Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Identification of Diabetes Predictive Factors with Combined Factor Analysis of Mixed Data and Machine Learning Methods

Ping Ouyang 1a, Ya Zhang 1ab, Addie Weaver1, and Arezoo Rojhani1
1 Western Michigan University, Kalamazoo, Michigan U.S.A, 49008

a co-first authors
b corresponding author

***Abstract* — Diabetes is a growing public health concern in the United States and worldwide, contributing significantly to morbidity and mortality rates. Previous research has identified risk factors for diabetes as both modifiable and nonmodifiable. Machine learning methods, including Neural Networks, Random Forest, and Support Vector Machines, have shown promise in predicting diabetes identification. This study evaluates the performance of these models to enhance early detection and prevention strategies for diabetes, ultimately aiming to improve public health outcomes and reduce the burden of diabetes-related complications and financial expenses. The analysis utilizes data from the 2022 Behavioral Risk Factor Surveillance System, a nationwide health survey conducted by the Centers for Disease Control and Prevention, incorporating 90 health-related variables and 147,472 participants. Factor Analysis for Mixed Data was used to identify seven diabetes-associated factors: General Health and Socioeconomic Status, Smoking Behavior and Smoking History, Healthcare Access and Preventive Care, Perceived Racial and Ethnic Classification, Mental Health and Social Support, Weight Status, and Substance Use and Well-being. The predictive effectiveness of these seven diabetes-associated factors was examined using the three machine-learning models. All models demonstrated a consistent accuracy of approximately 0.89. Notably, the Neural Networks classifier exhibited well-balanced performance across the key metrics, including accuracy, precision, recall, and F1-score. The top three diabetes-associated factors identified across all models were Mental Health and Social Support, Healthcare Access and Preventive Care, and General Health and Socioeconomic Status, underscoring their critical roles in diabetes prediction.**

***Keywords*—Diabetes; Predictive Factors; Machine Learning; Factor Analysis of Mixed Data; BRFSS**

I. INTRODUCTION

It is estimated that 11.6% of the total American population has diabetes, diagnosed or undiagnosed [1], and rates of diabetes are rising quickly in the U.S. [2]. Diabetes is the eighth leading cause of death in the U.S. [3] but causes poorer outcomes or increases risk of seven of the other nine leading causes of death including heart disease [4], cancer [5], COVID-19 [6], stroke [7], Alzheimer's disease [8], kidney disease [9], and liver diseases [10]. The high rates of diabetes also have a significant impact on the healthcare system, and the costs of those with diabetes are estimated 2.6 times higher than for those without diabetes, and a quarter of US health spending is on people with diabetes [11].

A. Risk Factors for Diabetes

Approximately 90% of diabetes cases are type 2 diabetes (T2DM) [12], and the known risk factors for T2DM can be characterized as modifiable and nonmodifiable. Modifiable risks include unhealthy diet [13], sedentary lifestyle [14], smoking [15], overweight and obesity [16], central adiposity [2], high blood pressure [17], sleep problems [18], low income and less education [19], and so on. Non-modifiable risks include older age [20], gender [1], family history of diabetes [21], non-white ancestry [1], and history of gestational diabetes [2].

It is estimated that up to 90% of T2DM cases could be prevented [12] through diet, exercise, weight loss [22], and medications [23]. Before a person is found to have T2DM, prediabetes is often observed, which indicates a high risk of developing T2DM [23]. It is estimated that 38% of U.S. adults have prediabetes [1], which is reversible with lifestyle changes and medications, but 31% to 41% of those will develop diabetes within 12 years [23]. In addition, it is estimated that a quarter to a third of all cases [1] are undiagnosed diabetes, and the time elapsed until diagnosis can be lengthy, from four to seven years on average [24, 25].

Prompt diagnosis and treatment of diabetes is critical because poorly controlled diabetes increases the risk of cardiovascular disease (CVD) [26], kidney disease and kidney failure [27], retinopathy [28], peripheral neuropathy [29], and mortality [3].

B. Machine Learning Applications in Diabetes Prediction

As technology advances in processing massive amounts of data, more and more researchers have evaluated and predicted the occurrence, development, and prognosis of diseases using machine learning methods. To learn more about the risk factors and consequences of some health problems, studies mine vast amounts of epidemiological data. For example, Ma and colleagues used a backpropagation artificial neural network (ANN) to establish a predictive model and predict childhood asthma [30], and a multivariable logistic regression model was developed to assess the prevalence of stroke in patients with prediabetes and diabetes and to identify predictors of stroke [31].

Chowdhury and colleagues analyzed 2021 BRFSS data with 20 selected variables to identify diabetes risk and investigate how sampling techniques impact the accuracy of models [32]. Sampling techniques to address imbalanced data included the

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Synthetic Minority Oversampling Technique (SMOTE), Edited Nearest Neighbors (ENN), and SMOTE-ENN. Then, logistic regression, gradient boosting, AdaBoost, and Random Forest (RF) machine learning algorithms were applied to each sampling technique. The authors found that sampling strategies improved accuracy in all models, with ENN being the most effective.

A study analyzed 22 features from the 2015 BRFSS dataset to detect risk factors for diabetes [33]. Machine learning models, including K-Nearest Neighbor (KNN), RF, XGBoost, bagging, and AdaBoost, were utilized with a Synthetic Minority Over-sampling Technique integrated with the Edited Nearest Neighbor (SMOTE-ENN) for treating imbalanced data. KNN was found to be the most accurate at 98.4%. The authors attributed the high accuracy to SMOTE-ENN.

The 2015 BRFSS dataset has also been utilized to identify risk factors for T2DM using various machine learning models, including Support Vector Machine (SVM), decision tree, logistic regression, random forest (RF), Gaussian Naïve Bayes, and neural networks (NN) [34]. Twenty-seven variables were selected for analysis, and imbalanced data was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). While all models demonstrated high accuracy, the NN achieved the highest accuracy at 82.4%.

Another study by Saeed [35] analyzed 21 variables from the 2022 BRFSS dataset and nine from the PIMA dataset to identify the most effective model for predicting diabetes risk among Decision Tree Classifier (DTC), AdaBoost, Gradient Boosting Classifier (GBC), and Extra Trees Classifier (ETC). ETC demonstrated the highest accuracy, achieving 0.89 for the PIMA dataset and 0.96 for the BRFSS dataset.

C. Machine Learning Techniques in the Current Study

In this study, three different machine learning techniques, including NN, RF, and SVM, have been tested. Neural networks, a type of machine learning, were first developed in the mid-1900s [36] to replicate human brain processes. They consist of interconnected neurons, or nodes, that act as signaling units. Each node receives one or more inputs, and if the weighted sum exceeds a threshold, it activates subsequent neurons. Nodes are organized into layers. The first layer is called the input layer, and the last layer is the output layer. Nodes between the input and output layers are in hidden layers, making the model a black box [36]. Models with more than three layers are classified as deep learning [37]. Neural networks learn by adjusting node weights to produce the desired outputs, making them effective for pattern recognition, latent factor discovery, and new data analysis.

