MINISTRY OF HIGHER EDUCATION,
SCIENCE AND INNOVATION OF THE
REPUBLIC OF UZBEKISTAN

TASHKENT STATE UNIVERSITY OF ECONOMICS

DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE: PROBLEMS, INNOVATIONS AND TRENDS

1st international scientific - practical conference

CONFERENCE PROGRAM

SEPTEMBER 11, TASHKENT 2024

IDENTIFYING HIGH RISK OF DIABETES BASED ON SURVEY RESEARCH QUESTIONNAIRE

Nurbek Fayzullayev,
Tashkent University of Information Technologies,
Uzbekistan,
nurbekfayzullayev89@gmail.com

***ABSTRACT* Diabetes is one of the most common diseases and the leading cause of death today. That's why a lot of work is being done all over the world to prevent it and save people with the disease. Later in this article we will briefly look at diabetes monitoring methods.**

KEYWORDS chronic diseases, healthcare systems, Self-management, patient’s condition, survey methodology.

INTRODUCTION

Patients with chronic diseases are spreading all over the world and in most cases, they are found among elderly people. The main reason for this is smoking, daily inactivity, alcohol and other reasons. The types of chronic diseases that are common nowadays are diabetes, heart disease, arthritis and cancer. Due to the complication of this disease and in many cases, it causes death, it is necessary for the patient to be under the constant supervision of a doctor and follow the daily changes. We will also analyze several methods in this article and find an effective solution in order to develop a mobile application. The reason is that almost everyone uses a phone, and it is much easier to maintain daily contact with the doctor. The next problem is which of the existing methods is easy and effective to develop a mobile application and requires to be under the control of cameras and it can cause financial problems for the patient. In other methods, the monitoring technologies can only be used by the doctor in the hospital, which causes the patient to waste more time. Therefore, we chose the Questionnaire-based method. The reason is that it is much cheaper and also allows you to chat with similar patients, and we want to add artificial intelligence to this method in the future.

METHODS

The Chronic Care Model (CCM). The Chronic Care Model was designed in 1996 and later revised in 1998 by Edward Wagner, MD, MPH. Wagner noticed individual providers rarely followed up on multiple chronic conditions, even though people with one chronic illness often have others. Wagner determined that physicians could manage chronic care more effectively and set out to create a system designed to manage chronic disease proactively. Wagner’s Chronic Care Model was further refined in 2003 by the Improving Chronic Illness Care (ICIC) program and became the framework we use today.

Internet of medical things (IoMT). The internet of medical things (IoMT) is the collection of medical devices and applications that connect to healthcare information technology systems through online computer networks. Medical devices equipped with Wi-Fi enable the machine-to-machine communication that is the basis of IoMT.

Examples of IoMT include the following:

- Using remote patient monitoring (RPM) for people with chronic diseases and long-term conditions.
- Tracking patient medication orders.
- Tracking the location of patients admitted to hospitals.
- Collecting data from patients' wearable mobile health devices.
- Connecting ambulances en route to medical facilities to healthcare professionals.

Types of IoMT devices:

- In-home IoMT
- Wearable IoMT
- Mobile IoMT
- Public IoMT
- In-hospital IoMT

Survey based applications and programs. Earlier detection of individuals at the highest risk of developing diabetes is crucial to avoid the disease’s prevalence and progression. Therefore, we aim to build a data-driven predictive application for screening subjects at a high risk of developing Type 2 Diabetes mellitus (T2DM) in the western region of Saudi Arabia. In this context, we designed and implemented a questionnaire-based cross-sectional study using conventional diabetes risk factors for studying the prevalence and the association between the outcomes and exposure (s).

