Determination of predisposition to diabetes mellitus based on Fuzzy neural network

Determinación de la predisposición a la diabetes mellitus basado en red neuronal Fuzzy

Ilyas Idrisovich Ismagilov, Doctor of Technical Sciences, Professor of Economic Theory and Econometrics Department of the Institute of Management,Economics and Finance of Kazan (Volga Region) Federal University; e-mail: iiismag@mail.ru; Scopus ID: 6603342575;
Dina Vladimirovna Kataseva, Graduate Student and Senior Lecturer of Information Security Systems Department of the Institute of Computer Technologies and Information Security of Kazan National Research Technical University named after A. N. Tupolev-KAI; e-mail: DVKataseva@kai.ru Scopus ID: 57193401954;
Alexey Sergeevich Katasev, Doctor of Technical Sciences, Professor of Information Security Systems Department of the Institute of Computer Technologiesand Information Security of Kazan National Research Technical University named after A. N. Tupolev-KAI; e-mail: ASKatasev@kai.ru; Scopus ID: 57193408902;
Anastasia Olegovna Barinova, Master's Student of Information Security Systems Department of the Institute of Computer Technologies and InformationSecurity of Kazan National Research Technical University named after A. N. Tupolev-KAI; e-mail: banderova2009@yandex.ru; Scopus ID: 57221537005; e-mail (contact email address): iiismag@mail.ru
Received: 06/24/2022 Accepted: 08/19/2022 Published: 09/25/2022 DOI: https://doi.org/10.5281/zenodo.7372918

Abstract

This article solves the problem of determining the occurrence of diabetes mellitus in pregnant women. The aim of the study is to find an effective method based on the machine learning method, which will simplify the procedure and increase the speed of determining this disease with a high degree of accuracy. The currently existing methods for determining diabetes mellitus cannot allow constantly monitoring the presence of diabetes mellitus in pregnant women due to the complexity of the procedure, as well as the long time it takes to process the analysis results. Therefore, the use of modern technologies, in particular machine learning methods, will get rid of these disadvantages and allow to continuously monitor the possible occurrence of the disease. In this article, the subject area is analyzed and the relevance of the research topic is considered, the initial data preparation and collection for a neuro-fuzzy model constructing are carried out. In addition, intelligent models based on various machine learning methods have been constructed. The best method based on a fuzzy neural network has been chosen, which allows to classify the available data with a high degree of accuracy. A use-case diagram for solving practical problems of determining the presence of diabetes mellitus has been developed.

Keywords: diabetes mellitus, diagnostics, modeling, classification, machine learning, fuzzy neural network, knowledge base.

Resumen

Este artículo resuelve el problema de determinar la aparición de diabetes mellitus en mujeres embarazadas. El objetivo del estudio es encontrar un método eficaz basado en el método de aprendizaje automático, que simplificará el procedimiento y aumentará la velocidad de determinación de esta enfermedad con un alto grado de precisión. Los métodos existentes actualmente para la determinación de diabetes mellitus no pueden permitir monitorear constantemente la presencia de diabetes mellitus en mujeres embarazadas debido a la complejidad del procedimiento, así como al largo tiempo que lleva procesar los resultados de los análisis. Por lo tanto, el uso de tecnologías modernas, en particular métodos de aprendizaje automático, eliminará estas desventajas y permitirá monitorear continuamente la posible aparición de la enfermedad. En este artículo se analiza el área temática y se considera la relevancia del tema de investigación, se realiza la preparación y recolección de datos iniciales para la construcción de un modelo neuro-borroso. Además, se han construido modelos inteligentes basados en varios métodos de aprendizaje automático. Se ha elegido el mejor método basado en una red neuronal difusa, que permite clasificar los datos disponibles con un alto grado de precisión. Se ha desarrollado un diagrama de casos de uso para resolver problemas prácticos de determinación de la presencia de diabetes mellitus.

Palabras clave: diabetes mellitus, diagnóstico, modelado, clasificación, aprendizaje automático, red neuronal difusa, base de conocimientos.