Random forest classification, introduced by Breiman in 2001 [38], is a machine learning method used to create predictive models. It consists of multiple decision trees trained using the Classification and Regression Tree (CART) algorithm, with randomly selected variables. Key parameters such as node size, the number of trees, and feature sampling are set before training. Results are validated using out-of-bag samples.

Support Vector Machines (SVMs), developed by Cortes and Vapnik in 1995, are designed for classification tasks [39]. SVMs utilize binary classification to identify an optimal hyperplane with a maximum margin classifier. SVM can be applied to non-linear problems with the kernel function [40]. They can handle non-linear problems using kernel functions [40]. Compared to NN and decision trees, SVMs are less prone to overfitting but are also less flexible and more difficult to interpret [37].

D. Growing Need for Improved Diabetes Prediction

Diabetes has long been a critical public health issue, and the COVID-19 pandemic has further exacerbated the situation. A systematic review and meta-analysis of nine studies involving 40 million participants found a statistically significant increase in the relative risk of both type 1 and type 2 diabetes following SARS-CoV-2 infection [41]. This elevated risk was observed across all age groups and genders, with the greatest risk occurring within the first three months after infection [41].

Under these new circumstances, it is crucial to identify the predictive factors for T2DM using the most comprehensive database available—BRFSS—and to raise public awareness. This study has two primary objectives. First, it aims to classify diabetes-related variables obtained from large-scale health data and identify key factors associated with diabetes. Second, it seeks to apply various machine learning classifiers to determine the most important predictors of diabetes, ultimately enhancing the accuracy of diabetes prediction models.

The overarching goal of this study is to provide empirical evidence that federal and state governments can use to improve public health management and policy guidance. The findings will help encourage high-risk individuals to undergo screening for prediabetes or diabetes, raising awareness of existing and potential health risks. Additionally, the study aims to inform preventive measures for those without diabetes, reducing their likelihood of developing the disease.

II. METHODS

A. Data Source

The Behavioral Risk Factor Surveillance System (BRFSS) is the largest health survey system in the world conducted by the United States Centers for Disease Control and Prevention (CDC). It gathers information related to health behaviors, chronic diseases, and disease prevention among U.S. adults through the telephone [42]. The yearly survey was started in 1984, and by 2001, all 50 states and territories participated [42]. Administered at a state and local level under the guidance of the CDC, the BRFSS surveys more than 400,000 people per year [42]. BRFSS data have been utilized to identify emerging health problems, track health aims, and assess public health problems, which has been consistently found to be a reliable and valid dataset [43].

This study utilizes data from the 2022 BRFSS, which includes 445,132 participants and 328 variables, encompassing a broad spectrum of demographic, lifestyle, and health-related factors. This extensive dataset provides a robust foundation for analyzing diabetes-associated factors and developing predictive models.

B. Data Preparation

Data preprocessing was conducted to ensure data quality and consistency. This involved the following steps:

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

1. Feature Selection:

The target variable, "(Ever told) you had diabetes," was analyzed with three categories: "Yes," "No," and "No, pre-diabetes or borderline diabetes." A total of 121 features associated with diabetes diagnosis—including demographic, lifestyle, and health indicators—were selected for analysis.

2. Handling Missing Values:

Features with more than 90% missing values were excluded from the analysis. Missing data in categorical features were addressed using multiple imputations, while median imputation was applied to continuous features. The final dataset comprised 147,472 participants and 90 features, including 17 numerical and 73 categorical features.

C. Factor Analysis for Mixed Data (FAMD)

Given the high dimensionality of the current dataset, Factor Analysis for Mixed Data (FAMD) was used for dimensionality reduction. FAMD extends Principal Component Analysis (PCA) to accommodate both numerical and categorical features simultaneously [44]. It handles these feature types automatically without the need for manual one-hot encoding of categorical features or normalization of numerical features. This technique has been shown to be effective in managing complex educational data by reducing the number of predictors [45, 46]. The optimal number of factors was determined based on the scree plot and the cumulative variance explained. The *prince* library in Python was used to perform FAMD. The factors extracted through FAMD were then used as inputs for the machine learning models.

D. Machine Learning Methods

1. Neural Networks (NNs):

A feedforward neural network (FNN) was implemented using TensorFlow/Keras in Python for multiclass classification. The model architecture consists of two hidden layers with the ReLU activation function. To improve generalization and prevent overfitting, regularization techniques such as Dropout and Batch Normalization were applied. The model was compiled using the Adam optimizer with a learning rate of 0.0005. Additionally, an early stopping mechanism was implemented to enhance training efficiency by preventing unnecessary epochs and retaining the best-performing model.

2. Random Forest (RF):

A Random Forest (RF) classifier was implemented using Scikit-Learn in Python. The model was fine-tuned by adjusting hyperparameters such as the number of trees (`n_estimators`), the maximum depth of the trees (`max_depth`), and by using `GridSearchCV` to perform an exhaustive search over a range of values to find the best combination of hyperparameters. Model performance was evaluated using classification metrics, including accuracy, precision, recall, and F1-score, with all metrics measured using a weighted average approach to account for class imbalances.

3. Support Vector Machine (SVM):

A Support Vector Machine (SVM) classifier was implemented and fine-tuned by adjusting the regularization parameter (`C`), kernel type, and gamma parameter. The final model was evaluated on the test set using various classification metrics, including accuracy, precision, recall, and F1-score.

III. RESULTS

A. FAMD

A combined visualization of the scree plot and cumulative percent of variance explained was created to aid in factor extraction (see Figure 1). According to the scree plot, the first seven factors were selected based on eigenvalues greater than 1. Following the Kaiser criterion, it is recommended to retain factors that cumulatively explain at least 70% of the variance [47]. In this analysis, the first seven factors accounted for a cumulative variance of 80%.

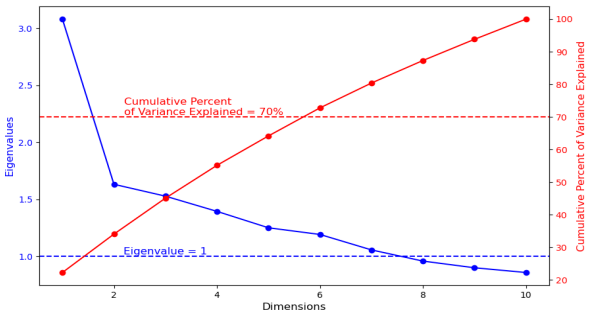

Figure 1. Combine scree plot and cumulative percent of variance explained

The most relevant features for each factor were identified based on a factor loading of 0.3 or higher, and each factor was labeled accordingly based on its loading features.

The first factor**,** General Health and Socioeconomic Status**,** is defined by features such as general health, physical health status, difficulty concentrating or remembering, employment status, physical limitations, education level, and income level.

The second factor, Smoking Behavior and Smoking History, is characterized by features such as age, smoking status, smoking frequency, and cigarette consumption.

The third factor, Healthcare Access and Preventive Care**,** includes features related to the length of time since the last routine checkup or flu shot, access to personal health care providers, and access to health insurance.

The fourth factor, Perceived Racial and Ethnic Classification, is identified by features such as individuals' self-reported ethnicity, racial perception, and treatment experiences.

The fifth factor, Mental Health and Social Support, encompasses features such as a history of depressive disorders, providing regular care for a family member or friend, number of days with poor mental health, overall mental health status, emotional support received, and perceptions of social isolation.