We used the Chi-Squared test and binary logistic regression to analyze and screen the most significant diabetes risk factor for T2DM risk prediction. Synthetic Minority Over-sampling Technique (SMOTE), a class-balancer, was used to balance the cross-sectional data. We used the balanced class data to screen the best performing classification algorithm to classify patients at high risk of diabetes with a higher F1 Score. The best performing classifier’s hyper-parameters were further tuned using 10-fold cross-validation for achieving an improved F1 Score. Additionally, we validated our proposed model with the existing models built using the National Health and Nutrition Examination Survey (NHANES) dataset and Pima Indian Diabetes (PID) dataset. The results of the Chi-squared test and binary logistic regression showed that the exposures, namely Smoking, Healthy diet, Blood-Pressure (BP), Body Mass Index (BMI), Gender, and Region, contributed significantly ($p < 0.05$) to the prediction of the Response variable (subjects at high risk of diabetes).

The tuned two-class Decision Forest (DF) model showed better performance with an average F1score of 0.8453 ± 0.0268. Moreover, the DF based model adapted reasonably well in different diabetes dataset. An Application Programming Interface (API) of the tuned DF model was implemented and deployed as a web service at https://type2-diabetes-risk-predictor.herokuapp.com, and the implementation codes are available at https://github.com/SAH-ML/T2DM-Risk-Predictor.

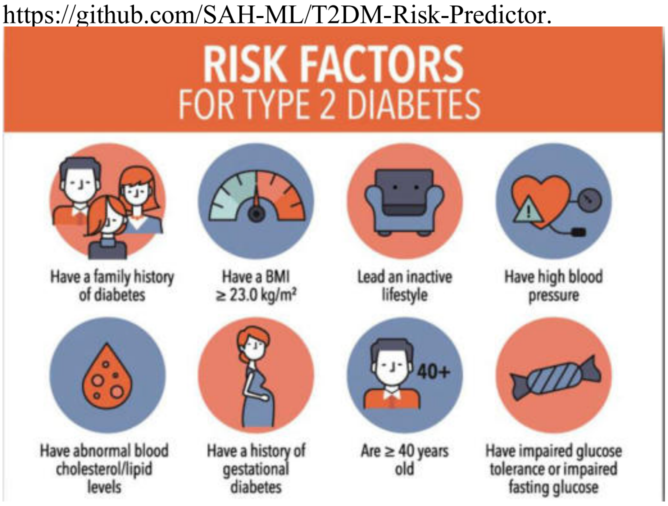
RISK FACTORS
FOR TYPE 2 DIABETES
Have a family history of diabetes
Have a BMI ≥ 23.0 kg/m²
Lead an inactive lifestyle
Have high blood pressure
Have abnormal blood cholesterol/lipid levels
Have a history of gestational diabetes
40+
Are ≥ 40 years old
Have impaired glucose tolerance or impaired fasting glucose

Figure-1. This picture shows the risk factors for type 2 diabetes and obesity is the main cause.

RESULTS

Cross-sectional study

In a cross-sectional survey, the researcher measures the exposure(s) in the population, the outcome and may study their relationship concurrently. The cross-sectional survey studies are typically economical and more rapidly implemented. These kinds of observational studies give us on-time information about the frequency of exposure (s) or outcomes. Thus, the information obtained from the retrospective cross-sectional study will be useful as a baseline for a cohort study. In our study, we intend to use a cross-sectional diabetes survey to estimate the prevalence of the disease in the western region of the Kingdom of Saudi Arabia (KSA). Furthermore, the strength of association between the outcomes and exposure (s), i.e., Odds Ratios (OR), will also be analyzed.

1) Survey research questionnaire

In this cross-section survey, we focused on the following closed-ended research questioners for identifying participants at high risk of diabetes:

a) Choose the region of your residence.

b) How old are you?

c) What is your Gender?

d) What is your Body Mass Index (BMI)? Use the height and weight table to find your BMI (The table will appear upon clicking the button).

e) What is your Waist size? Measured below the ribs (usually at the level of the navel)

f) Do you daily engage in at least 30 minutes of physical activity?

g) How often do you eat fruits and vegetables?

h) Have you ever taken hypertension medicine?

i) Have any members of your family been diagnosed with diabetes?

j) Have you ever had high blood glucose (for example, in a health examination, during an illness, during pregnancy)?

