Deep learning-based genetic detection and pathogenesis of diabetes

Sihan Chen,
University of Toronto, Toronto, Canada
veronica.chen@mail.utoronto.ca

Linshan Cao
Northeast Forestry University, Harbin, China
clover20020121@qq.com

***Abstract*—In the development of modern science and technology innovation, how to use deep learning algorithms to diagnose diseases is the main problem in artificial intelligence research. As one of the major diseases affecting human physical and mental health, diabetes is a disease caused by high glucose levels in the blood, with high morbidity, high treatment cost and high mortality. If it is not effectively controlled for a long time, it will directly damage various human body organs. In this paper, after understanding the research status of diabetes gene testing and morbidity association, combined with the deep learning concept in the field of artificial intelligence, the main research is how to prevent and control the complications of diabetes at the source, how to build a scientific and effective disease prediction system, to create convenient conditions for medical diagnosis, and truly achieve the development goal of early detection and early treatment.**

Keywords—Deep learning; Diabetes mellitus; Genetic testing; Morbidity association; Prediction system

I. INTRODUCTION

With the gradual development of China's medical technology level, most third-class hospitals have built a complete electronic medical system for hospitals. On the one hand, it can help medical staff accurately record various data information and obtain more effective content during data development. On the other hand, it can protect patients' privacy and provide quality services for patient treatment while mining data resources. According to the current domestic and foreign medical conditions and diagnostic levels, the treatment of diabetes is still unable to achieve eradication; only effective control measures can be taken to avoid excessive deterioration of the disease. Traditionally, diabetes treatment combines the medical staff's clinical experience and the patient's relevant report and quickly develops the appropriate consultation and treatment plan. At present, both home and abroad, the diagnosis and treatment of diabetes are mainly based on prevention, and no thorough treatment plan has been proposed. However, with the further aggravation of patients' disease, the corresponding treatment plan is also different in the face of relatively complex complications. According to the accumulated experience of practice, early prevention of diabetes is crucial, as it can help patients prevent or delay the occurrence of complications. From the point of view of experimental analysis, if people adhere to medication, diet control and daily exercise, then diabetes can be effectively controlled. Among them, older patients with early prevention and detection work than the age of patients without prediction and protection of the probability of complications is lower.

.

II. RESEARCH AND DEVELOPMENT OF DIABETES PATHOGENIC GENES

A. Common types of diabetes

Generally speaking, the existing types of diabetes can be divided into four types: first, type 1 diabetes, also known as insulin-dependent diabetes mellitus, is caused by autoimmune system dysfunction, environmental or genetic factors, mainly in young people, need lifelong dependence on insulin treatment, often appear weight loss, excessive drinking and eating symptoms; Secondly, type 2 diabetes mellitus, also known as non-insulin-dependent diabetes mellitus, is caused by insulin resistance and progressive insufficiency of insulin secretion. It often occurs in people over 40 years old, and the patients are relatively obese. The disease of this type of patient is relatively mild, mainly showing symptoms such as thirst and dry mouth. Thirdly, gestational diabetes refers to the first occurrence of abnormal glucose metabolism in patients after pregnancy, which is a very common complication during pregnancy. Gestational diabetes is the result of a relative lack of insulin, which is unable to maintain the normal growth of the mother and fetus. If the disease is not effectively controlled, it is likely to result in large babies, deformed fetuses, maternal and infant death and other adverse phenomena. Finally, a special type of diabetes is caused by environmental, disease, genetic and other factors. For example, in patients with pancreatitis, after pancreas resection, insulin secretion will gradually decline, which will lead to diabetes. After a clear diagnosis, medical staff will use general treatment, medication, and other relief methods.

Nowadays, the diagnosis of diabetes in China mainly relies on the detection of blood sugar content and glycosylated hemoglobin content in the human body to determine whether the two meet the human standard. According to the World Health Organization's diagnostic criteria for type 2 diabetes, blood glucose should be greater than 7.0mmol/L on fasting, ≥11.1mmol/L two hours after a meal, and ≥11.1mmol/L in random detection. At present, whether to diagnose diabetes only based on blood glucose value is still controversial. The real danger of type 2 diabetes, for example, is its complications, which vary from stage to stage, making it impossible to apply a single diagnostic criterion.