The sixth factor, Weight Status, includes features such as body mass index (BMI) categories and overweight or obesity classification.

The seventh and final factor, Substance Use and Well-being, is defined by features such as alcohol use, marital status, and life satisfaction.

After the factors were determined, it was essential to assess the internal consistency of the features that load onto each factor to ensure reliability in factor identification. To this end,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Cronbach's Alpha was used to evaluate how well a set of features align with a single factor. A Cronbach's Alpha of 0.7 or higher typically indicates that the features are reasonably cohesive, and thus, the identified factors are appropriate for further analysis.

The reliability of the extracted factors is summarized in Table 1. As shown, the Cronbach's Alpha values ranged from 0.8416 to 0.9759, suggesting that the items defining each factor are consistent and reliable.

Table 1. Internal reliability of the extracted factors

Component	Extracted Factors	Items	Cronbach's Alpah
Comp. 1	General Health and Socioeconomic Status	11	0.9456
Comp. 2	Smoking Behavior and Smoking History	4	0.8576
Comp. 3	Healthcare Access and Preventive Care	5	0.8997
Comp. 4	Perceived Racial and Ethnic Classification	5	0.9638
Comp. 5	Mental Health and Social Support	5	0.8416
Comp. 6	Weight Status	3	0.9759
Comp. 7	Substance Use and Well-Being	4	0.8790

B. Machine Learning

The transform function was used to project the original data into a lower-dimensional space defined by the principal factors. The factors extracted through FAMD were then used as input features for training three machine learning classifiers (NN, RF, SVM). Model performance metrics, including accuracy, precision, recall, and F1-score, were calculated to evaluate their performance. These metrics are commonly used in classification tasks, especially when dealing with imbalanced classes.

Accuracy is the proportion of correct predictions (both true positives and true negatives) out of all predictions. Precision refers to the proportion of correctly predicted positive instances out of all instances predicted as positive, indicating how many of the predicted positives are actually true positives. Recall measures the proportion of correctly predicted positive instances out of all actual positive instances, highlighting how well the model identifies all positive cases. Lastly, the F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of the model's performance when both false positives and false negatives are of importance.

The comparisons of the performance metrics of the three ML classifiers are summarized in Table 2.

Table 2. Comparison of machine learning model performance

Metric	Neural Network (NN)	Random Forest (RF)	Support Vector Machine (SVM)
Accuracy	0.8886	0.8819	0.8881
Precision	0.8413	0.8417	0.8561
Recall	0.8646	0.8619	0.8603
F1-score	0.8528	0.8510	0.8598

1. Neural Networks (NNs):

The model was trained for up to 30 epochs with a batch size of 64. The training and validation losses are presented in Figure 2. The model appears to have effectively learned from the training data and generalizes well to unseen data, as evidenced by the stable and relatively low validation loss. Both the training and validation losses decrease and stabilize, indicating that the model is not overfitting.

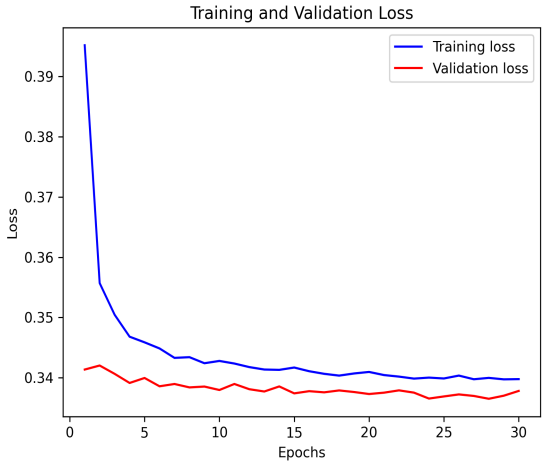

Figure 2. Training and validation losses

The model achieved an accuracy of 0.8886, indicating it correctly classified approximately 89% of the instances. The precision of 0.8413 suggests that when the model predicted a positive class, it was correct 84.13% of the time. The recall of 0.8646 indicates that the model was able to identify 86.46% of the actual positive cases. Finally, the F1-score of 0.8528 balances precision and recall, providing a strong measure of the model's overall performance. These metrics suggest that the NN model performs well in both correctly identifying positive cases and minimizing false positives.

Permutation Importance was calculated to evaluate factor values and their importance in the NN model, as presented in Table 3. The top three predictive factors for diabetes diagnosis are Healthcare Access and Preventive Care, Weight Status, and Mental Health and Social Support.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Table 3. Factor importance of the NN model

Factor	Importance
Healthcare Access and Preventive Care	0.0217
Weight Status	0.0130
Mental Health and Social Support	0.0049
General Health and Socioeconomic Status	0.0036
Substance Use and Well-Being	0.0018
Smoking Behavior and Smoking History	0.0013
Perceived Racial and Ethnic Classification	0.0002

2. Random Forest (RF):

The RF model achieved an accuracy of 0.8819, correctly classifying approximately 88.19% of the instances. The NN model slightly outperforms the RF in overall accuracy. The precision of 0.8417 is comparable to that of the NN model. The RF model's recall of 0.8619 indicates that it correctly identified 86.19% of all actual positive cases, which is slightly lower than the NN model's performance. With an F1-score of 0.8510, the RF model demonstrates a balance between precision and recall. While both the NN and RF models perform similarly, the NN model has a marginal advantage in overall performance.

The feature importance of the RF model is presented in Table 4. The top three predictive factors for diabetes diagnosis were found to be Mental Health and Social Support, Healthcare Access and Preventive Care, and General Health and Socioeconomic Status.

Table 4. Factor importance of the RF model

Factor	Importance
Mental Health and Social Support	0.1620
Healthcare Access and Preventive Care	0.1604
General Health and Socioeconomic Status	0.1542
Smoking Behavior and Smoking History	0.1374
Weight Status	0.1314
Perceived Racial and Ethnic Classification	0.1273
Substance Use and Well-Being	0.1273

3. Support Vector Machine (SVM):

An SVM classifier with a hyperparameter `C` of 0.1 and an `rbf` kernel was trained. The model achieved an accuracy of 0.8881, a precision of 0.8561, a recall of 0.8603, and an F1 score of 0.8598. The SVM model performs similarly to the RF and NN models, with a notable advantage in model precision.

The feature importance for the SVM model is presented in Table 5. The top three predictive factors for diabetes diagnosis were consistent with those identified by the RF model.

Table 5. Factor importance of the SVM model

Factor	Importance
Mental Health and Social Support	0.0222
Healthcare Access and Preventive Care	0.0189
General Health and Socioeconomic Status	0.0157
Smoking Behavior and Smoking History	0.0073
Perceived Racial and Ethnic Classification	0.0031
Weight Status	0.0015
Substance Use and Well-Being	0.0007

IV. DISCUSSION

The United States has a high prevalence of undiagnosed diabetes, accounting for an estimated 25% to 33% of all cases [1]. Timely diagnosis and treatment are crucial, as poorly managed diabetes significantly increases the complications and mortality. Conversely, early diagnosis and appropriate treatment have been shown to improve short-term glycemic control [48], slow disease progression [49], and reduce the risk of complications such as cardiac events [50, 51] and retinopathy [52]. Despite the inconsistencies in diabetes screening guidelines among public health agencies, earlier identification and treatment of T2DM could improve health outcomes and reduce healthcare costs.

