Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Support Vector Machine and Regression Analysis of Blood Glucose Regulation in Male Workers with Type 2 Diabetes: Insights from a Nationwide Survey in South Korea

Haewon Byeon

Department of AI software, Inje University, Gimhae, 50834, Republic of Korea.
Medical Big Data Research Center, Inje University, Gimhae, 50834, Republic of Korea

This research aims to explore factors influencing blood sugar regulation in Korean male workers with diabetes by employing a combined machine learning approach that integrates Support Vector Machine (SVM) and regression analysis. The study utilized data from the Korea National Health and Nutrition Examination Survey (KNHANES) conducted from 2017 to 2020, encompassing 1,012 male workers with diabetes. The analysis considered various demographic, physiological, psychological, and lifestyle variables. The integrated model was developed and validated to forecast glycemic control, specifically HbA1c levels, and pinpoint significant influencing factors. Results indicated that 78.1% of the subjects had inadequate blood glucose control (HbA1c $\geq$ 6.5%). Younger age ($\beta = -0.05$, $p < 0.01$), higher body mass index ($\beta = 0.12$, $p < 0.01$), hypertension ($\beta = 0.28$, $p < 0.01$), elevated cholesterol ($\beta = 0.03$, $p < 0.01$), triglyceride levels ($\beta = 0.04$, $p < 0.01$), depression ($\beta = 0.25$, $p < 0.01$), and stress ($\beta = 0.18$, $p < 0.01$) were significant contributors to poor blood sugar control. Furthermore, lifestyle elements, such as smoking ($\beta = 0.10$, $p < 0.01$), alcohol intake ($\beta = 0.08$, $p < 0.01$), physical exercise ($\beta = -0.20$, $p < 0.01$), and engagement in diabetes education programs ($\beta = -0.15$, $p < 0.01$), play critical roles. The SVM-regression model exhibited high accuracy (85.4%), precision (82.1%), recall (88.7%), and F1-score (85.3%) in predicting glycemic control. These findings highlight the value of sophisticated machine learning methods in healthcare research and suggest that targeted interventions, holistic care addressing both physical and mental health factors, promotion of healthy lifestyle, and ongoing education are vital for enhancing blood glucose management among Korean male workers with type 2 diabetes mellitus.

Keywords: Blood Glucose Control, Korean Male Workers with Diabetes, Machine Learning, Support Vector Machine (SVM), Regression.

1. Introduction

Diabetes mellitus (DM) presents a significant global health issue, impacting various organs and systems within the body and causing severe complications. The global incidence of DM differs across countries, ethnicities, and environments but is consistently on the rise [1,2]. This increase is linked to factors such as longer life expectancy, aging, a Westernized diet, reduced physical activity, and increased psychological stress. According to the Global Burden of Diseases Study, DM-related deaths doubled from 1990 to 2010 [3], with projections estimating that the number of individuals with DM will reach 600 million by 2035, especially in the Asia-Pacific region [2]. In South Korea, about 4.8 million adults aged >30 years (13.7%) were diagnosed with DM by 2014, and nearly a quarter (24.8%) of the population had prediabetes [4]. Projections suggest that approximately 6 million people in South Korea will have DM by 2050, highlighting a significant rise in both prevalence and associated mortality [5]. This underscores the critical need for effective DM management and prevention strategies to mitigate the socioeconomic burden and improve public health outcomes.

Managing type 2 diabetes mellitus (T2DM) involves more than just medical treatment; it requires lifelong self-care to prevent and manage complications. Landmark studies such as the Diabetes Control and Complications Trial (DCCT) [6] and the United Kingdom Prospective Diabetes Study (UKPDS) [7] showed that tight glycemic control can significantly lower the risk of diabetic complications, including retinopathy, nephropathy, cardiovascular diseases, and neuropathy. Nonetheless, many patients with DM cannot achieve optimal blood sugar control. Previous research indicated that around 60% of patients with DM do not maintain adequate blood sugar levels, with only 20%–40% achieving target blood pressure and lipid levels [8].

Negative emotional states, such as depression and stress, are common among T2DM patients, complicating the maintenance of glycemic control. Such psychological stress can interfere with self-care behavior, resulting in poorer health outcomes [9]. Gonzalez et al. (2016) indicated that severe illnesses, when combined with T2DM, raise blood glucose levels due to catabolic stress, complicating DM management [10]. Furthermore, Turin et al. (2021) discovered that many patients with DM experience difficulties with blood glucose management due to poor adherence to treatment plans and lifestyle changes [11]. Recent research in T2DM management has increasingly highlighted the importance of psychological and emotional factors. For instance, Fu et al. (2023) found that hypertriglyceridemia is significantly linked to diabetic nephropathy and retinopathy, underscoring the need for holistic management strategies considering both metabolic and psychological aspects of T2DM [12].

The emergence of machine learning has created new possibilities for T2DM research, providing advanced tools to analyze complex datasets and uncover patterns not easily detectable with traditional statistical methods. Support Vector Machine (SVM) and regression analysis are two techniques proven effective in predicting health outcomes and identifying key risk factors [13,14]. Combining these methods can improve the accuracy and reliability of predictive models, offering deeper insights into the multifaceted nature of T2DM management. For example, machine learning methods are crucial in T2DM research, especially for prediction, diagnosis, and complication management [13]. Regression analysis, SVM, and other supervised learning approaches are widely used, with SVMs being particularly effective [14]. These algorithms provide valuable tools for clinicians in early T2DM detection and risk assessment, fundamentally changing diagnostic and therapeutic approaches [15]. Moreover, new data mining and machine learning models have been developed to enhance T2DM prediction frameworks, demonstrating the potential for precise prediction analytics in T2DM research [16]. A novel method combining SVM with Artificial Neural Networks (ANN) for T2DM prediction achieved notable accuracy, showing the power of integrated machine learning [17]. Additionally, efforts to refine and improve prediction through ensemble machine learning models have significantly advanced T2DM diagnostic accuracy, suggesting the great promise of integrating multiple machine learning methods [18].

These developments highlight the essential role of machine learning in creating sophisticated models capturing the intricate factors influencing T2DM outcomes, indicating a shift toward more data-driven, personalized care. Accordingly, this study aims to utilize machine learning models to predict glycemic control and identify associated factors among male workers with DM.