Introduction

Diabetes mellitus is an endocrine disease characterized by high blood sugar levels due to impaired ability of the body to process glucose[1]. The disease depends on insulin, which the pancreas cannot produce in sufficient quantities.

The most common symptoms of diabetes mellitus are weight loss, increased thirst, and hunger[2]. At different stages of the disease, symptoms can develop at different quickness. For example, in 1st type of diabetes, they can develop rapidly, and in 2nd type of diabetes - extremely slowly and may be invisible or absent[3,4].

Over time, diabetes can cause serious health problems such as heart disease, vision loss, and kidney disease. The currently existing methods for determining diabetes mellitus do not allow constantly monitoring the presence of diabetes mellitus[5], in particular, in pregnant women due to the complexity of the procedure, as well as the large amount of time it takes to process the analysis results. Therefore, the use of modern technologies, in particular machine learning methods[6-9], will eliminate these shortcomings and make it possible to control the occurrence of the disease in real time.

To solve the problem of timely diagnosis of diabetes mellitus, in this work, it is proposed to use a machine learning algorithm, with the help of which, based on blood analysis data, it will be possible to make a preliminary diagnosis of diabetes mellitus presence in a pregnant woman in the early stages, which will reduce the risk of developing the disease. Also, the use of this approach will simplify and significantly reduce the diagnosis time from several hours to several minutes.

Methods

As a material for the intelligent models analysis and construction, a dataset containing information about analyzes taken from pregnant women over 20 years old, who had diabetes mellitus, was used. These data were obtained from the open source of the US National Institutes of Health[10]. The complete dataset, including 768 records, consisted of 9 columns, 8 of which were input parameters and 1 output:

1) the number of pregnancies;

2) plasma glucose concentration;

3) blood pressure;

4) skin thickness;

5) the level of insulin in the blood serum;

6) body mass index;

7) diabetic pedigree function;

8) age;

9) the presence or absence of diabetes.

From the initial data, a training sample was formed to construct intelligent models for diabetes mellitus determining. The sample includes prepared data presented in the format "inputs - parameters for determining diabetes mellitus, output - the presence or absence of diabetes".

The initial data preparation for analysis was carried out on the Loginom Studio Community analytical platform basis[11]. For this, a special tool of the platform "Preprocessing" was used, on the basis of which the following stages were implemented:

1) loading the initial data sample;

2) filling in the gaps in the data;

3) identification and elimination of duplicates and inconsistencies in the data;

4) editing outliers and extreme values in the data;

5) data smoothing and highlighting the trend component;

6) conducting correlation analysis to identify the relationship between the input and output parameters in the data;

7) uploading prepared data suitable for analysis.

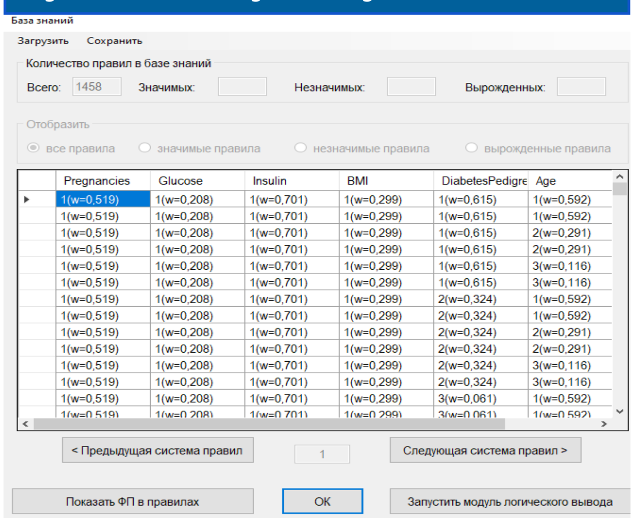
The implementation of these stages made it possible to identify 12 duplicates and 12 contradictions, which were excluded from the sample, 76 outliers and 6 extreme values were detected and processed. In this case, the outlier was considered such data that differed by 3 standard deviations from their mathematical expectation. To determine the extremes, 5 standard deviations from the average value were used. The smoothing procedure using the Hodrick-Prescott filter[12] was applied to all found outliers and extreme values. This filter is widely used for smoothing time series and highlighting the trend component in them [13, 14]. Further, a correlation analysis was carried out using the Pearson method. The smallest correlation with the output parameter is found for the "blood pressure" and "skin thickness" input parameters. Their correlation was less than 0.1, so these parameters were excluded from the analysis. Thus, the obtained data for analysis contained 744 lines, included 6 input parameters and 1 output. A testing data sample of 100 lines was randomly generated from this data. The rest of the data were used for training.