The top seven factors accounted for approximately 80% of the variance, with General Health and Socioeconomic Status emerging as the most influential. Diabetic complications are major disease-specific determinants of quality of life, significantly affecting both physical and mental well-being [53]. This explains why individuals with diabetes can often be identified by their general health status, physical limitations, mental health challenges, and vision problems. Diabetes and its complications can lead to illness, job loss, and reduced workforce participation, further impacting socioeconomic stability.

Socioeconomic status (SES) plays a crucial role in both physical and mental health. Studies have shown that individuals from lower SES backgrounds are at a higher risk of developing diabetes and experiencing additional diabetes-related complications compared to those from higher SES backgrounds [54]. A study in Great Britain found that unemployment rates were significantly higher among individuals with diabetes compared to those without (22% vs. 8% in males, 12% vs. 5% in females; $p < 0.001$) [55].

Furthermore, employment status and income level influence economic stability and, consequently, the risk of developing T2DM. A 2022 study using data from the Dutch Lifeline prospective cohort reported that individuals with a monthly income below €1000 had a significantly higher risk of developing T2DM (OR 1.71 [95% CI 1.30–2.26]) [56]. Similarly, a 2023 retrospective cohort study found that children and adolescents from very low- and low-income families had an increased risk of T2DM diagnosis (aHR 1.55; 95% CI 1.41–1.71 and aHR 1.34; 95% CI 1.27–1.41, respectively) compared to those from higher-income families [57].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Our findings align with a 2005 BRFSS dataset analysis, which reported that individuals from households with an annual income below $15,000 experienced more days of poor physical and mental health in the past month compared to those in households earning $50,000 or more [19]. These findings underscore the strong link between socioeconomic status, health disparities, and diabetes outcomes.

The second factor, Smoking Behavior and Smoking History, is characterized by smoking status, smoking frequency, and cigarette consumption. Cigarette smoking is associated with many health problems, including diabetes [58, 59, 60]. A prospective cohort study of 78,212 Koreans demonstrated that the risk of developing diabetes increased in a dose-dependent manner, with longer smoking duration significantly correlating with higher risk ($p < 0.05$) [61]. Tobacco smoke contains over 4,000 chemicals, including heavy metals that can induce insulin resistance [58]. A review study identified several mechanisms through which smoking contributes to diabetes by disrupting carbohydrate metabolism: 1) elevated abdominal adiposity and metabolic syndrome, 2) oxidative stress, endothelial dysfunction, and nicotine-induced insulin resistance via multiple neurotransmitters and catecholamines, 3) increased inhibition of insulin receptor substrate-1 phosphorylation, leading to reduced glucose absorption, and 4) alkaloids in tobacco altering sex hormone balance [58].

The third factor is Healthcare Access and Preventive Care. Given the higher likelihood of developing multiple complications due to diabetes, individuals with diabetes require more medical assistance, such as regular check-ups, visits to healthcare providers, flu vaccinations, and financial support from health insurance. The lived experience of regular check-ups for people with T2DM has shown a universally positive impact on patients' lifestyles by enhancing their knowledge and ability to manage daily life [62].

A 2024 study of 625,279 elderly participants with hypertension found significant reductions in CVD-related mortality and all-cause mortality for those who adhered to regular health check-ups (HR: 0.442, 95% CI: 0.434-0.450; HR: 0.441, 95% CI: 0.435-0.448) [63]. However, even when controlling for similar diabetes and multiple chronic conditions, this population—particularly racial minorities—faced an increased risk of losing access to certain services during the COVID-19 pandemic [64]. Furthermore, Doucette et al. [65] reported that individuals with private insurance were more likely to receive flu vaccinations (OR=1.75, 95% CI: 1.37-2.25), diabetes education (OR=1.36, 95% CI: 1.06-1.74), and standard diabetes services compared to uninsured individuals with diabetes.

The fourth factor, Perceived Racial and Ethnic Classification, reflects an individual's self-reported ethnicity, racial perceptions, and experiences of treatment. The prevalence of diabetes varies among different racial groups. According to the 2011-2016 U.S. National Health and Nutrition Examination Surveys (NHANES), among individuals aged 20 and older, Hispanic Americans have the highest diabetes prevalence (22.1%), followed by non-Hispanic Black adults (20.4%), non-Hispanic Asian Americans (19.1%), with non-Hispanic White Americans having the lowest prevalence (12.1%) [66].

Unfortunately, the COVID-19 pandemic exacerbated the link between racism and health, placing Black, Indigenous, and People of Color (BIPOC) at an even higher risk for mental health challenges and substance abuse [67]. During the pandemic, a survey conducted in April 2020 among American Indian, Alaska Native, Asian, Black, and Latino youth revealed that 72% of respondents had experienced at least one form of ethnic/racial discrimination, with Black individuals reporting significantly higher levels of exposure [67]. A separate U.S. national online survey conducted in April 2020 of young adults aged 18-25 years found that Asian American and Black American groups scored significantly higher on the Coronavirus Racial Bias Scale compared to other groups, such as American Indian/Alaskan Natives and Latinx populations [68].

Racial discrimination can contribute to mental distress, and our results, based on data collected in 2022 during the ongoing COVID-19 pandemic, suggest that racial perceptions and treatment experiences play a significant role in predicting the diabetes population. This indicates that, in addition to managing health issues, many people with diabetes were also experiencing stress related to their racial identity during the pandemic.

The fifth factor, Mental Health and Social Support, includes overall mental health status, history of depressive disorders, and perceptions of social isolation. Our results indicate that mental health and emotional factors play a crucial role in predicting diabetes. Jacobson and colleagues reported that patients with either type of diabetes, who had higher comorbidity severity, scored lower on all domains of the 36-Item Short Form Survey measuring health status and quality of life, including general mental health, social functioning, energy/vitality, and role limitations due to both physical and mental health [69].

People with chronic diseases, including diabetes, face additional challenges to their health and financial well-being, especially due to the COVID-19 pandemic. Gregg and colleagues found that diabetes was the leading cause of severe morbidity in COVID-19 patients, and conversely, COVID-19 had a devastating impact on individuals with diabetes [70]. An online survey of 2,176 U.S. adults conducted between May 29, 2020, and June 30, 2020, assessed depression, anxiety, resilience, perceived stress, and diabetes-related distress. The study found that people with T2DM experienced significantly greater depressive symptoms ($p < 0.05$) and significantly lower levels of resilience ($p < 0.05$) compared to those without diabetes [71].