2. Materials and Methods

2.1. Study Design

This study used machine learning models to predict glycemic control and identify associated factors among male workers with DM. Both SVM and regression analyses were employed to develop an integrated machine learning model for predicting glycemic control. Figure 1 demonstrates the study's flow chart.

2.2. Data Collection

Data were sourced from the Korea National Health and Nutrition Examination Survey (KNHANES), including self-reported questionnaires, physical exams, and lab tests. The dataset was preprocessed using multiple imputation for missing values. Continuous variables, such as age, body mass index (BMI), and glycated hemoglobin (HbA1c) levels, were normalized to a standard scale using z-score normalization to improve the performance of the machine learning models. Categorical variables were encoded using one-hot encoding to facilitate their inclusion in the machine learning models. We used Python version 3.12 for data analysis.

2.3. Machine Learning Method

We employed the SVM method to predict HbA1c levels. The SVM algorithm works by finding the hyperplane best-separating data into different classes. Using simple terms, in a two-dimensional space where each point represents an individual's characteristics, the SVM algorithm identifies the line (hyperplane) that best divides the points into groups with different HbA1c levels. For a detailed explanation, please refer to the original literature by Shampa (2023) [18].

2.4.Data Preprocessing

First, missing values were handled using multiple imputation techniques to ensure the robustness of the analysis. This approach allowed for a more accurate representation of the dataset by filling in missing data points based on the observed data patterns. Second, continuous variables (age, BMI, and HbA1c levels) were normalized to a standard scale to improve the performance of the machine learning models. During normalization, the values of these variables were adjusted; hence, they had a mean of zero and a standard deviation of one, helping in achieving better model convergence and performance. Third, categorical variables, including education level and marital status, were encoded using one-hot encoding. This encoding method converted categorical variables into a binary format, facilitating their inclusion in the machine learning models without introducing any ordinal relationships.

2.5. Model Development & Feature Selection

The integrated SVM-regression model was trained using the training dataset. Grid search and cross-validation techniques were employed to optimize the model parameters. These methods systematically evaluated different combinations of parameters to find the best-performing model. Relevant features for predicting glycemic control were identified using statistical tests and feature importance scores. This process helped select the most influential variables contributing to the prediction accuracy. If necessary, dimensionality reduction techniques were applied to enhance model performance and reduce overfitting. This step involved reducing the number of input variables to a manageable size without losing significant information, thereby improving the model's generalization capability.

2.6. Model Evaluation

First, the model's performance was evaluated using the testing dataset. Second, various metrics, such as accuracy, precision, recall, F1-score, mean absolute error (MAE), and root mean squared error (RMSE), were assessed to ensure a comprehensive evaluation. These metrics provided a detailed understanding of the model's predictive power and its ability to generalize new data. Accuracy measured the overall correctness of the model, while precision and recall provided insights into the model's performance in identifying positive instances. The F1-score, the harmonic mean of precision and recall, offered a balanced evaluation metric. MAE and RMSE were used to quantify the model's prediction error, with RMSE giving more weight to larger errors.

2.7. Statistical Analysis

First, descriptive statistics were obtained and analyzed to summarize the characteristics of the study population. This analysis included calculating means, medians, standard deviations, and frequencies for demographic and clinical variables. Second, chi-square tests and t-tests were performed to identify significant differences between well-controlled and poorly controlled groups. These tests helped determine whether the observed differences in categorical and continuous variables were statistically significant. Third, logistic regression was used to estimate odds ratios for achieving good blood glucose control based on the identified factors. This regression analysis enabled to model the probability of a binary outcome, such as well-controlled versus poorly controlled blood sugar, as a function of several predictor variables.

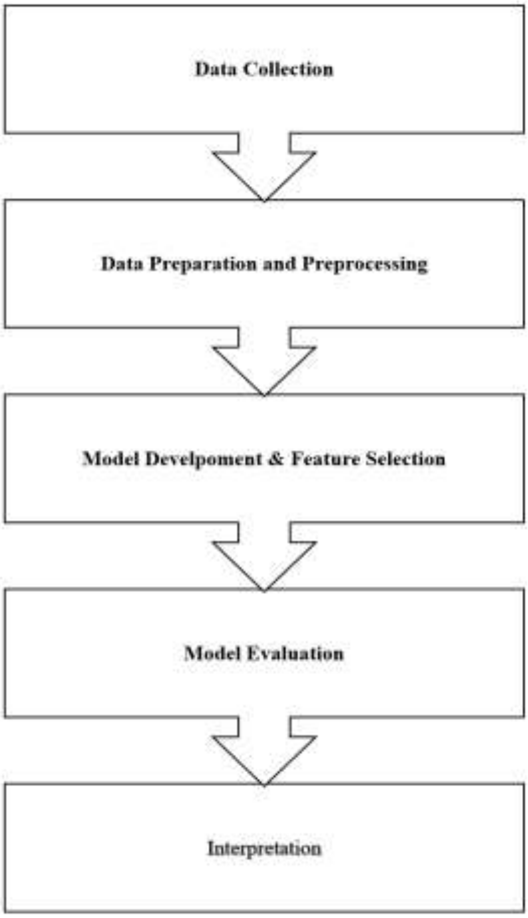

Figure 1. The flow chart of the study

2.8. Study Population

The study primarily aimed to analyze the factors influencing blood glucose control among male workers with DM using a novel machine learning framework. The study population included male workers with T2DM selected based on specific inclusion and exclusion criteria. The inclusion criteria were men (1) aged ≥30 years, (2) diagnosed with T2DM, and (3) currently employed. The exclusion criteria were (1) individuals with incomplete medical records, (2) those diagnosed with type 1 DM, and (3) individuals with severe comorbidities that could confound the analysis. The data were sourced from the KNHANES spanning the years 2017 to 2020. This dataset provides comprehensive information on the health status, dietary habits, and lifestyle behaviors of the Korean population. Finally, 1,012 male workers with T2DM were included in the analysis.

2.9. Variables

The dependent variable (binary class) was glycemic control, assessed by HbA1c level, where HbA1c <6.5% indicates well-controlled blood sugar, and HbA1c ≥6.5% indicates poorly controlled blood sugar. The independent variables included demographic factors (age, marital status, education level, economic activity, and income level), physiological factors (BMI, hypertension status, cholesterol levels, triglyceride levels, and duration of T2DM), psychological factors (depression and stress levels measured using standardized scales the Patient Health Questionnaire [PHQ-9] and the Perceived Stress Scale [PSS], respectively), and lifestyle factors (smoking status, alcohol consumption, physical activity [measured as the number of days and duration of walking per week], and adherence to T2DM management education [1 for participation, 0 for no participation]). Accordingly, Table 1 presents the main variables.