Based on the data obtained, a fuzzy neural network was trained and a knowledge base was formed to determine the presence or absence of diabetes mellitus in humans. The neural fuzzy model consisted of 6 input and 1 output neurons. The knowledge base contained 481 fuzzy rules.

The generated knowledge base can be presented in the form of a table containing the conditional parts of the rules and the state of a person in terms of the presence or absence of the desired disease [15]. The numbers in the output column cells indicate the classification result (diagnosis). The table also provides information on the weight of the conditions in the rules and the validity of each rule. The last column is output and contains information about the classification result.

A fragment of the generated knowledge base is shown in Figure 1.

Figure 1. The knowledge base fragment



Based on the formed knowledge base, studies were carried out to assess the accuracy of determining diabetes mellitus in humans.

Results and Discussion

The diabetes mellitus determining accuracy assessment was carried out on a testing data sample using the generated knowledge base. For this, a logical inference module was used based on the rules of the knowledge base[15]. The 100-line testing data included data on 50 non-diabetic pregnant women and 50 diagnosed diabetic women. It should be noted that the testing data were not duplicated in the training set and were not used in training a fuzzy neural network and forming a knowledge base.

The classification accuracy for the testing sample was 94%. Consequently, the model has acquired the ability to generalize to accurately determine the state of the object and can be used to solve practical problems of determining diabetes mellitus in humans.

Let's also compare the classification algorithm used in this article with the following algorithms:

- multilayer neural network;

- decision tree with algorithm C4.5;

- linear regression;

- logical regression.

Data classification based on these methods was carried out using the Deductor Studio analytical platform[16]. The construction of the corresponding classification models was carried out on the training data set. The accuracy of the models was evaluated on the testing data sample.

The result of comparing the classification accuracy by different methods is presented in Table 1.

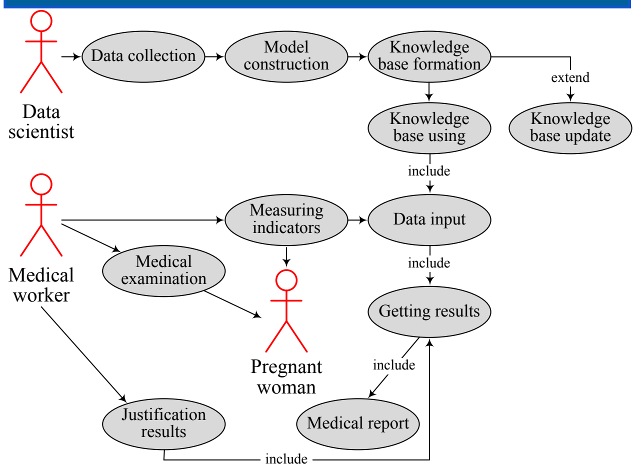
Table 1. Comparison of the diabetes mellitus determining accuracy by different methods

Classification method	Diabetes mellitus determination accuracy (%)
Multilayer neural network	86,5
Decision tree C4.5	89
Linear regression	88,75
Logistic regression	80
Neural Fuzzy Model	94

Thus, when using the neuro-fuzzy model and, accordingly, the formed knowledge base, the highest accuracy of determining diabetes mellitus in pregnant women is achieved.

The knowledge base formed on the basis of the fuzzy neural network training is proposed to be used as part of an intelligent system for determining diabetes mellitus in humans. To solve this problem, a use-case model of passing a medical examination using such a system has been developed (see Fig. 2).