Social support is critical to health improvement, particularly during the pandemic. A meta-analysis evaluating 11 studies revealed that people with T2DM who had low social support were twice as likely to develop depression compared to those with high social support [72]. Additionally, social isolation and loneliness contribute to the onset of various health problems, including diabetes, cardiovascular disease, immune system issues, cancer, and mental health disorders [73]. The UK Biobank cohort study, with 423,503 adult participants and an average follow-up of 13.5 years, along with the China Health and Retirement Longitudinal Study, which involved 13,800 adult participants with an average follow-up of 5.8 years, both showed that individuals who felt lonely and did not actively engage in leisure or social activities had a higher risk of developing T2DM [74].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

The sixth factor, Weight Status, refers to body mass index (BMI) and classifications of overweight or obesity. According to the World Health Organization [75], in 2022, 43% of adults aged 18 and older were overweight, while 16% were obese worldwide. Overweight and obesity have become the fifth leading cause of global deaths and contribute to 44% of the diabetes burden, 23% of ischemic heart disease cases, and approximately 7–41% of certain cancers [76].

The situation is even more concerning in the United States. According to the 2015-2016 NHANES data, 31.8% of U.S. adults (≥ 20 years old) were overweight, 39.8% were obese, and 7.6% were classified as extremely obese [77]. These conditions have been identified as major risk factors for numerous severe health issues, including all-cause mortality, hypertension, T2DM, dyslipidemia, heart disease, stroke, multiple cancers, and reduced quality of life [78]. A study analyzing data from 310,000 Chinese participants found that when BMI reached 24 kg/m² or higher, the risk of diabetes and metabolic syndrome increased significantly. However, maintaining a BMI below 24 kg/m² was associated with a 45–50% reduction in risk [79].

The final factor, Substance Use and Well-Being, includes elements such as alcohol use, marital status, and life satisfaction. Similar to cigarette smoking, alcohol consumption has also been shown to predict diabetes in this study. While light or moderate alcohol consumption has traditionally been associated with a reduced risk of CVD and mortality, the amplification of these protective effects has been questioned due to systematic selection bias in studies [80] and misleading information spread by the alcohol industry [81]. Similar findings in Chinese [82] and Indian populations [83] did not confirm these protective effects. Another study compared the risks of heart failure, stroke, hypertensive disease, and aortic aneurysm in individuals consuming 100 g of alcohol per week versus 0-25 g per week, showing a positive linear association rather than a J-shaped distribution [84].

Additionally, marital status or cohabitation has been shown to be associated with diabetes. Studies have demonstrated that people living alone (single, divorced, or widowed) have higher rates of prediabetes [85] and diabetes [86, 87, 88]. This may be attributed to elevated stress, unhealthy diets, and a lack of social support for those living alone [89]. As discussed earlier, diabetes-related complications can negatively impact both physical and mental health [53]. Consequently, life satisfaction may decrease due to the reduced quality of life and psychosocial distress [90]. In addition to managing disease treatment and control, support and care from family, community, and society are essential to improving the life satisfaction of individuals with diabetes.

The variables identified in this study include age, gender, exercise, sleep, educational level, prediabetes status, and more. While previous research has examined the relationships between these factors and diabetes, our analysis of the 2022 BRFSS dataset highlights several key predictors. Many diabetes-related variables clustered into broader factors, with Mental Health and Social Support, Healthcare Access and Preventive Care, and General Health and Socioeconomic Status emerging as the top three predictors of diabetes. Additionally, smoking, body weight, alcohol consumption, marital status, and life satisfaction were strongly associated with diabetes risk. A randomized clinical trial among U.S. adults at high risk for T2DM found that lifestyle interventions and medication (metformin) reduced the incidence of diabetes by 58% and 31%, respectively, regardless of gender, age, race, or ethnicity [91]. Our results further confirm that lifestyle plays a crucial role in diabetes prediction.

These findings can help inform targeted interventions, particularly when modifiable factors such as smoking, social interaction, weight management, alcohol consumption, and marital status are involved.

V. STRENGTHS AND LIMITATIONS

To our knowledge, this is the first study to use machine learning to analyze nearly 100 health- and lifestyle-related factors from epidemiological data to predict diabetes. The three tested models—NN, RF, and SVM—all demonstrated a high predictive accuracy of approximately 0.89, indicating that diabetes can be effectively predicted based on individuals' health and lifestyle information. This study specifically utilized BRFSS 2022 data to examine diabetes-associated factors in the context of the COVID-19 pandemic.

Despite the extensive dataset used, several limitations should be acknowledged. First, as the data were collected from the U.S. population, the findings may not be generalizable to populations with significantly different lifestyle habits. Second, BRFSS data were obtained through telephone surveys of noninstitutionalized adults, excluding individuals without landlines or those residing in long-term care facilities. Third, while the BRFSS includes many diabetes risk factors, some crucial information—such as diet and family history of diabetes—were not captured in the survey. Incorporating this information into analytical models could enhance the accuracy of diabetes prediction, particularly for individuals whose condition has not yet been diagnosed. Additionally, all data were self-reported, which may introduce recall and reporting biases; however, previous research has confirmed the reliability of BRFSS data. Fourth, the dataset is imbalanced, with a smaller proportion of participants reporting diabetes, which could affect model performance. Lastly, individuals with higher socioeconomic status, who are more likely to engage in routine health check-ups, may be overrepresented in the prediabetic or diabetic population.

VI. CONCLUSIONS

All three models—Neural Networks, Random Forest, and Support Vector Machines—consistently achieved a diabetes prediction accuracy of 0.89. Notably, the Neural Networks classifier demonstrated well-balanced performance across key metrics, including accuracy, precision, recall, and F1-score. Seven key diabetes-associated factors were identified across a broad range of variables: General Health and Socioeconomic Status, Smoking Behavior and History, Healthcare Access and Preventive Care, Perceived Racial and Ethnic Classification, Mental Health and Social Support, Weight Status, and Substance Use and Well-being. Among these, the strongest predictors across all three classifiers were Mental Health and Social Support, Healthcare Access and Preventive Care, General Health and Socioeconomic Status, and Body Weight.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

REFERENCES

[1] CDC (2024a). National diabetes statistics report. https://www.cdc.gov/diabetes/php/data-research/?CDC_AAref_Val=https://www.cdc.gov/diabetes/dat a/statistics-report/index.html

[2] Chen L., Magliano D. J., & Zimmet P. J., "The worldwide epidemiology of type 2 diabetes mellitus-present and future perspectives." Nature Reviews Endocrinology, 8, 228-236. 2012. https://doi.org/10.1038/nrendo.2011.183

[3] National Center for Health Statistics. Data Brief 492. Mortality in the United States, 2022. CDC. 2022. https://www.cdc.gov/nchs/data/databriefs/db492-tables.pdf#4

[4] Leon B.M. & Maddox T.M. "Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research." World Journal of Diabetes, 6(13), 2046-1258. 2015. https://doi.org/10.4239%2Fwjd.v6.i13.1246

[5] Giovannucci E., Harlan D.M., Archer M.C., Bergenstal R.M., Gapstur S.M., Habel L.A., Pollak M., Regensteiner J.G., & Yee, D., "Diabetes and cancer: A consensus report." Diabetes Care, 33(7), 1647-1685. 2010. https://doi.org/10.2337%2Fdc10-0666

[6] Guo W., Li M., Dong Y., Zhou H., Zhang Z., Tian C., Qin R., Wang H., Shen Y., Du K., Zhao L., Fan H., Luo S., & Hu D., "Diabetes is a risk factor for the progression and prognosis of COVID-19." Diabetes/ Metabolism Research and Reviews, 36(7), e3319. 2020. https://doi.org/10.1002/dmrr.3319