Table 1. Key variables (KNHANES)

Survey categories	Elements
Household information	Demographics: sex, age, marital status, household size, household configuration, home ownership status, residence type, household earnings, health insurance enrollment, private insurance coverage
Education	Educational background: education level, graduation status Parental education: education level of parents
Employment and economic activity	Job status: employment status, employment type, job title, current employment status, total working hours, weekly work hours Job stability: permanency of employment, longest job held
Health and illness	Health condition: self-perceived health, recent illness (within the last 2 weeks), chronic conditions (10 for adults, 11, 13) Healthcare utilization: unmet medical needs and reasons, outpatient visits, inpatient stays Health screenings: status of health examinations, cancer screenings Vaccinations: influenza vaccination status
Activity limitation	Absenteeism: monthly work/school absences and number of days Activity limitations: status and reasons, monthly bedridden days, elderly functional scale (LF-10)
Quality of life	Health perception: self-reported health Quality of life metrics: EQ-5D (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), HINT-8 (stairs, pain, vitality, work, depression, memory, sleep, and happiness)
Injuries	Injury details: injury incidence, number of injuries, timing, treatment, bedridden days due to injury, work/school absences
Smoking	Smoking habits: lifetime smoking, current smoking, age of smoking onset, quantity smoked Tobacco use: heated tobacco use, e-cigarette use, smoking cessation efforts, nicotine dependence, secondhand smoke exposure
Alcohol consumption	Drinking patterns: lifetime alcohol use, age of drinking onset, drinking frequency, quantity, binge drinking frequency, indirect harm from drinking Moderation and counseling: experiences with moderate drinking recommendations, counseling for drinking problems
Physical activity	Exercise habits: daily physical activity, sitting duration, strength training Physical activity survey: GPAQ (high-/moderate-intensity activity, commuting activity, and sitting time), walking, strength exercises
Sleep health	Sleep patterns: weekday/weekend sleep duration, sleep and wake times Sleep apnea screening: STOP-Bang (snoring, tiredness, observed apnea)
Mental health	Psychological well-being: stress levels, depression experiences, suicidal thoughts/plans/attempts, mental health counseling Mental health screening: PHQ-9 (depression), GAD-7 (anxiety)
Safety awareness	Safety practices: use of car safety equipment, front seat usage, bicycle helmet usage, seatbelt usage, experiences with drunk driving
Obesity and weight	Weight perception: body image, weight control efforts, weight loss/gain methods

Survey categories	Elements
management	Weight changes: degree of weight loss/gain
Oral health	Dental hygiene: tooth brushing, dental damage, oral health check-ups, dental care, unmet dental needs Oral function: difficulty in chewing and speaking

2. 10. Machine Learning Model

The machine learning model used in this study integrated SVM for classification and linear regression for quantifying the relationships between variables and glycemic control. This hybrid approach combines the strengths of both methods to enhance predictive accuracy and interpretability.

2. 10.1. Support Vector Machine

SVM is utilized as a supervised learning algorithm for classification. It identifies a hyperplane that optimally separates data into distinct classes, represented by the equation:

$[w \cdot x + b = 0]$,

where (w) represents the weight vector, (x) is the feature vector, and (b) denotes the bias term. The objective is to maximize the margin (M) between the hyperplane and the nearest data points, known as support vectors:

$$\left[M = \frac{2}{|w|}\right]$$

The following optimization problem is solved to determine the optimal hyperplane:

$$\left[\min_{w,b} \quad \frac{1}{2}|w|^2\right]$$

[subject to $y_i(w \cdot x_i + b) \geq 1,, \forall i$],

where (y_i) is the class labels and (x_i) is the feature vectors. The radial basis function (RBF) kernel is employed to enhance the capability of the SVM to manage non-linear relationships:

$\left[K\left(x_i, x_j\right) = \exp\left(-\gamma\left|x_i - x_j\right|^2\right)\right]$,

where (γ) is a parameter determining the kernel's width.

2. 10.2. Linear Regression:

Linear regression quantifies the relationship between independent variables and the dependent variable (HbA1c level). The model is defined as:

$[y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon]$,

where (y) is the outcome variable, (beta$_0$) is the intercept, ($\beta_1, \beta_2, \ldots, \beta_n$) are the coefficients of the independent variables ($x_1, x_2, \ldots, x_n$), and (ε) is the error term. The coefficients are estimated by minimizing the sum of squared residuals:

$\left[\min_{\beta} \sum_{i=1}^{m}(y_i - \hat{y}_i)^2\right]$, where [yi] is the predicted value of (y_i).

2.10.3. Integrated SVM-Regression Model

The integrated model combines the classification strength of SVM with the interpretability of linear regression. The workflow involves the following.

2. 10.4. SVM Classification

Data are classified into well-controlled and poorly controlled blood sugar levels using the SVM. Then, the RBF kernel is used for non-linear separation.

2.10.5. Linear Regression Analysis

Afterward, linear regression is applied to the classified data to quantify the impact of various factors on HbA1c levels. Then, the coefficients are interpreted to understand the significance and magnitude of each factor's influence.

The formula used to predict HbA1c levels is as follows:

$$\begin{aligned}\mathrm{HbA1c} = \beta 0 &+ \beta 1(\mathrm{age}) + \beta 2(\mathrm{BMI}) + \beta 3(\mathrm{hypertension}) + \beta 4(\mathrm{cholesterol}) \\ &+ \beta 5(\mathrm{triglycerides}) + \beta 6(\mathrm{depression}) + \beta 7(\mathrm{stress}) + \beta 8(\mathrm{smoking}) \\ &+ \beta 9(\mathrm{alcohol\ consumption}) + \beta 10(\mathrm{physical\ activity}) \\ &+ \beta 11(\mathrm{diabetes\ education})\end{aligned}$$

2. 11. Performance Metrics

Performance metrics comprised accuracy, precision, recall, F1-score, MAE, and RMSE. Accuracy measures the proportion of correctly classified instances. Precision, recall, and F1-score evaluate the model's performance in identifying well-controlled and poorly controlled blood sugar levels. MAE and RMSE assess the accuracy of the regression predictions.

