

# Based on data Mining, Analysis of Diabetes disease Risk Prediction and Diabetes Medication Pattern

## Abstract

After heart disorders and cancerous tumours, diabetes mellitus is the second most prevalent disease. The number of diabetic patients is rising quickly and displaying a tendency of youth due to the ongoing acceleration of people's living standards and life rhythms. According to a recent study, China has 114 million adult diabetics, a high prevalence rate, but low levels of awareness, medication adherence, and compliance. Diabetes can lead to a number of complications, including cardiovascular, cerebrovascular, and diabetic foot problems, which not only have a significant impact on the patient's survival but also put a lot of strain on the patient's family and society. These complications can be prevented if diabetes is treated and controlled early on. Therefore, controlling and preventing diabetes is a crucial method to conserve medical resources and lower medical expenses. In order to construct a prediction model of diabetes and investigate the law of medication for diabetic patients based on this analysis, we primarily read a lot of literature and gathered some significant theoretical knowledge to clarify the fundamental principles and methods of data mining. We also referred to the research findings of other scholars.

**Keywords:** Adverse drug reactions • Itemset mining • Diabetes • health care • Disease prediction • Diabetes treatment

## Introduction

Chronic hyperglycemia is the primary characteristic of diabetes mellitus (DM), a metabolic imbalance syndrome brought on by the combination of hereditary and environmental variables. Our people's quality of life has been significantly impacted by diabetes as their standard of living rises, and the disease's risks are now a big public health concern. The International Diabetes Federation recently released a study on diabetes at the World Diabetes Conference 2015, noting that 415 million adults globally, or 1 in 11, had the disease. Diabetes is the most prevalent disease in the world, affecting 110 million people in China. By 2040, the population of diabetics will have geometrically increased absent action. Diabetes prevalence in China has climbed from 0.67% to 11.6% over the past 30 years [1].

According to Preventing Type 2 Diabetes in China (2013 Edition) 3, type 2 diabetes is currently the primary cause of death in China, affecting 90% of middle-aged and elderly individuals. Studies with sufficient data indicate that patients 60 years of age and older have a prevalence of diabetes of more than 20%. Its prevalence is ten times greater than that of the 20–30-year-old youth. The incidence of diabetes increases by 68% for every additional decade of growth, adjusted for other factors. The following variables are primarily to blame for the country's rapid rise in diabetes incidence. Changes in lifestyle come first. Diabetes is becoming more common as a result of the lack of manual labour in society and the heavy use of medical resources. The second is dietary changes. The incidence of diabetes rises along with the prevalence of obesity as people's standard of life rises, followed by the population's ageing and hereditary factors, etc. [2].

As a common statistical method for categorizing binary dependent variables a common statistical approach for categorizing binary dependent variables based on independent

## Ilaria Cavallari\*

Departments of Biochemistry and Internal Medicine, University of Cardiovascular Science, Campus Bio-Medico University of Rome, Italy

\*Author for correspondence:

Ilaria.cavallari@gmail.com

Tel: +397892582304

**Received:** 01-Feb-2023, Manuscript No. actvr-23-88005; **Editor assigned:** 04-Feb-2023, PreQC No. actvr-23-88005(PQ); **Reviewed:** 18-Feb-2023, QC No. actvr-23-88005; **Revised:** 21-Feb-2023, Manuscript No. actvr-23-88005(R); **Published:** 28-Feb-2023; DOI: 10.37532/ACTVR.2023.13(1).01-03

factors is logistic regression. But the existence of a linear decision surface between the dependent and independent variables is a crucial assumption in logistic regression. Furthermore, there should be little to no multicollinearity between the independent variables when using logistic regression. Real-world circumstances frequently contain linearly separable data and multicollinearity. Using logistic regression, complex correlations are challenging to ascertain. The performance of this approach can be readily surpassed by more potent and succinct algorithms like discriminant analysis [3].

Data visualization, dimensioning, and classification all employ discriminant analysis. In a study of discriminant analysis, which has been applied to numerous classification issues, the classification was investigated. Users can debate classification techniques while performing discriminant analysis in which two or more groups and one or more independent variables are added to one of the assessed attributes. Medical researchers look at the differences between groups (defined by blood pressure, blood glucose levels, and age) across distinct variables. In order to predict whether a patient would survive based on other criteria, one study employed discriminant analysis to identify whether patients had previously experienced a heart attack. Linear discriminant analysis (LDA) and quadratic discriminant analysis are two intriguing discriminant analyses (QDA) [4].

## Materials and Method

This study was an open, approved the study protocol and patient permission forms (2015KY07). Every patient gave their signed, informed consent to take part. The procedures were carried out in accordance with the principles outlined in the Declaration of Helsinki, including any pertinent information [5].