Figura 2. Modelo de caso de uso para la determinación de diabetes mellitus



In this model, the analyst collects data, constructs a neuro-fuzzy model[17-19], forms a knowledge base[20-24], and updates it if necessary. The finished knowledge base is used in the software package.

The medical worker assesses the general condition of the pregnant woman and enters the following data into the software package:

- glucose level according to the last analysis;

- the age of the person;

- diabetic pedigree function;

- the number of pregnancies;

- body mass index;

- the result of the analysis of the determination of the level of insulin in the blood serum.

These data are transmitted to a system that determines the susceptibility of a pregnant woman to diabetes. A medical worker writes a conclusion on the basis of the data received. If there is no predisposition, then the patient does not need additional examinations. If there is a predisposition, a pregnant woman is advised to follow a diet. In addition, she is assigned additional tests to measure her blood sugar.

It should be noted that the use of such a system may be only one of the stages of a medical examination. Additionally, classical laboratory observations can be carried out.

Summary

In this work, the aim of which was to increase the efficiency of determining diabetes mellitus in pregnant women based on the neuro-fuzzy model construction, the following tasks were solved:

- the subject area was analyzed and the relevance of the problem being solved was substantiated, including it was determined that the existing methods for diabetes mellitus determining are ineffective due to the complexity and time of obtaining the analysis result, which excludes the possibility of promptly controlling the occurrence of this disease;

- initial data collecting and preparating for a neuro-fuzzy model building;

- based on the prepared data and various machine learning methods, models were constructed, which made it possible to choose the best method in terms of classification accuracy.

The result of all work is a formed knowledge base that can be used for the prompt determination of diabetes mellitus in humans, which will reduce the risks of further disease development.

Conclusions

Thus, the work has solved the urgent problem of determining diabetes mellitus in humans based on the neuro-fuzzy model construction and the knowledge base formation. The obtained testing results based on the knowledge base rules using the inference algorithm on the rules showed the proposed approach effectiveness to solving the problem. The classification accuracy based on the generated knowledge base exceeds the accuracy of other classification methods. This indicates its effectiveness and the possibility of practical use as a part of an intelligent system for solving the task.

Acknowledgements
This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program.

References

1. Liebl A. Practical implementation of insulin therapy for type 2 diabetes: Simple and complicated cases. Diabetes Aktuell. 2021;19(5):204-212.

2. Aghajanzadeh H, Abdolmaleki M, Ebrahimzadeh MA, Mousavi T, Izad M. Methanolic extract of sambucus ebulus ameliorates clinical symptoms in experimental type 1 diabetes through anti-inflammatory and immunomodulatory actions. Cell Journal. 2021;23(4):465-473.

3. Muller M, Wheeler BJ, Blackwell M, Medlicott NJ, Al-Sallami HS. The influence of patient variables on insulin total daily dose in paediatric

inpatients with new onset type 1 diabetes mellitus. Journal of Diabetes and Metabolic Disorders. 2018;17(2):159-163.

4. Verma S, Alam R, Ahmad I, Ali K, Hussain ME. Effect of glycemic control and disease duration on cardiac autonomic function and oxidative stress in type 2 diabetes mellitus. Journal of Diabetes and Metabolic Disorders. 2018;17(2):149-158.

5. Bartolic L, Zoric B, Martinovic G. E-Gluko: A Ubiquitous System for Health Status Monitoring and Tracking in Diabetes Patients. Proceedings of International Conference on Smart Systems and Technologies. 2018:153-158.

6. Dagaeva M, Garaeva A, Anikin I, Makhmutova A, Minnikhanov R. Big spatiotemporal data mining for emergency management information systems. IET Intelligent Transport Systems. 2019;13(11):1649-1657.

7. Perfilieva IG, Yarushkina NG, Afanasieva TV, Romanov AA. Web-based system for enterprise performance analysis on the basis of time series data mining. Advances in Intelligent Systems and Computing. 2016; 450:75-86.

8. Ismagilov II, Khasanova SF, Katasev AS, Kataseva DV. Neural network method of dynamic biometrics for detecting the substitution of computer. Journal of Advanced Research in Dynamical and Control Systems. 2018;10(10 Special Issue):1723-1728.