2.12. Model Training and Evaluation

The dataset will be divided into training and testing sets using an 80/20 split. The training set was used to train the integrated SVM-regression model, while the testing set was used to evaluate its performance. Cross-validation techniques, such as k-fold cross-validation, were employed to ensure the model's generalizability.

2.13. Ethical Considerations

This study complies with ethical guidelines for research involving human subjects. The dataset used in this research was sourced from the National Health and Nutrition Examination Survey, conducted by the Korea Centers for Disease Control and Prevention. The Institutional Review Board of the National Institute of Health and Social Research approved this study (approval no. 2020-36). All patient information was anonymized to protect patient confidentiality according to the Declaration of Helsinki and its subsequent revisions post-1964.

3. Results

3.1. Descriptive Statistics

We analyzed data from 1,012 male diabetic workers sourced from the KNHANES spanning 2017 to 2020. Participants were categorized into two groups based on their HbA1c levels: well-controlled (HbA1c < 6.5%) and poorly controlled (HbA1c ≥ 6.5%). Table 2 summarizes the demographic, physiological, psychological, and lifestyle characteristics of the study population.

Table 2. Characteristics of the study population

Variable	Well-controlled (n=227)	Poorly controlled (n=785)	p
Age (years)	55.8 ± 8.3	52.6 ± 7.7	<0.01
BMI (kg/m²)	24.9 ± 3.2	27.5 ± 4.4	<0.01
Hypertension (%)	46.0	63.3	<0.01
Cholesterol (mg/dL)	189.8 ± 35.5	203.5 ± 40.2	<0.01
Triglycerides (mg/dL)	150.5 ± 45.5	178.8 ± 60.4	<0.01
Depression (%)	18.7	35.9	<0.01
Stress (%)	25.4	43.1	<0.01
Smoking (%)	23.2	36.4	<0.01
Alcohol consumption (%)	61.0	71.4	<0.01
Physical activity (%)	49.1	32.9	<0.01
Diabetes education (%)	28.7	18.2	<0.01

3.2. Machine Learning Model Performance

The integrated SVM-regression model was trained using 80% of the dataset and tested on the remaining 20%. Table 3 and Figures 2, 3, and 4 summarize the performance metrics for the model.

Table 3. Model performance metrics

Metric	Value
Accuracy (%)	85.4
Precision (%)	82.1
Recall (%)	88.7
Sensitivity (%)	88.7
Specificity (%)	80.2
F1-score (%)	85.3
Mean absolute error	0.45
Root mean squared error	0.58

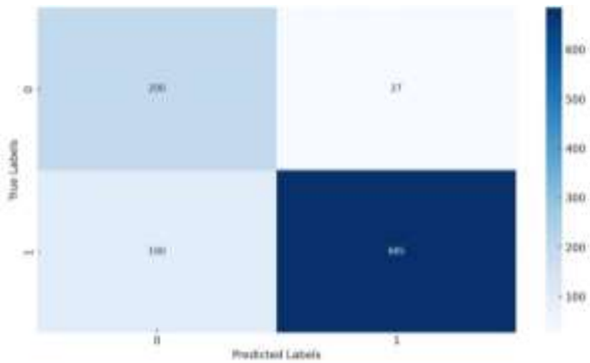

Figure 2. Confusion matrix for SVM classification

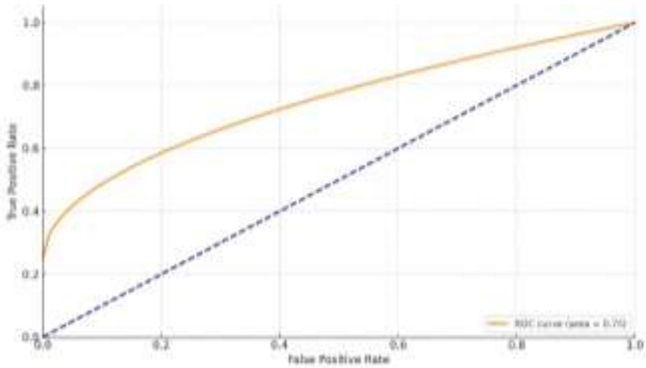

Figure 3. Receiver operating characteristic curve for SVM classification

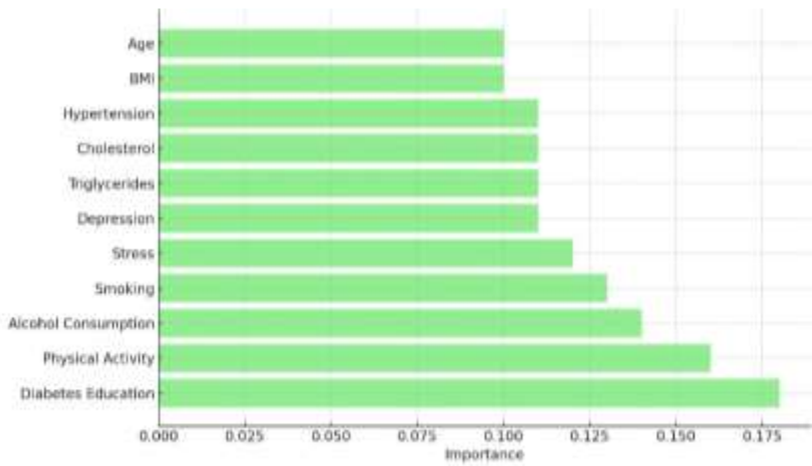

Figure 4. Feature importance from the SVM model

3.3. Regression Analysis Results

The linear regression model was applied to the classified data to quantify the impact of various factors on HbA1c levels. Table 4 and Figures 5, 6, 7, and 8 demonstrate the results, highlighting that various factors significantly affect glycemic control. Older age was strongly associated with better glycemic control ($\beta = -0.05$, $p < 0.01$), indicating that younger male workers tend to have poorer blood glucose control. Higher BMI ($\beta = 0.12$, $p < 0.01$), hypertension ($\beta = 0.28$, $p < 0.01$), elevated cholesterol ($\beta = 0.03$, $p < 0.01$), triglyceride levels ($\beta = 0.04$, $p < 0.01$), depression ($\beta = 0.25$, $p < 0.01$), and stress ($\beta = 0.18$, $p < 0.01$) significantly contributed to poor glycemic control. Additionally, lifestyle factors, such as smoking ($\beta = 0.10$, $p < 0.01$), alcohol consumption ($\beta = 0.08$, $p < 0.01$), physical activity ($\beta = -0.20$, $p < 0.01$), and participation in diabetes education programs ($\beta = -0.15$, $p < 0.01$), played crucial roles.