[7] Chen R., Ovbiagele B., & Feng W. "Diabetes and stroke: Epidemiology, pathophysiology, pharmaceuticals and outcomes." American Journal of Medical Science, 351(4), 380-386. 2016. https://doi.org/10.1016%2Fj.amjms.2016.01.011

[8] Barbagallo M. & Dominguez L. J., "Type 2 diabetes mellitus and Alzheimer's disease." World Journal of Diabetes, 5(6), pp. 889-893, 2014. https://doi.org/10.4239%2Fwjd.v5.i6.889

[9] Kumar M., Dev S., Khalid M.U., Siddenthi S.M., Noman M., John C., Akubuiro C., Haider A., Rani R., Kashif M., Varrassi G., Khatri M., Kumar S., & Mohamad T. "The bidirectional link between diabetes and kidney disease: Mechanisms and management." Cureus, 15(9). 2023. https://doi.org/10.7759%2Fcureus.45615

[10] Garcia-Compean D., Jaquez-Quintana J.O., Gonzalez-Gonzalez J.A., & Maldonado-Garza H., "Liver cirrhosis and diabetes: Risk factors, pathophysiology, and clinical implications and management." World Journal of Gastroenterology, 15(3), 280-288. 2009. https://doi.org/10.3748%2Fwjg.15.280

[11] Parker E.D., Lin J., Mahoney T., Ume N., Yang G., Gabbay R.A., ElSayed N.A., & Bannuru R.R. "Economic costs of diabetes in the U.S. in 2022." Diabetes Care, 47(1), 26-43. 2024. https://doi.org/10.2337/dci23-0085

[12] DeFronzo R.A., Ferrannini E., Groop L., Henry R. R., Herman W. H., Holst J. J., Hu F. B., Kahn R., Raz I., Shulman G. I., Simonson D. C., Testa, M. A., & Weiss, R., "Type 2 diabetes mellitus." Nature Reviews Disease Primers. 1, 1-22. 2015. http://dx.doi.org/10.1038/nrdp.2015.19

[13] Salas-Salvado J., Martinez-Gonzalez M.A., Bullo M., & Ros E., "The role of diet in prevention of type 2 diabetes." Nutrition, Metabolism and Cardiovascular Disease, 21(2), B32-B48. 2011. https://doi.org/10.1016/j.numecd.2011.03.009

[14] Hamilton M.T., Hamilton D.G., & Zderic T.W. "Sedentary behavior as a mediator of type 2 diabetes." Medicine and Sport Science. 2014. https://doi.org/10.1159/000357332

[15] Maddatu J., Anderson-Baucum E., & Evans-Molina C. "Smoking and the risk of type 2 diabetes." Translational Research, 184, 101-107. 2017. https://doi.org/10.1016/j.trsl.2017.02.004

[16] Schwarz P.E., Greaves C.J., Lindstrom J., Yates T., & Davies M.J., "Nonpharmacological interventions for the prevention of type 2 diabetes mellitus." Nature Reviews Endocrinology, 8, 363-373. 2012. https://doi.org/10.1038/nrendo.2011.232

[17] Petrie J.R., Guzik T.J., & Touyz R.M., "Diabetes, hypertension, and cardiovascular disease: Clinical insights and vascular mechanisms." Canadian Journal of Cardiology, 34(5), 575-584. 2018. https://doi.org/10.1016%2Fj.cjca.2017.12.005

[18] Darraj A., "The link between sleeping and type 2 diabetes: A systematic review." Cureus. 15(11). 2023. https://doi.org/10.7759%2Fcureus.48228

[19] Campbell H.M., Khan N., Cone C., & Raisch D.W., "Relationship between diet, exercise habits, and health status among patients with diabetes." Research in Social and Administrative Pharmacy, 7 (2), 151-61, 2011. doi: 10.1016/j.sapharm.2010.03.002.

[20] Yan Z., Cai M., Han X., Chen Q., & Lu H., "The interaction between age and risk factors for diabetes and prediabetes: A community-based cross-sectional study." Diabetes, Metabolic Syndrome and Obesity, 16, 85-93. 2023. https://doi.org/10.2147%2FDMSO.S390857

[21] Hariri S., Yoon P. W., Qureshi N., Valdez R., Scheuner M. T., & Khoury M. J. "Family history of type 2 diabetes: A population-based screening tool for prevention?" Genetics in Medicine, 8, 102-108. 2006. https://doi.org/10.1097/01.gim.0000200949.52795.df

[22] Hamman R.F., Wing R.R., Edelstein S.L., Lachin J.M., Bray G.A., Delahanty L., Hoskin M., Kriska A.M., Mayer-Davis E.J., Pi-sunyer X., Regensteiner J., Venditti B., & Wylie-Rosett J., "Effect of weight loss with lifestyle intervention on risk of diabetes." Diabetes Care, 29(9), 2102-2107. 2006. https://doi.org/10.2337%2Fdc06-0560

[23] Echouffo-Tcheugui J.B., Perreault L., Ji L., & Dagogo-Jack S., "Diagnosis and management of prediabetes: A review." JAMA, 329(14), 1206-1216. 2023. https://doi.org/10.1001/jama.2023.4063

[24] Harris M.I., Klein R., Welborn T.A., & Knuiman M.W., "Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis." Diabetes Care, 15(7), 815-819. 1992. https://doi.org/10.2337/diacare.15.7.815

[25] Samuels T.A., Cohen D., Brancati F.L., Coresh J., & Kao W., "Delayed diagnosis of incident type 2 diabetes mellitus in the ARIC study." The American Journal of Managed Care, 12(12), 717-724. 2007.

[26] Raghavan S., Vassy J.L., Ho Y., Song R.J., Gagnon D.R., Cho K., Wilson P.W.F., & Phillips L.S. "Diabetes mellitus-Related all-cause and cardiovascular mortality in a national cohort of adults." Journal of the American Heart Association, 8(4). 2019. https://doi.org/10.1161/JAHA.118.011295

[27] Coca S. G., Ismail-Beigi F., Haq N., Krumholz H. M., & Parikh C. R., "Role of intensive glucose control in development of renal endpoints in type 2 diabetes: Systematic review and meta-analysis. ' Archives of Internal Medicine, 172(10), 761-769. 2012. https://doi.org/10.1001%2Farchinternmed.2011.2230

[28] Wright W.S., Eshaq R.S., Lee M., Kaur G., & Harris N.R., "Retinal physiology and circulation: effect of diabetes." Comprehensive Physiology, 10(3), 933-974. 2020. https://doi.org/10.1002%2Fcphy.c190021

[29] Galiero R., Caturano A., Vetrano E., Beccia D., Brin C., Alfano M., Di Salvo J., Epifani R., Piacevole A., Tagliaferri G., Rocco M., Iadicicco I., Docimo G., Rinaldi L., Sardu C., Salvatore T., Marfella R., & Sasso F.C., "Peripheral Neuropathy in Diabetes Mellitus: Pathogenetic Mechanisms and Diagnostic Options." Int J Mol Sci. 10;24(4):3554. 2023. doi: 10.3390/ijms24043554.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

[30] Ma X.Y., Zhang H., & Zhao Y.H. "Building childhood asthma prediction model with artificial neural network and BRFSS database." Data Analysis and Knowledge Discovery 2(8): 10-15. 2018. https://doi.org/10.11925/infotech.2096-3467.2018.0205

[31] Khan M.M., Roberson S., Reid K., Jordan M., & Odoi A., "Prevalence and predictors of stroke among individuals with prediabetes and diabetes in Florida." BMC Public Health. 6;22(1):243. 2022. https://doi: 10.1186/s12889-022-12666-3.