9. Ismagilov II, Molotov LA, Katasev AS, Kataseva DV. Construction and efficiency analysis of neural network models for assessing the financial condition of enterprises. Journal of Advanced Research in Dynamical and Control Systems. 2019;11(8 Special Issue):1842-1847.

10. Cimino JJ, Ayres EJ. The clinical research data repository of the US National Institutes of Health. Studies in Health Technology and Informatics. 2010;160(PART 1):1299-1303.

11. Mitina OA, Yurchenkov IA. Data Classification in Medicine and Healthcare Service. Lecture Notes in Networks and Systems. 2021;229:396-403.

12. Djehiche B, Hilbert A, Nassar H. On the functional Hodrick-Prescott filter with non-compact operators. Random Operators and Stochastic Equations. 2016;24(1):33-42.

13. Trimbur TM. Detrending economic time series: A Bayesian generalization of the hodrick-prescott filter // Journal of Forecasting. 2006;25(4):247-273.

14. Li Z, Tanaka G. HP-ESN: Echo State Networks Combined with Hodrick-Prescott Filter for Nonlinear Time-Series Prediction. Proceedings of the International Joint Conference on Neural Networks. – 2020, 9206771.

15. Katasev AS. Neuro-fuzzy model of fuzzy rules formation for objects state evaluation in conditions of uncertainty. Computer Research and Modeling. 2019;11(3):477-492.

16. Nagy Z, Soper DE. Jets and threshold summation in Deductor. Physical Review D. 2018;98(1):14-35.

17. Chupin MM, Katasev AS, Akhmetvaleev AM, Kataseva DV. Neuro-fuzzy model in supply chain management for objects state assessing. International Journal of Supply Chain Management. 2019;8(5):201-208.

18. Vlasenko A, Vynokurova O, Vlasenko N, Peleshko M. A Hybrid Neuro-Fuzzy Model for Stock Market Time-Series Prediction. Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP. 2018:352-355.

19. Smagin SV. Method for inductive formation of medical diagnostic knowledge bases. CEUR Workshop Proceedings. 2015; 1536:48-56.

20. Grigoriev OG, Kobrinskii BA, Osipov GS, Molodchenkov AI, Smirnov IV. Health Management System Knowledge Base for Formation and Support of a Preventive Measures Plan. Procedia Computer Science. 2018; 145:238-241.

21. Diaz CI, Rey PJ, Ramírez PM, Herrera PE, Verduga DJ, Jara DA, Aveiga RA, Rivera RH, Zambrano CL, Mendoza JJ, Suárez JE. Características clínico-epidemiológicas de los pacientes amputados ingresados a la unidad de pie diabético del Hospital Abel Gilbert Pontón, Ecuador. Archivos Venezolanos de Farmacología y Terapéutica. 2019;38(2):40-3.

22. Gualaquiza González R, Pérez Granja A, Tapia Caisaguano A, Legña Tibanta D, Bastidas Jiménez E, Gaibor Ortiz A, Bastidas Haro T, Allauca Yumiseba M, Bravo Bohórquez G, Miranda Buenaño F, Castañeda Morales D. Incidencia y características clínicas de lactantes menores con neumonía adquirida en la comunidadingresados en el Hospital Pediátrico "Baca Ortiz", Ecuador. Archivos venezolanos de farmacología y terapéutica. 2020;39(4):260-3.

23. Zambrano AM, Díaz CE, Zamora EE, Laverde AJ, Adame FT, Ronda JR, Quintanilla SB, Uvidia VC. Salud mental de los pacientes con enfermedades crónicas durante la pandemia por COVID-19. Sindrome Cardiometabólico. 2020;10(1):20-4.

24. Andrade AV, Cruz JG, Jiménez SI, Demera JA, Ronda JR, Uzhca BS, Tinoco YJ, Lara JC. Promoción de la actividad física en los pacientes con enfermedad cardiovascular durante el confinamiento por covid-19. Sindrome Cardiometabólico. 2020;10(1):16-9.