B. Research and Development of Diabetes Gene Detection and Pathogenesis

2024 Fourth International Conference on Digital Data Processing (DDP) | 979-8-3315-1570-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/DDP64453.2024.00011

979-8-3315-1570-6/24/$31.00 ©2024 IEEE
DOI 10.1109/DDP64453.2024.00011

Since the 1980s, people have become increasingly aware of safety, and countries have begun to apply several advanced technologies to build disease prediction models. According to their own primary national conditions and the development of major diseases, they have mastered many empirical detection and morbidity research results. Among them, as one of the global diseases, domestic and foreign scholars mainly focus on three aspects when studying diabetes prediction. First, the screening model of diabetes is studied to facilitate the early detection and treatment of patients. Second, the risk of diabetes identify the influence of various factors on the composition of diabetes and early intervention. Finally, the complications of diabetes were studied, and the intervention plan was formulated according to patients' physical quality.

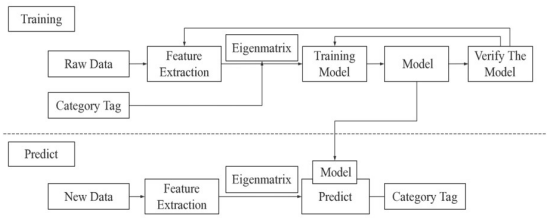
For example, some scholars have used BP neural network and binary Logistic regression to study the specific manifestations of type 2 diabetes in different onset stages. Some scholars have used machine learning algorithms to explore the prediction model of diabetes, which can accurately distinguish which type of diabetes patients belong to. Some scholars have used multilayer neural networks to predict the complications of type 2 diabetes and can accurately distinguish between nephropathy and retinopathy. Some scholars used a genetic algorithm to optimize the BP neural network model, which enhanced the stability of the prediction model, and selected the influencing factors of 18 kinds of complications as input indicators, and finally could accurately predict and analyze whether the patient was diabetic nephropathy or neuropathy, but the actual accuracy was low. Some scholars selected high-risk groups and developed effective prediction methods. Logistic regression, random forest, gradient enhancement and other algorithms were used to predict and analyze the two types of diabetes, cardiovascular disease and kidney disease. Some scholars used logistic regression and random forest algorithms to predict whether patients suffer from diabetes and retinopathy mainly, and the actual accuracy rate could reach 80.8%. Some scholars chose MexicAn-American and non-Hispanic Caucasians as research groups, and built a complete prediction model after comprehensively mastering patients' basic information, mainly analyzing the risk assessment probability of diabetes patients in the population. These findings further prove the practical significance of deep learning in diabetes gene detection and concurrent association research.

III. Methods

A. Deep Learning

As a new research direction in the field of collective learning, in-depth learning is closer to the initial development goal, which is mainly to learn the internal law and representation level of sample data, to obtain more valuable information and data, and to provide more help for diabetes gene detection and concurrent association analysis. Deep learning has made excellent achievements in multimedia learning, natural language processing, machine learning, data mining and other fields. It mainly uses machines to imitate human activities, such as audio-visual and thinking, to solve complex pattern recognition problems and make artificial intelligence technology achieve excellent results. Combined with the deep learning process analysis shown in Figure 1 below, it can be seen that the overall operation is mainly divided into two parts: one is training, the other is prediction, the former includes features extraction, training model, model verification and other links, while the latter includes new data, feature extraction, data prediction, classification label and other links.

FIGURE I. FLOW CHART OF DEEP LEARNING MODEL



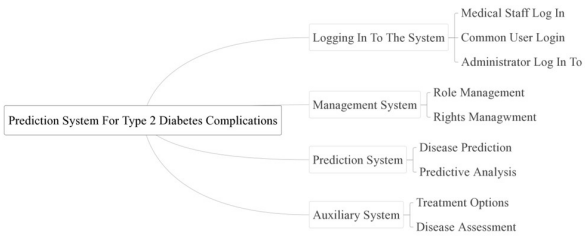
.

B. System Structure

According to the requirements of different users for system operation, to better meet the requirements of testing, detection and prediction analysis of type 2 diabetes complications, the overall system design mainly includes several modules, such as user login, information input, system management, auxiliary diagnosis, auxiliary treatment, etc. The specific structure is shown in Figure 2 below:

.