[32] Chowdhuy M. M., Ayon R. S., & Hossain, M. S., "An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset." Healthcare Analytics, 5. 2023. https://doi.org/10.1016/j.health.2023.100297

[33] Ullah Z., Saleem F., Jamjoom M., Fakieh B., Kateb F., Ali A. M., & Shah B., "Detecting high-risk factors and early diagnosis of diabetes using machine learning methods." Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/2557795

[34] Xie Z., Nikolayeva O., Luo J., & Li D., "Building risk prediction models for type 2 diabetes using machine learning techniques." Preventing Chronic Disease, 16(130). 2019. https://doi.org/10.5888/pcd16.190109

[35] Saeed M.A.H. "Diabetes type 2 classification using machine learning algorithms with up-sampling technique." Journal of Electrical Systems and Information Technology, 10(8). 2023. https://doi.org/10.1186/s43067-023-00074-5

[36] Prieto A., Prieto B., Ortigosa E. M., Ros E., Pelayo F., Ortega J., & Rojas I. "Neural networks: An overview of early research, current frameworks, and new challenges." Neurocomputing, 214, 242-268. 2016. https://doi.org/10.1016/j.neucom.2016.06.014

[37] IBM. What are support vector machines (SVMs)? 2023. https://www.ibm.com/topics/support-vector-machine

[38] Breiman L., "Random Forests." Machine Learning 45, 5–32. 2001. https://doi.org/10.1023/A:1010933404324

[39] Cortes C. & Vapnik V. "Support-vector networks." Mach Learn 20, 273–297. 1995. https://doi.org/10.1007/BF00994018

[40] Valkenborg D., Rosseau A., Geubbelmans M., & Burzykowski T., "Support vector machines. American Journal of Orthodontics & Dentofacial" Orthopedics, 166(4). 754-757. 2023. https://doi.org/10.1016/j.ajodo.2023.08.003

[41] Zhang T., Mei Q., Zhang Z., Walline J.H., Liu Y., Zhu H., & Zhang S., "Risk for newly diagnosed diabetes after COVID-19: a systematic review and meta-analysis." BMC Med. 15;20(1):444. 2022. https://doi: 10.1186/s12916-022-02656-y.

[42] CDC (2024b): Behavioral Risk Factor Surveillance System. https://www.cdc.gov/brfss/

[43] Rolle-Lake L., & Robbins E., "Behavioral Risk Factor Surveillance System." 2023 In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.

[44] Pagès J. "Analyse factorielle de données mixtes". Rev. Statistique Appliquée LII (4) 93–111. 2004.

[45] El Jihaoui M., Abra O.E.K., & Mansouri K., "Factors Affecting Student Academic Performance: A Combined Factor Analysis of Mixed Data and Multiple Linear Regression Analysis." IEEE Access. 2025.

[46] Pereira N., "Factor Analysis of Mixed Data (FAMD) and Multiple Linear Regression in R" Dissertations. 212. 2019. https://arrow.tudublin.ie/scschcomdis/212

[47] Yeomans K.A. & Golder P.A., "The Guttman-Kaiser criterion as a predictor of the number of common factors." The statistician, 221-229. 1982.

[48] Li S.Q., Guthridge S., Lawton P., & Burgess P., "Does delay in planned diabetes care influence outcomes for aboriginal Australians? A study of quality in health care." BMC Health Services Research, 19. 2019. https://doi.org/10.1186/s12913-019-4404-7

[49] Zhang J., Zhang Z., Zhang K., Ge X., Sun R., & Zhai X., "Early detection of type 2 diabetes risk: limitations of current diagnostic criteria." Frontiers in Endocrinology. 2023. https://doi.org/10.3389%2Ffendo.2023.1260623

[50] Herman W.H., Ye W., Griffin S.J., Simmons R.K., Davies M.J., Khunti K., Rutten G.E.H.M., Sandbaek A., Lauritzen T., Borch-Johnsen K., Brown M. B., & Wareham N.J., "Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: A simulation of the results of the Anglo-Danish-Dutch study of intensive treatment in people with screen-detected diabetes in primary care (ADDITION-Europe)." Diabetes Care, 38(8), 1449-1455. 2015. https://doi.org/10.2337/dc14-2459

[51] Paul S.K., Klein K., Thorsted B.L., Wolden M.L., & Khunti K., "Delay in treatment intensification increases the risks of cardiovascular events in patients with type 2 diabetes." Cardiovascular Diabetology, 14(100). 2015. https://doi.org/10.1186/s12933-015-0260-x

[52] Porta M, Boscia F, Lanzetta P, Mannucci E, Menchini U, Simonelli F. "Systematic screening of Retinopathy in Diabetes (REaD project): an Italian implementation campaign." Eur J Ophthalmol. 10;27(2):179-184. 2017. doi: 10.5301/ejo.5000912.

[53] Peyrot M. & Rubin R.R., "Levels and risks of depression and anxiety symptomatology among diabetic adults." Diabetes Care. 20: 585±590. 1997.

[54] Hill-Briggs F., Adler N.E., Berkowitz S.A., Chin M.H., Gary-Webb T.L., Navas-Acien A., Thornton P.L., & Haire-Joshu D., "Social determinants of health and diabetes: a scientific review." Diabetes Care 2020; 44: 258–79.

[55] Robinson N., Yateman N.A., Protopapa L.E., & Bush L., "Unemployment and diabetes." Diabet Med. 6(9):797-803. 1989. doi: 10.1111/j.1464-5491.1989.tb01282.x.

[56] Duan M.J.F., Zhu Y., Dekker L.H. Mierau J. O., Corpeleijn E., Bakker S.J.L., & Navis G., "Effects of Education and Income on Incident Type 2 Diabetes and Cardiovascular Diseases: a Dutch Prospective Study." J GEN INTERN MED 37, 3907–3916 2022. https://doi.org/10.1007/s11606-022-07548-8

[57] Yen F.S., Wei J.C.C., Liu J.S., Hwu C.M., & Hsu C.C., "Parental Income Level and Risk of Developing Type 2 Diabetes in Youth." JAMA Netw Open. 1;6(11):e2345812. 2023. doi: 10.1001/jamanetworkopen.2023.45812.

[58] Rouland A., Thuillier P., Al-Salameh A., Benzerouk F., Bahougne T., Tramunt B., Berlin I., Clair C., Thomas D., Le Faou A.L., Vergès B., & Durlach V., "Smoking and diabetes. Ann Endocrinol" (Paris). 85(6):614-622. 2024. doi: 10.1016/j.ando.2024.08.001.

[59] Song B.J., Akbar M., Abdelmegeed M.A., Byun K., Lee B., Yoon S.K., & Hardwick J.P., "Mitochondrial dysfunction and tissue injury by alcohol, high fat, nonalcoholic substances and pathological conditions through post-translational protein modifications." Redox Biol. 3:109-23. 2014. doi: 10.1016/j.redox.2014.10.004.