2) Dataset collection, transformation, and variable characterization

The cross-sectional survey dataset consists of 4896 subjects or participants (990 diabetic cases and 3906 non-diabetic cases). Among the ten diabetes risk factors considered for data analysis, region, gender, and age were demographic by nature. Region was categorized into ten different regions labeled as Abwa = 1, Jeddah = 2, Khulays = 3, Medina = 4, Masturah = 5, Mecca = 6, Rabigh = 7, Sabar = 8, Thual = 9, Yambu =10. The gender was coded as Female = 0, and Male = 1. Age was divided into three categories labeled as 0 =< 40 Years, 1 = 40 - 49 Years, 2 = 50 - 59 Years, and 3 => 60 Years. Body Mass Index (BMI) is calculated as body weight in kilograms divided by the square of body height in meters.

The BMI was divided in to three levels labeled as 0 =< 25 Kg/m2, 1 = 25 – 30 Kg/m2, 2 => 30 Kg/m2. Waist size for male and female divided into three categories each labeled as 0Male =< 94 cm (37'') or 0Female =< 80 cm (31.5''), 1Male = 94 – 102 cm (37'' – 40'') or 1Female = 80 -88 cm (31.5'' - 35''), 2Male => 102 cm (40'') or 2Female => 88 cm (35''). Physical activity is defined as daily at least 30 minutes of exercise or physical activity labeled as Yes = 0 and No = 1. A healthy diet indicated how regularly the subject eats fruits and vegetables labeled as ''0'' = every day and ''1'' = Not Every day. Subjects not undertaking medication for Blood Pressure (BP) labeled as ''0''. Whereas the subjects who were taking BP medicines were labeled as ''1''. Family history of diabetes is defined as ''do any members of the subject family been diagnosed with diabetes.'' The attribute Family history is categorized into three categories labeled as ''0'' = No family history of diabetes, ''1'' = Yes: Grandparents, and ''2'' = Yes: Parents. Smoking habits were categorized into two categories, non-smokers were labeled as ''0'' and smokers labeled as ''1''. Finally, the dataset included a response variable (diabetic and non-diabetic) based on subject exposure to fasting plasma glucose = 5.6 mmol/L [23] in a health examination or pregnancy.

We collected a large set of cross-sectional diabetes data over time and eventually developed a cross-sectional diabetes survey dataset comprising KAU subjects. The research questionnaire from the above-mentioned Q1 to Q9 includes the explanatory variables (predictors) and is categorical. While the attribute of high Fasting Blood Glucose level, was selected as a categorical response variable. The samples with a response of ''YES'' for the dichotomous class (high blood glucose) will fall in the category of ''high risk'' of diabetes, and conversely, the samples with a response of ''NO'' for the response variable will fall in the category of ''low risk'' of diabetes. Our cross-sectional survey dataset has been uploaded and is available at https://ieee-dataport.org/open-access/cross-sectional type-2-diabetes-survey-saudi-arabia-western-proven.

Pearson's chi-square test of independence

Step 1: Stating the Hypothesis

Null Hypothesis (Ho): There is no significant association between the two categorical variables {explanatory variables (risk factors) and the dependent variable (high or low risk of diabetes)}.

Alternate Hypothesis (H1): There is a significant association between the two categorical variables. {Explanatory variables (risk factors) and the dependent variable (high or low risk of diabetes)}.