Male patients without a history of drug use who had been given a diagnosis of LADA or T2D by research clinicians at Nanjing First Hospital met the inclusion criteria. The Chinese Diabetes Society put forth the following criteria for diagnosing LADA: (1) the presence of any islet autoantibodies (glutamic acid decarboxylase antibody (GADA), and the combination of tyrosine phosphatase-like protein antibody (IA-2A), insulin autoantibody (IAA), and

autoantibodies to zinc transporter 8 (ZnT8A), which can increase the detection rate); (2) the years of onset; and (3) a minimum of six months of The following patients were excluded from the study: (a) those who had taken systemic steroidal anti-inflammatory drugs or any other medications that could have an impact on T level within the previous three months; (b) those who had an acute infection; (c) those who had an acute diabetes complication; and (d) those who had a severe systemic disease or any other condition that the researchers deemed inappropriate for this study. Between March 2016 and January 2019, all registered patients were admitted to Nanjing First Hospital in China, and the observation duration was 12 months [6].

## Discussion

Other than metabolic disease, patients with T2D may require hospitalisation for a variety of causes. The primary metabolic factors that led to the research participants' admission to this community-based teaching hospital included poor glycemic control, the beginning, or a rapid development of diabetic complications. As a result, such acute illnesses would have less of an impact on the association studies conducted in this sample. These hospitalised T2D patients had later stages of the disease than the general T2D population, greater levels of TyG index, and vascular consequences. The sample was typical of different metabolic states due to the huge number of participants gathered over a five-year period, and this allowed for the investigation of the relationships between the TyG index and various vascular outcomes [7].

A serious microvascular consequence for those with diabetes, diabetic nephropathy is a complex illness. Our findings demonstrated that the TyG index had a strong connection with MAU but not with CKD. According to a prior study, TyG index had a greater association with urine MAU than HOMA2-IR, although there was no clear correlation with eGFR [8]. These findings revealed that the early stages of diabetic nephropathy were more adversely affected by insulin resistance than the late stages, which are marked by the development of renal function insufficiency. The early onset of renal function abnormalities in diabetes individuals may have been triggered by insulin resistance symptoms

including hyperglycemia and dysfunctional lipid oxidation and consumption [9]. This is one explanation for the occurrence. Along with insulin resistance, other complicated metabolic abnormalities such hyperglycemia-induced metabolic acidosis and persistent inflammatory response may develop as dysglycemia progresses. These conditions place a heavy strain on the fragile kidney, worsening renal failure. More research should only be done to understand the processes driving the progression of renal insufficiency at this later stage [10].

### Conclusions

The effective use of medical data is a topic of discussion due to the exponential expansion of medical data and the development of data mining tools. In order to analyse the disease and drug use of diabetic patients in clinical data, this paper combines data mining models with medical data features. Based on the experience of many researchers, it proposes a set of reasonable and appropriate prediction models for the risk of diabetes in high-risk groups. The model has been examined using the WEKA platform, and it has been discovered that, in comparison to earlier research, its predicted accuracy has greatly increased.

These data were compared with those of other researchers. The dataset utilised in this work, the Pima Indian Diabetes dataset, is the standard for many studies on diabetes-related data mining. The trial data's advantages and disadvantages were carefully pre-processed to make sure they were accurate, reasonable, and uniform. The analysis of diabetes medication patterns can be performed by keyword design during the model experiments, yielding the highest frequency of Danshen, occurrence, after comparing the improved K-means and Logistic regression methods, which have a prediction accuracy

of up to 95.42% and a conclusion of better results in terms of average prediction.

### Conflict of Interest

None

### Acknowledgement

None

### References

1. Tonesk X, Buchanan RG. An AAMC pilot study by 10 medical schools of clinical evaluation of students. *J Med Educ.* 62, 707–18 (1998).
2. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ.* 320, 50–2 (2000).
3. Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med.* 74, 842–9 (1999).
4. Siminoff LA, Zhang A, Colabianchi N *et al.* Factors that predict the referral of breast cancer patients onto clinical trials by their surgeons and medical oncologists. *J Clin Oncol.* 18, 1203–11 (2000).
5. Ross S, Grant A, Counsell C *et al.* Prescottt RJ. Barriers to participation in randomised controlled trials. *J Clin Epidemiol.* 52, 1143–56 (1999).
6. Ding EL, Song Y, Malik VS *et al.* Sex differences of endogenous sex hormones and risk of type 2 diabetes. *Jama.* 295:1288-1299 (2006).
7. Mohammed M, Al-Habori M, Abdullateef A *et al.* Impact of metabolic syndrome factors on testosterone and SHBG in type 2 diabetes mellitus and metabolic syndrome. *J Diabetes Res.* 492:78-98 (2018).
8. Malipatil NS, Yadegarfar G, Lunt M *et al.* 14-year prospective outcome in 550 men with type 2 diabetes. *Endocrinol Metab.* 2: 3-4 (2019).
9. Kelsey MM, Bjornstad P, McFann K *et al.* Testosterone concentration and insulin sensitivity in young men with type 1 and type 2 diabetes. *Pediatr Diabetes.* 17:184-190 (2016).
10. Juneja R, Palmer JP. Type 1 1/2 diabetes: myth or reality? *J Autoimmun.* 29:65-83 (2009).