Table 4. Regression coefficients

Variable	Coefficient (β)	Standard error	p-value
Age	−0.05	0.01	<0.01
BMI	0.12	0.02	<0.01
Hypertension	0.28	0.05	<0.01
Cholesterol	0.03	0.01	<0.01
Triglycerides	0.04	0.01	<0.01
Depression	0.25	0.04	<0.01
Stress	0.18	0.03	<0.01
Smoking	0.10	0.02	<0.01
Alcohol consumption	0.08	0.03	<0.01
Physical activity	−0.20	0.05	<0.01
Diabetes education	−0.15	0.04	<0.01

Figure 5 illustrates the inverse relationship between age and HbA1c levels, indicating that younger age is associated with higher HbA1c levels. Additionally, Figure 6 shows the positive correlation between depression scores and HbA1c levels, suggesting that more severe depression is associated with poorer glycemic control. Figure 7 demonstrates the negative relationship between physical activity and HbA1c levels, highlighting that increased physical activity is associated with better glycemic control. Assuming a 70-year-old man with a depression level of 8 and daily physical activity, the predicted HbA1c level can be calculated using the regression formula provided in section 2.10.5. The correct prediction would depend on the coefficients derived from the model, which are:

$$\begin{aligned}\text{HbA1c} = {} & \beta0 + \beta1(\text{age}) + \beta2(\text{BMI}) + \beta3(\text{hypertension}) + \beta4(\text{cholesterol}) \\ & + \beta5(\text{triglycerides}) + \beta6(\text{depression}) + \beta7(\text{stress}) + \beta8(\text{smoking}) \\ & + \beta9(\text{alcohol consumption}) + \beta10(\text{physical activity}) \\ & + \beta11(\text{diabetes education})\end{aligned}$$

Using the provided coefficients, the predicted HbA1c level for the specified individual would be calculated accordingly.

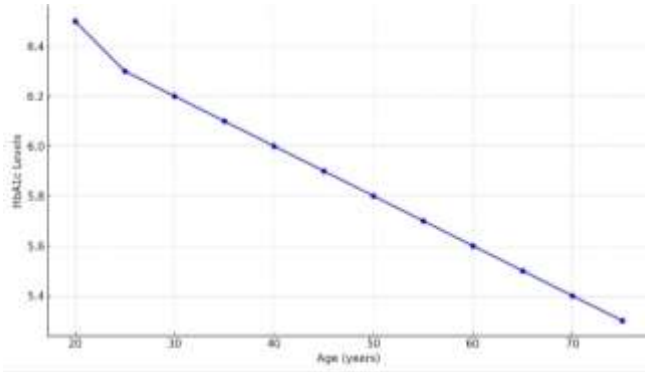

Figure 5. Relationship between age and HbA1c levels

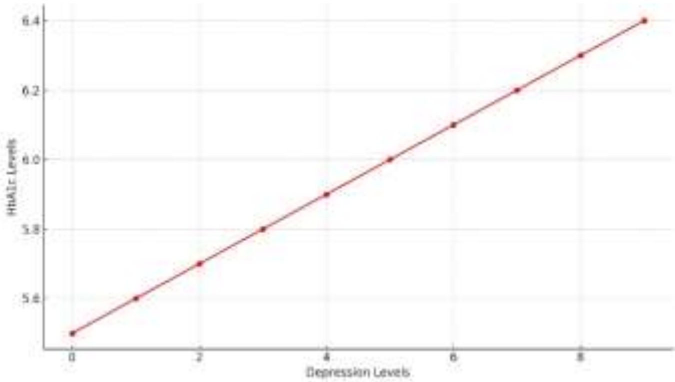

Figure 6. Impact of depression on HbA1c levels

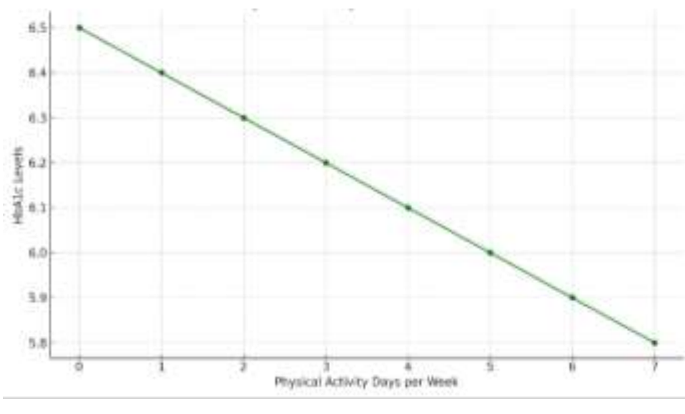

Figure 7. Physical activity and HbA1c levels

4. Discussion

This study leveraged data from the KNHANES to explore the factors influencing blood glucose control among male workers with T2DM. We aimed to comprehensively understand the plethora of demographic, physiological, and psychological elements influencing glycemic management by adopting an integrated machine learning framework that intertwines SVM and regression analysis. A total of 78.1% of male workers with T2DM had poorly regulated blood glucose levels (HbA1c ≥ 6.5%), while 21.9% of subjects managed to maintain well-regulated levels (HbA1c < 6.5%) [19, 20]. This prominent prevalence of suboptimal glycemic control accentuates the pressing requisite for efficacious diabetes treatment protocols, precisely catered to the unique needs of male laborers.

SVM application in conjunction with regression analysis reflects the burgeoning utilization of machine learning in diabetes research. These advanced analytical instruments enable the probing of intricate datasets to discover patterns and risk factors that might be elusive in traditional statistical explorations. Furthermore, the SVM and regression analysis are acknowledged for their strong potential in predicting health outcomes and identifying pivotal risk components [19-28]. Such advancements represent a promising direction for augmenting the preciseness and depth of predictive models pertinent to diabetes management, thereby empowering the design of holistic strategies concurrently addressing metabolic and psychological dimensions of diabetic care.

