclic ether mushrooms, flavonoids and other medicinal components, with antioxidant, immune regulation, hypoglycemic and a variety of health care effects.

#### 4 Example Analysis

Insulin resistance at the cellular level is defined as the decrease in the number of insulin receptors on the cell surface or the defect after the receptor, resulting in a certain concentration of insulin can not make the cell reach the expected metabolic level. Liver and peripheral tissues, such as skeletal muscle and fat, are the main sites of insulin resistance. In this experiment, human hepatoma cell line HepG2 was selected, and the high affinity insulin receptor expressed on its surface met the standard required by typical insulin receptor. Some researchers have stimulated HepG2 cells with high concentration insulin ( $1 \times 10^7$  mol/L), there are functional defects of insulin receptor and post receptor, and the receptor self phosphorylation function and insulin stimulated glucose, lipid and egg self metabolism function are reduced. Therefore, HepG2 is an ideal cell model for studying insulin resistance [1-531]. The results showed that when different concentrations of inducers induced insulin resistance in HepG2 cells, the palmitic acid induced model had the best effect and stability at 250  $\mu$ mol/L for 24 h. In the model induced by insulin and glucose,  $1 \times 10^{-7}$  mol/L and 55 mmol/L in 24 h were better. Therefore, based on the above situation, it is determined that the best scheme of insulin resistance HepG2 cell model is: 250  $\mu$ mol/L palmitic acid acts on cells for 24 h. At this time, the consumption of glucose by cells is reduced, that is, obstacles to glucose uptake and utilization occur, indicating that the establishment of insulin resistance model is successful. The established cell model was used to screen the multi components of *Eucommia ulmoides* leaves. It was found that the ethyl acetate layer (100, 200, 400  $\mu$ G/ml), n-butanol layer (200, 400, 800  $\mu$ g/ml), neochlorogenic acid (50, 100  $\mu$ G/ml), isoquercetin (50, 100  $\mu$ G/ml), Morin (50, 100  $\mu$ g/ml), myricetin (50, 100  $\mu$ G/ml) has good glucose utilization activity and no cytotoxicity, indicating that they are the main active compounds of *Eucommia ulmoides* leaves to improve insulin resistance.

The extract of *Eucommia ulmoides* leaves had a good effect on glucose metabolism and lipid metabolism of HepG2 insulin resistant cells. The indexes were determined



by HK, PK, TC and TG kits. The activities of HK and PK could be increased by the intervention of drug administration at the extraction site ( $P < 0.05$ ); The extracts from *Eucommia ulmoides* leaves could significantly reduce the accumulation of TG in cells ( $P < 0.05$ ) and clear TC ( $P < 0.05$ ). The results showed that the extract of *Eucommia ulmoides* leaves promoted the entry of glucose into liver cancer cells and accelerated the oxidative decomposition of glucose, so as to regulate glucose metabolism and improve insulin resistance. The extract of *Eucommia ulmoides* leaves may play a role in lipid clearance by reducing the accumulation of lipids in cells. Four small molecular monomer compounds in *Eucommia ulmoides* leaves increased HK activity and PK activity, and decreased TG content and TC content, but there was no significant difference. The improvement of glucose and lipid metabolism in *Eucommia ulmoides* leaves may be related to the synergistic effect of a variety of compounds and the activation of a certain pathway. When the three small molecule compounds were administered alone, there was no significant difference, which may be due to the dose or the need to cooperate with other monomers to activate PI3K Akt mTOR pathway. Traditional Chinese medicine materials have the advantages of multi-component and multi-target. Combined with network pharmacology methods, relevant pathways are predicted. WB experiment explains the activated gene protein pathway from the protein level. To study the gene and protein expression changes of key targets in PI3K Akt mTOR insulin signal transduction pathway can explain the causes of insulin resistance to a certain extent. If drug intervention can effectively regulate the key targets in this signal pathway, it can also improve the internal molecular mechanism of insulin resistance to a certain extent.