Step 2: The Idea of the Chi-Square Test

How different is the observed count (our data) from the expected count when the explanatory and dependent variables are independent. Our cross-sectional data's observed count is shown in the respective Crosstabulation table of the exposure (s) and the outcome variable. The expected count was calculated using the formula shown below in ''equation 1'':

$$Expected\ Count = \frac{Column\ Total\ X\ Row\ Total}{Table\ Total} \quad (1)$$

Measurement of association between variables

The Chi-square is a tool to determine a significant association between the two categorical variables and should be followed with a statistical test to measure the strength of the relationship between the variables. For the Chi-square, the generally employed strength estimation test is the Cramer's V test. Cramer's V is a form of correlation and hence is interpreted similarly. The Cramer's V test was calculated using the formula shown below in ''equation 2'':

$$V = \sqrt{\frac{\varphi^2}{t}} = \sqrt{\frac{x^2}{n*t}} \qquad (2)$$

Here in ''equation 3'', ''t'' is the lesser of the total number of columns (c) minus one or the total number of rows (r) minus one, and ''n'' is equal to the sample size, then:

$$t=Minimum\{(r-1), or\ (c-1)\} \quad (3)$$

The Cramer's V test value ranges from 0 to 1. Where ''0'' means no correlation between the variable and on the other hand, ''1'' signifies a strong correlation between the variables, regardless of the sample size and dimensions of the contingency table.

DISCUSSION

In today's modern world, there is almost no area where telephones, computers and other technologies have not penetrated. Because with these devices, people's work has decreased, and their efficiency has increased. Therefore, large-scale work is being done in the field of medicine. is increasing day by day, and by itself, waiting in line in hospitals is increasing. Therefore, solving such problems with modern devices remains the optimal solution. Now we will consider a short and effective solution through one chronic disease. You can get information about it above. To treat a chronic disease, the doctor and the patient must be in constant contact and exchange information. This requires the patient to attend the hospital every day. Nowadays, almost everyone has a phone or a computer device. Because of this, we thought that creating a software application for monitoring chronic diseases was the most correct decision and developed several methods. The reason is that if the patient uses this application, he does not need to go to the hospital every day, he can analyze his condition with the doctor remotely while sitting at home. It will be possible to get it. Above, we studied the advantages and disadvantages of the main 3 methods and found the most suitable software application based on the Questionnaire, and thus we aimed to create an application for patients. The reason is that the patient's daily disease information and symptoms are stored and goes to the doctor in a graphic form, besides, it is possible to exchange information with patients suffering from the same disease through chat. Such software applications are in great demand all over the world because they provide more convenient opportunities for patients.

V. CONCLUSION

This article will serve as a foundation for our future work and develop the most effective and useful mobile application for patients

References

[1] https://www.chesco.org/357/Chronic-Diseases

[2] Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study

[3] N. Nasimova "Semi-Empiric Future Importance Dependency Evolution Method to Increase Diagnosis Accuracy of Chronic Disease", "Innovative Development in The Global Science", International Scientific and Technical Conference, Boston, USA, 2022, B. 96-100

[4] R. H. Nasimov, N. M. Nasimova "Xronik kasalliklarni monitoring qilishda ko'maklashuvchi dasturlar tahlili", "Kompyuter ilmlari va muhandislik Texnologiyalari" mavzusidagi Xalqaro miqyosidagi ilmiy-texnik anjuman materiallari to'plami, 14-15 oktyabr 2022-yil, 2-qism, B. 217-219

[5] https://www.mdpi.com/2306-5354/10/9/1031

[6] R. H. Nasimov, N. M. Nasimova "Surunkali kasalliklarni monitoring qilishda grafik usullardan samarali foydalanishning ahamiyati", "Интеллектуал ахборот ва коммуникация технологияларининг долзарб масалалари" мавзусидаги халқаро анжуман, Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети, 2022 йил 23-24 ноябрь

[7] R. H. Nasimov, N. M. Nasimova "Surunkali kasalliklarni kunlik monitoring qiluvchi mobil ilovani ishlab chiqish", "Matematik modellashtirish, axborot-kommunikatsiya texnologiyalarning dolzarb masalalari" mavzusida Respublika ilmiy-texnik anjuman, Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteta Nukus filiali, 2022 yil 17-18 noyabr.

[8] https://pubmed.ncbi.nlm.nih.gov/37760133/