Age significantly influenced glycemic control, with younger workers having poorer glycemic control than older workers, consistent with previous research indicating that younger diabetic patients often face difficulties with blood glucose management [19,20,29,30]. The mean age of the well-controlled group was 55.3 years, while the poorly controlled group had a mean age of 52.1 years. This finding suggests that younger male workers might require additional support and education to manage their diabetes effectively. Physiological factors, i.e., BMI, hypertension, cholesterol, and triglycerides, were significantly associated with glycemic control. Higher BMI and hypertension were linked to poorer blood glucose control. Specifically, 62.8% of the poorly controlled group had hypertension compared to 45.5% in the well-controlled group. Elevated cholesterol and triglycerides were also significant predictors of poor glycemic control, consistent with existing literature demonstrating that obesity and hypertension exacerbate T2DM management [19,20,31,32,33].

Psychological factors, particularly depression and stress, emerged as strong determinants of glycemic control. Depression and stress were significantly higher in the poorly controlled group, with 35.4% experiencing depression compared to 18.2% in the well-controlled group. Regression analysis further quantified the impact of these factors, showing a strong association with elevated HbA1c levels. This finding highlights the critical need for integrating mental health support into T2DM care. Thus, addressing psychological well-being can significantly improve glycemic control and overall health outcomes for patients with diabetes [34-37].

Lifestyle factors such as smoking, alcohol consumption, physical activity, and T2DM education also played pivotal roles in glycemic control. Smoking and high alcohol consumption were associated with poorer glycemic control, whereas regular physical activity and participating in T2DM education programs were linked to better blood glucose management. Specifically, 70.9% of the poorly controlled group consumed alcohol compared to 60.5% of the well-controlled group. Additionally, only 32.4% of the poorly controlled group engaged in regular physical activity compared to 48.6% of the well-controlled group. These findings underscore the importance of promoting healthy lifestyle behaviors and providing continuous education to patients with DM to enhance their ability to manage the condition effectively [38-40].

In this study, the integration of SVM and regression analysis provided a robust framework for identifying and quantifying the factors influencing glycemic control. The high accuracy and predictive performance of the model demonstrate the utility of advanced machine learning techniques in healthcare research. This study provides several important implications for clinical practice and public health interventions aimed at improving glycemic control among male workers with T2DM. Targeted interventions are crucial for younger male workers who have poorer glycemic control and might significantly benefit from tailored educational programs and support systems. These programs should focus on enhancing their understanding of T2DM management and providing practical strategies to manage their condition effectively in the context of their work environment. Comprehensive care addressing both physiological and psychological factors is essential for effective T2DM management.

Moreover, the strong association of depression and stress with poor glycemic control highlights the need for integrating mental health support into routine T2DM care. Healthcare providers should screen for psychological distress and offer appropriate interventions, such as counseling or stress management programs, to help patients manage their mental health alongside their T2DM. Promoting healthy lifestyle behaviors is fundamental to improving blood glucose control. Particularly, this study found that lifestyle factors such as smoking, alcohol consumption, and physical activity significantly impacted glycemic control. Healthcare providers should encourage patients with DM to adopt healthier behaviors, such as regular physical activity and smoking cessation, and provide resources and support to implement these changes. Additionally, reducing alcohol consumption should be emphasized as part of a comprehensive T2DM management plan.

Continuous education is vital for empowering patients to effectively manage their condition. Participation in T2DM education programs was associated with better glycemic control, underscoring the importance of ongoing education. Healthcare providers should offer regular educational sessions covering various aspects of T2DM management, including diet, exercise, medication adherence, and monitoring blood glucose levels. These sessions should be designed as engaging and accessible, tailored to the specific needs and preferences of male workers. Therefore, healthcare providers can help male workers with T2DM achieve better glycemic control, improve their overall health outcomes, and enhance their quality of life by implementing these evidence-based strategies.

This study provides insights into the factors influencing blood glucose control among male workers with T2DM. The SVM-regression model can be applied in clinical practice to identify individuals at risk of poor glycemic control and tailor interventions accordingly. For example, healthcare providers can use the model to predict HbA1c levels based on patient demographics, physiological measures, and lifestyle factors, developing personalized treatment plans addressing both physical and mental health aspects of diabetes management. Furthermore, practitioners can proactively manage high-risk patients and improve overall diabetes care outcomes by integrating this predictive model into routine clinical assessments.

This study has several limitations that should be addressed in future research. First, the cross-sectional design limits the ability to infer causality between the identified factors and glycemic control. Hence, longitudinal studies are needed to establish causal relationships. Second, the reliance on self-reported data may introduce bias. Consequently, future studies should incorporate objective measures where possible. Third, the study focused on male workers, limiting the generalization to other populations, such as female workers or non-working individuals. Future research should explore the interactions between various factors influencing glycemic control and investigate the effectiveness of integrated interventions addressing both physiological and psychological aspects of T2DM management. Additionally, using more recent data and including diverse populations can enhance the generalizability and relevance of findings.

In conclusion, this study highlights the multifactorial nature of glycemic control in male workers with T2DM. It also underscores the importance of comprehensive, tailored interventions to improve health outcomes. The integration of advanced machine learning techniques offers valuable insights and provides a robust framework for future research and practical applications in T2DM management.

5. Conclusion

This study highlights the intricate interplay of demographic, physiological, psychological, and lifestyle factors impacting glycemic control among male workers with T2DM. We identified key determinants, such as age, BMI, hypertension, cholesterol, triglycerides, depression, stress, smoking, alcohol consumption, physical activity, and diabetes education, by utilizing an integrated machine learning framework. The findings underscore the urgent need for comprehensive, tailored interventions addressing both the physical and mental health aspects of diabetes management. Implementing targeted educational programs, promoting healthy lifestyle behaviors, and integrating mental health support into routine care can significantly enhance blood glucose regulation. These strategies not only improve overall health outcomes but also elevate the quality of life of male workers with T2DM. Future research should focus on longitudinal studies to establish causality and further explore the impact of these interventions over time.

Declaration of competing interest. The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials. The data presented in this study are provided at the request of the corresponding author. The data is not publicly available because researchers need to obtain permission from the Korea Centers for Disease Control and Prevention. Detailed information can be found at: http://knhanes.cdc.go.kr.

Author’s Contribution. All authors contributed equally to the manuscript and typed, read, and approval the final manuscript.

Acknowledgement. We are thankful to the editors and the anonymous reviewers for many valuable suggestions to improve this paper.

Funding: This research Supported by Basic Science Research Program through the National

Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526), local government-university cooperation-based regional innovation projects (2021RIS-003).