The marker proteins PI3K and Akt in this pathway are the objects we continue to observe. Studies have shown that PI3K/Akt mTOR pathway affects cell growth, proliferation, differentiation, movement, survival and metabolism through signal transduction. Insulin signaling has also been confirmed to be transmitted through this pathway. Insulin exerts its biological effects through receptors on target cells and post receptor signal transduction system. Insulin acts on cells, first binds to insulin receptor and transmits insulin signal from extracellular to intracellular. Insulin receptor binding consists of two  $\alpha$  subunits and two  $\beta$  A heterotetramer composed of subunits. Once insulin binds to subunit A, it immediately causes  $\beta$  The tyrosine residues of subunits are self phosphorylated, and catalyze the disc acidification of multiple tyrosine residues of insulin receptor binding receptor substrate to transmit insulin signal down. This pathway mainly regulates glycogen synthesis. In conclusion, *Eucommia ulmoides* leaves can improve insulin resistance and regulate glucose and lipid metabolism through PI3K Akt mTOR pathway.

## 5 Conclusion

The role of association rule mining and classification analysis in mature data mining algorithms is studied. Firstly, the association analysis model is used to extract relevant rules and generate frequent itemsets. Then, highly relevant rules are generated from frequent project sets, and interesting relationships and models between project groups are found in a large amount of data. Based on the relevant rules of the Apriori algorithm, the role of the two intermediate leaves in the prevention and treatment of diabetes is analyzed, and the relevant rules are used to extract data for analysis, which has important reference value.

## References

1. Tykaowski, B., Miaek, M., Kowalczyk, J., et al.: Phytoncides in the prevention and therapy of blackhead disease and their effect on the turkey immune system. *Int. Metabol.* **76**(11), 5701–5713 (2021)
2. Ferraris, S., Warkomicka, F., Barberi, J., et al.: Contact guidance effect and prevention of micro-fouling on a beta titanium alloy surface structured by electron-beam technology. *Nanomaterials* **11**(4), 21–34 (2021)
3. Lin, S., Liu, Z., Qian, J., et al.: Inertant effects and mechanism of Al(OH)<sub>3</sub> powder on polyethylene dust explosions based on flame propagation behavior and thermal analysis. *Fire Saf. J. Mater.* **124**, 103392 (2021)
4. Chenwei, S., Wei, L., Yahui, L., et al.: Alleviating the effect of quinoa and the underlying mechanism on hepatic steatosis in high-fat diet-fed rats. *Nutr. Metabol.* **77**(4), 479–489 (2021)
5. Sharma, V., Patial, V.: Food mycotoxins: dietary interventions implicated in the prevention of mycotoxicosis. *Int. J. Dietary* **44**(3), 289–299 (2021)
6. Pan, M.X., Zheng, C.Y., Deng, Y.J., et al.: Hepatic protective effects of Shenling Baizhu powder, a herbal compound, against inflammatory damage via TLR4/NLRP3 signalling pathway in rats with nonalcoholic fatty liver disease. *J. Diet* **19**(5), 11 (2021)
7. Williams, D.A., Horsburgh, K.J., Schultz, D.M., et al.: Proudman resonance with tides, bathymetry and variable atmospheric forcings. *Nat. Hazards. J. Int. Soc. Prevent. Mitigat. Nat. Hazards* 106–119 (2021)
8. Han, R., Kong, X.H., Zhao, F., et al.: Auricular acupressure for myopia prevention and control in children and its effect on choroid and retina: a randomized controlled trial protocol. *Trials* **22**(1), 52–68 (2021)
9. Kumari, R., Verma, A., Sharma, D., et al.: Mechanism ethnomedicinal and phytochemical effect of ayurvedic plants in prevention and management of senile dementia of Alzheimer's disease: a review. *Int. J. Med.* **33**(11), 222–234 (2021)