Figure 2. STRUCTURE DIAGRAM OF THE SYSTEM MODULE



Different roles have different operation rights when the system is running, and users can use different function modules after logging in. The details are as follows: First, log in to the system. The functional module of medical staff includes a login terminal, modification terminal and viewing terminal. In contrast, the functional module of patients includes personal information login, password modification, information input, result viewing, treatment plan, etc. Secondly, management system. This functional module is mainly divided into two parts: role management and authority management. Patients cannot log in to the work interface of medical staff, and staff cannot log in to the information interface of patients; they belong to two independent systems,

and both have the right to modify the password. Third is the prediction system. This functional module is mainly divided into predictive analysis and complication prediction. The system will complete data preprocessing according to patients' input information. The information will be processed by the particle swarm optimization algorithm DBN, and the final result will be displayed on the prediction interface. Finally, the auxiliary system. This functional module is mainly divided into two parts: one refers to the treatment plan, and the other hand refers to the condition detection; medical staff should develop the treatment plan according to the changes in the patient's condition, and then real-time monitoring of the patient's condition, and do a good job of disease assessment and analysis.

IV. ANALYSIS AND RESULTS

To verify the application effect of deep learning in type 2 diabetes gene detection and morbidity association research, this study conducted validation analysis on the system structure proposed above, mainly combining the confusion matrix, accuracy, recall rate and other evaluation models. The research results of the confusion matrix are shown in Table 1 below:

Table 1 RESEARCH RESULTS OF THE CONFUSION MATRIX

	Correlated (positive class)	independent (negative class)
Be retrieved	TP	FP
Not retrieved	PN	TN

The calculation formula for accuracy is as follows:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+FN+TN}$$

The formula for calculating the recall rate is as follows:

$$\text{Recall}=\frac{TP}{TP+FN}$$

In the above formula, TP+TN represents the correct number in the classification, and TP+FP+FN+TN represents the total number of samples. The higher the calculation result's accuracy, the higher the final classification accuracy can be guaranteed.

V. DISCUSSION

In this paper's research experiment, confusion matrix evaluation is used to analyze four models: support vector machine, grid-optimized support vector machine, deep placement network, and PSO-DBN. Figure 3 represents the algorithm flow of the support vector machine, and Figure 4 represents the particle swarm optimization algorithm flow.

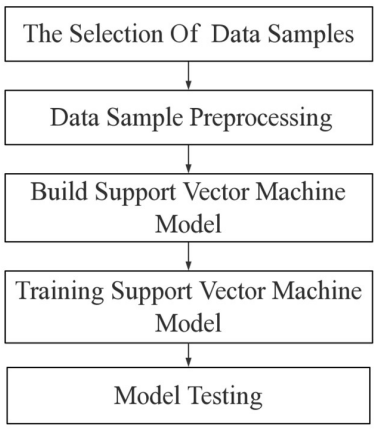


FIGURE. 3 ALGORITHM FLOW CHART OF SUPPORT VECTOR MACHINE

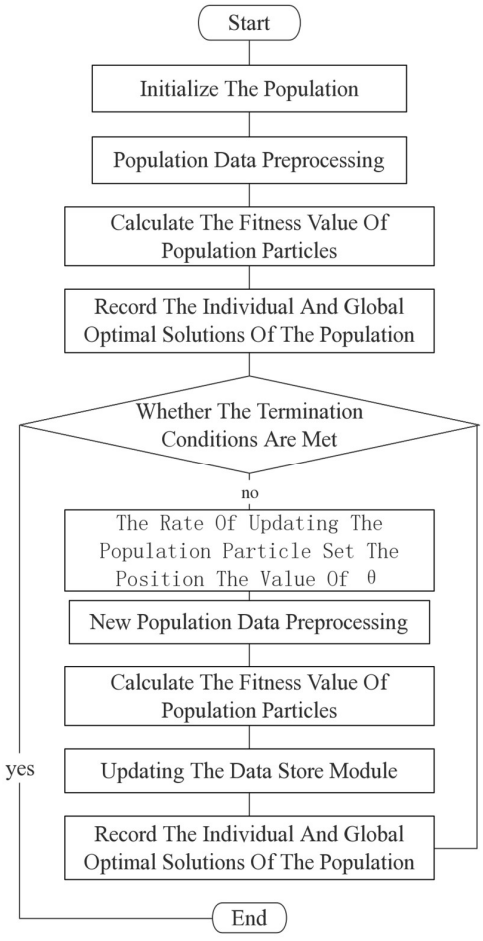


FIGURE 4 FLOW CHART OF PARTICLE SWARM OPTIMIZATION

In the confusion matrix of SVM after optimization, the prediction accuracy of kidney disease can reach 71%, and the accuracy of eye disease is the highest, followed by neuropathy, the actual accuracy can reach 82%.