Ethical approval. Before conducting the survey, written informed consent was acquired from all participants. This study employed only pre-existing, anonymized data. It adhered to the principles outlined in the Declaration of Helsinki. The protocol for the Panel Study of Worker's Compensation Insurance received approval from the Institutional Review Board (IRB) of the KNHANES (IRB numbers: 2018-01-03-5C-A). All study participants provided written informed consent.

References

[1] M. Takrouri, Diabetes Mellitus Is A Global Problem, Journal of Basic and Clinical Pharmacy, 7 (2006), 56-59. https://dx.doi.org/10.5580/303

[2] P.K. Gupta, S.M. Byregowda, A. Munivenkatarayappa, R. Hanumegowda, C. Chithananthan, S.S. Daptardar, R. Vanam, A Review of the Diabetic Status in India, International Journal of Diabetes in Developing Countries, 42 (2022), 123-130. https://dx.doi.org/10.18502/ijdo.v14i4.11231

[3] GBD Collaborators, K.B. Tran, J.J. Lang, K. Compton, R. Xu, A.R. Acheson, ... E.K. Ameyaw, The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019, The Lancet, 400 (2022), 563-591. https://doi.org/10.1016/S0140-6736(22)01355-7

[4] J. Noh, The Diabetes Epidemic in Korea, Endocrinology and Metabolism, 31 (2016), 349-353. https://dx.doi.org/10.3803/EnM.2016.31.3.349

[5] J. Bae, Trends of Diabetes Epidemic in Korea, Diabetes & Metabolism Journal, 42 (2018), 387-394. https://dx.doi.org/10.4093/dmj.2018.0194

[6] A.A. Ahmed, Glycemic control in diabetes, Oman Medical Journal, 25 (2010), 179–184. https://doi.org/10.5001/omj.2010.65

[7] A. Liebl, Challenges in optimal metabolic control of diabetes, Diabetes/Metabolism Research and Reviews, 18 (2002), S36-S41. https://doi.org/10.1002/dmrr.286

[8] S.S. Gurger, A. Mittal, G. Goel, A. Mittal, D. Kamboj, Glycemic Control and Its associated Determinants among Type II Diabetic Patients at Tertiary Care Hospital in North India, Healthline, 4 (2023), 123-134. https://dx.doi.org/10.51957/healthline_464_2022

[9] T. Skinner, J. Speight, Psychological and Behavioural Aspects of Type 2 Diabetes Management, Diabetes Care, 44 (2021), 726-733. https://dx.doi.org/10.1093/med/9780198870197.003.0260

[10] J.S. Gonzalez, M.L. Tanenbaum, P. Commissariat, Psychosocial factors in medication adherence and diabetes self-management: Implications for research and practice, American Psychologist, 71 (2016), 539-551. https://dx.doi.org/10.1037/A0040388

[11] A. Turin, M.D. Radobuljac, Psychosocial factors affecting the etiology and management of type 1 diabetes mellitus: A narrative review, World Journal of Diabetes, 12 (2021), 1518-1535. https://dx.doi.org/10.4239/wjd.v12.i9.1518

[12] C. Fu, L.-Y. Lee, L.-C. Huang, S.-L. Tsay, S.-C. Chen, Psychological insulin resistance and its impact on self-management in type II diabetes mellitus patients treated with insulin therapy, International Journal of Nursing Practice, 29 (2023), e13190. https://dx.doi.org/10.1111/ijn.13190

[13] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal, 15 (2017), 104-116. https://dx.doi.org/10.1016/j.csbj.2016.12.005

[14] L.P. Nguyen, D.D. Tung, D.T. Nguyen, H.N. Le, T.Q. Tran, T.V. Binh, D.T.N. Pham, The Utilization of Machine Learning Algorithms for Assisting Physicians in the Diagnosis of Diabetes, Diagnostics, 13 (2023), 2087. https://dx.doi.org/10.3390/diagnostics13122087

[15] S. Harnal, A. Jain, Anshika, A. Singh Rathore, V. Baggan, G. Kaur, R. Bala, Comparative Approach for Early Diabetes Detection with Machine Learning, Proceedings of the 2023 IEEE Student Conference on Electric Machines and Systems, (2023), 10100186. https://dx.doi.org/10.1109/ESCI56872.2023.10100186

[16] R. Krishnamoorthi, S. Joshi, H.Z. Almarzouki, P. Shukla, A. Rizwan, C. Kalpana, B. Tiwari, A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques, Journal of Healthcare Engineering, 2022 (2022), 1684017. https://dx.doi.org/10.1155/2022/1684017

[17] U. Ahmed, G.F. Issa, M.A. Khan, S. Aftab, M.F. Khan, R.A. Said, T.M. Ghazal, M. Ahmad, PREDICTION OF DIABETES EMPOWERED WITH FUSED MACHINE LEARNING, IEEE Access, 10 (2022), 3142097. https://dx.doi.org/10.1109/access.2022.3142097

[18] S.A. Shampa, M.S. Islam, A. Nesa, Machine Learning-based Diabetes Prediction: A Cross-Country Perspective, Proceedings of the National Conference on Innovations in Computing and Management, (2023), 10212596. https://dx.doi.org/10.1109/NCIM59001.2023.10212596

[19] D. Park, E. Jang, Convergence Factors Related to Glycemic Control in Workers with Diabetes Mellitus: using the Korean National Health and Nutrition Examination Survey, 2009-2013, Journal of Korean Clinical Society, 6 (2015), 95-101. https://dx.doi.org/10.15207/JKCS.2015.6.6.095

[20] J. Jeon, D. Kim, S.-H. Ko, H. Kwon, S. Lim, S. Choi, C. Kim, J. An, N. Kim, J. Won, J.H. Kim, B. Cha, K. Song, Current Status of Glycemic Control of Patients with Diabetes in Korea: The Fifth Korea National Health and Nutrition Examination Survey, Diabetes & Metabolism Journal, 38 (2014), 197-203. https://dx.doi.org/10.4093/dmj.2014.38.3.197

[21] J.-H. Park, Y.-E. Lee, Effects of exercise on glycemic control in type 2 diabetes mellitus in Koreans: the fifth Korea National Health and Nutrition Examination Survey (KNHANES V), Journal of Physical Therapy Science, 27 (2015), 3559-3564. https://dx.doi.org/10.1589/jpts.27.3559

[22] A. Lee, H.-J. Kim, The Effect of Vitamin D on the Glycemic Control in Patients with Diabetes: From the Fifth (2010-2012) Korea National Health and Nutrition Examination Survey, Journal of Korean Biological Nursing Science, 22 (2020), 53-60. https://dx.doi.org/10.7586/jkbns.2020.22.1.53

[23] N. Sihlangu, R. Millham, Analysis of machine learning methods to determine the best data analysis method for diabetes prediction, 2023 IEEE International Conference on Technology and Sustainability (ICTAS), (2023), 10082727. https://dx.doi.org/10.1109/ICTAS56421.2023.10082727

[24] N. Aggarwal, C.B. Basha, A. Arya, N. Gupta, A Comparative Analysis of Machine Learning-Based Classifiers for Predicting Diabetes, 2023 IEEE International Conference on Advanced Computing and Communication Technologies (ICACCTech), (2023), 61146:00105. https://dx.doi.org/10.1109/ICACCTech61146.2023.00105

[25] Z. Shaukat, W. Zafar, W. Ahmad, I. Haq, G. Husnain, M. Al-Adhaileh, Y. Ghadi, A. Algarni, Revolutionizing Diabetes Diagnosis: Machine Learning Techniques Unleashed, Healthcare, 11 (2023), 2864. https://dx.doi.org/10.3390/healthcare11212864

[26] S. Sakib, N. Yasmin, I.K. Tasawar, A. Aziz, M.A.B. Siddique, M.M.R. Khan, Performance Analysis of Machine Learning Approaches in Diabetes Prediction, 2021 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), (2021), 53172:9641737. https://dx.doi.org/10.1109/R10-HTC53172.2021.9641737

[27] S.M. Kuriakose, P. Pati, T. Singh, Prediction of Diabetes Using Machine Learning: Analysis of 70,000 Clinical Database Patient Record, 2022 IEEE International Conference on Computing,

Communication and Networking Technologies (ICCCNT), (2022), 54827:9984264. https://dx.doi.org/10.1109/ICCCNT54827.2022.9984264
[28] T. Tominanto, A. Syukur, Affandy, A. Marjuni, Performance Analysis of Machine Learning Algorithms for Diabetes Diagnosis Prediction, 2023 International Seminar on Application for Technology of Information and Communication (iSemantic), (2023), 59612:10295368. https://dx.doi.org/10.1109/iSemantic59612.2023.10295368
[29] S. Shamshirgaran, A. Mamaghanian, A. Aliasgarzadeh, N. Aiminisani, M. Iranparvar-Alamdari, J. Ataie, Age differences in diabetes-related complications and glycemic control, BMC Endocrine Disorders, 17 (2017), 21. https://dx.doi.org/10.1186/s12902-017-0175-5
[30] A.H. Ayoub, M.M. AbuShady, N. Youssef, Implications of age on diabetes control and management, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 11 (2017), S147-S152. https://dx.doi.org/10.1016/j.dsx.2017.09.007
[31] K. Ezhilarasi, R. Rajendran, A.J. Vanisree, Association of BMI with glycemic control in type 2 diabetes patients: A cross-sectional study in Chennai, South India, Biomedical and Pharmacology Journal, 11 (2018), 913-919. https://dx.doi.org/10.13005/bpj/1472
[32] S.K. Sharma, D.K. Vishwakarma, N. Modi, M. Sharma, Obesity and hypertension: An ongoing saga, Journal of the American Society of Hypertension, 13 (2019), 127-134. https://dx.doi.org/10.1016/j.jash.2019.01.007
[33] S. Gupta, S. Kapoor, Lipid abnormalities and efficacy of emerging LDL lowering treatments, Journal of Lipids, 2018, 8657316. https://dx.doi.org/10.1155/2018/8657316
[34] M. Papelbaum, R. Moreira, W. Coutinho, R. Kupfer, L. Zagury, S. Freitas, J.C. Appolinario, Depression, glycemic control and type 2 diabetes, Diabetology & Metabolic Syndrome, 3 (2011), 26. https://dx.doi.org/10.1186/1758-5996-3-26
[35] H. Ascher-Svanum, A. Zagar, D. Jiang, D.L. Schuster, H. Schmitt, E. Dennehy, D. Kendall, J. Raskin, R. Heine, Associations Between Glycemic Control, Depressed Mood, Clinical Depression, and Diabetes Distress Before and After Insulin Initiation: An Exploratory, Post Hoc Analysis, Diabetes Therapy, 6 (2015), 447-459. https://dx.doi.org/10.1007/s13300-015-0118-y
[36] H. Ahmadieh, DIABETES AND DEPRESSION: RELATION TO GLYCEMIC CONTROL, Health & Well-being, 2 (2020), 22-28. https://dx.doi.org/10.54729/2789-8288.1012
[37] B. Crispín-Trebejo, M.C. Robles-Cuadros, A. Bernabé-Ortiz, Association between depression and glycemic control among type 2 diabetes patients in Lima, Peru, Applied Psychology: Health and Well Being, 7 (2015), 269-282. https://dx.doi.org/10.1111/appy.12190
[38] M. Khanal, P. Bhandari, R. Dhungana, Y. Gurung, L. Rawal, G. Pandey, M. Bhandari, S. Devkota, M. Courten, B. Courten, Poor glycemic control, cardiovascular disease risk factors and their clustering among patients with type 2 diabetes mellitus: A cross-sectional study from Nepal, PLoS One, 17 (2022), e0271888. https://dx.doi.org/10.1371/journal.pone.0271888
[39] Y. Li, L. Yu, Z. Liu, S. Jia, Q. Man, Q. Zhu, C. Li, Y. Yang, B. Liu, J. Zhang, Dietary Pattern Associated with the Risk of Poor Glycemic Control in Chinese Diabetic Adults: Results from the China Nutrition and Health Surveillance 2015–2017 Survey, Nutrients, 15 (2022), 56. https://dx.doi.org/10.3390/nu15010056
[40] H. Masuda, D. Ishiyama, M. Yamada, F. Iwashima, Y. Kimura, Y. Otobe, N. Tani, M. Suzuki, H. Nakajima, Relationship Between Long-Term Objectively Measured Physical Activity and Glycemic Control in Type 2 Diabetes Mellitus Patients: A Prospective Cohort Study, Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 14 (2021), 2023-2033. https://dx.doi.org/10.2147/DMSO.S307070