The analysis of the confusion matrix of the deep location network showed that the prediction accuracy of nephropathy

could reach 91%, that of retinopathy could reach 84%, and that of neuropathy was the lowest, also reaching 71%.

By analyzing the PSO-DBN model, it was found that the actual prediction accuracy could reach 91%, 100% for eye disease, 97% for neuropathy and 76% for nephropathy, and the overall value was significantly higher than other models.

VI. CONCLUSION

Based on the analysis of the above research system structure and experimental results, it can be seen that when predicting and analyzing various data of type II diabetes and its complications, the prediction model constructed is in line with the current data research and analysis requirements, and can accurately predict the complications of diabetes and help doctors diagnose and treat as soon as possible. Therefore, in the future innovation and development of science and technology, the Chinese medical field should strengthen the research and application of deep learning theory and pay attention to constructing a complete predictive model system. Only in this way can diabetes gene detection and concurrent association research be done well.

REFERENCES

[1] XueTao Wang, YanPing Zhu. Association between gene detection and morbidity of type 2 diabetes mellitus. Journal of Integrated Traditional Chinese and Western Medicine Cardio-Cerebrovascular Diseases, 2017, 15(18):3.

[2] Shuai Ming. Research status and Prospect of deep learning artificial intelligence assisted diagnosis of diabetic retinopathy. Chinese Journal of Experimental Ophthalmology, 2019, 37(8):5.

[3] Huiying Pan, Jia Liu, Yue Liu, et al. Study on the relationship between high-density lipoprotein cholesterol and diabetes mellitus in middle-aged and elderly people in Beijing. Chinese Journal of Cardiology, 2021, 26(5):5.

[4] Summary of Ming Shuai, proofread by Lei Bo. Diagnosis of diabetic retinopathy by artificial intelligence based on deep learning. Chinese Journal of Experimental Ophthalmology, 2019(008):037.

[5] Feiling Ai, Xue Cao, Xiaochun Li, et al. A meta-analysis of the association between smoking and risk of type 2 diabetes in Asian adults based on a prospective cohort study. Chinese Journal of Health Management, 2020, 14(3):9.

[6] Xue Du, Ying Yang, Hui Peng, et al. Research progress on the correlation between diabetes mellitus and tumor. Advances in Modern Biomedicine, 2016, 16(35):4.

[7] Xin Jiao, Xianlin Zhang. Study on the correlation between BRAFV600E and papillary thyroid carcinoma, Advances in Clinical Medicine, 2022, 12(12):6.

[8] Caiying Yu. Research progress of the association between environmental endocrine disruptors and diabetes mellitus. Smart Health, 2021, 7(24):3.

[9] Xiuting Li, Jun Wang, Liangliang Zhao, et al. Research progress on the association between environmental endocrine disruptors and diabetes mellitus. Journal of Environment and Health, 2018, 35(5):5.

[10] Jian Zhang, Yanqiu Yu, Yu Yuan, Mei An He, Tangchun Wu. Prospective cohort study on the relationship between alcohol consumption and type 2 diabetes mellitus in middle-aged and elderly men. Chinese Journal of Preventive Medicine, 2017, 051(010):922-926.

[11] Bojie Wang. Study on the association between different hypoglycemic drugs and cancer risk. Journal of Clinical Medical Literatures Electronic, 2020(0A0):007.

[12] Xiuting Li, Jun Wang, Liangliang Zhao, et al. Research progress on the association between environmental endocrine disruptors and diabetes mellitus. Journal of Environment and Health, 2018, 35(05):93-97. (in Chinese)

[13] Yijuan Qiao, Junhong Leng. Research progress on the association between gestational diabetes mellitus and progeny obesity. China Chronic Disease Prevention and Control, 2020(1):3.

[14] Zhang Ying Zhang, Lin Zhang, XianLiang Zhou. FBN1 gene detection and genotype phenotype association in 55 patients with Marfan syndrome. 2022:29-29.

[15] Lijuan Ke, Yingfeng Zhuang, Bingrong Wang, et al. Relationship between gene detection and clinicopathological features of 90 cases of epithelial ovarian cancer. Fujian Journal of Medicine, 2022(003):044.