[60] Yang Y.S. & Sohn T.S., "Smoking as a Target for Prevention of Diabetes." Diabetes Metab J. 44(3):402-404. 2020. doi: 10.4093/dmj.2020.0126.

[61] Kim J.H., Seo D.C., Kim B.J., Kang J.G., Lee S.J., Lee S.H., Kim B.S., & Kang J.H., "Association between cigarette smoking and new-onset diabetes mellitus in 78,212 Koreans using self-reported questionnaire and urine cotinine." Diabetes Metab J, 44: 426-38 2020.

[62] Edwall L.L., Hellström A.L., Ohrn I., & Danielson E., "The lived experience of the diabetes nurse specialist regular check-ups, as narrated by patients with type 2 diabetes." J Clin Nurs.;17(6):772-81. 2008. doi: 10.1111/j.1365-2702.2007.02015.x.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

[63] Li Z., Wu J., Wen Q., Fu S., Sun X., He T., Zhang W., Lu Y., Yuan H., & Cai J., "Association of regular health check-ups with a reduction in mortality in 625,279 elderly participants with hypertension: A population-based cohort study." Public Health. 237:458-465. 2024. doi: 10.1016/j.puhe.2024.10.024.

[64] Clements J.M., "Access to care by Medicare beneficiaries in the U.S. with diabetes and multiple chronic conditions during the COVID-19 pandemic. J Diabetes Complications." Dec; 36(12): 108355. 2022. doi: 10.1016/j.jdiacomp.2022.108355.

[65] Doucette E.D., Salas J., Wang J., & Scherrer J.F., "Insurance coverage and diabetes quality indicators among patients with diabetes in the US general population." Prim Care Diabetes. 11(6):515-521. 2017. doi: 10.1016/j.pcd.2017.05.007.

[66] Cheng Y.J., Kanaya A.M., Araneta M.R.G., Saydah S.H., Kahn H.S., Gregg E.W., Fujimoto W.Y., & Imperatore G., "Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016." JAMA. 24;322(24):2389-2398. 2019. doi: 10.1001/jama.2019.19365.

[67] Tao X., Yip T., & Fisher C.B., "Psychological Well-Being and Substance Use During the COVID-19 Pandemic: Ethnic/Racial Identity, Discrimination, and Vigilance." J Racial Ethn Health Disparities. 11(1):62-71. 2024. doi: 10.1007/s40615-022-01497-y.

[68] Fisher C.B., Tao X., & Yip T., "The effects of COVID-19 victimization distress and racial bias on mental health among AIAN, Asian, Black, and Latinx young adults." Cultur Divers Ethnic Minor Psychol. 29(2):119-131. 2023 doi: 10.1037/cdp0000539.

[69] Jacobson A.M., de Groot M., & Samson J.A., "The evaluation of two measures of quality of life in patients with type I and type II diabetes." Diabetes Care, 17: 267-274. 1994.

[70] Gregg E.W., Sophiea M.K., & Weldegiorgis M., "Diabetes and COVID-19: Population Impact 18 Months Into the Pandemic." Diabetes Care. Sep;44(9):1916-1923. 2021. doi: 10.2337/dci21-0001.

[71] Myers B.A., Klingensmith R., & de Groot M., "Emotional Correlates of the COVID-19 Pandemic in Individuals With and Without Diabetes." Diabetes Care. 1;45(1):42-58. 2022. doi: 10.2337/dc21-0769.

[72] Azmiardi A., Murti B., Febrinasari R.P., & Tamtomo D.G., "Low Social Support and Risk for Depression in People With Type 2 Diabetes Mellitus: A Systematic Review and Meta-analysis." J Prev Med Public Health. Jan;55(1):37-48. 2022. doi: 10.3961/jpmph.21.490.

[73] Pai N. & Vella S.L., "The physical and mental health consequences of social isolation and loneliness in the context of COVID-19." Current Opinion in Psychiatry 35(5):p 305-310, 2022. doi: 10.1097/YCO.0000000000000806

[74] Song Y., Zhu C., Shi B., Song C., Cui K., Chang Z., Gao G., Jia L., Fu R., Dong Q., Feng L., Zhu C., Yin D., Manson J.E., & Dou K., "Social isolation, loneliness, and incident type 2 diabetes mellitus: results from two large prospective cohorts in Europe and East Asia and Mendelian randomization." EClinicalMedicine. 64:102236. 2023. doi: 10.1016/j.eclinm.2023.102236.

[75] World Health Organization: Obesity and overweight https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

[76] World Health Organization: Fact Sheet No.311 (May 2012). www.who.int/mediacentre/factsheets/fs311/en/

[77] Fryar C.D., Carroll M.D., & Ogden C.L., "Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2015–2016". National Center for Health Statistics. Health E-Stats. 2018.

[78] CDC (2024c): How Overweight and Obesity Impacts Your Health. https://www.cdc.gov/healthy-weight-growth/food-activity/overweight-obesity-impacts-health.html

[79] Jia W.P. "Diabetes mellitus and obesity." Zhonghua Yi Xue Za Zhi. Nov 2;84(21):1761-2. 2004.

[80] Naimi T.S., Stockwell T., Zhao J., Xuan Z., Dangardt F., Saitz R., Liang W., & Chikritzhs T. "Selection biases in observational studies affect associations between 'moderate' alcohol consumption and mortality." Addiction. 112(2):207-214. 2017. doi: 10.1111/add.13451.

[81] Arora M., ElSayed A., Beger B., Naidoo P., Shilton T., Jain N., Armstrong-Walenczak, J K., Mwangi Y., Wang, J.L., Eiselé, F.J., & Pinto Champagne B.M., "The Impact of Alcohol Consumption on Cardiovascular Health: Myths and Measures." Glob Heart. Jul 22;17(1):45. 2022. doi: 10.5334/gh.1132.

[82] Schooling C.M., Sun W., Ho S.Y., Chan W.M., Tham M.K., Ho K.S., Leung G.M., & Lam T.H., "Moderate alcohol use and mortality from ischaemic heart disease: a prospective study in older Chinese people." PLoS One. 4;3(6):e2370. 2008. doi: 10.1371/journal.pone.0002370.

[83] Roy A., Prabhakaran D., Jeemon P., Thankappan K.R., Mohan V., Ramakrishnan L., Joshi P., Ahmed F., Mohan B.V.M., Saran R.K., Sinha N., & Reddy K.S., "Sentinel Surveillance in Industrial Populations Study Group. Impact of alcohol on coronary heart disease in Indian men." Atherosclerosis. 210: 531–535.2010.doi: https://doi.org/10.1016/j.atherosclerosis.2010.02.033)

[84] Wood A.M., Kaptoge S., Butterworth A.S., et al. "Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies." The Lancet. 391. 2018. doi: 10.1016/S0140-6736(18)31168-1.

[85] Ford K.J., & Robitaille A., "How sweet is your love? Disentangling the role of marital status and quality on average glycemic levels among adults of 50 years and older in the English Longitudinal Study of Ageing." BMJ Open Diab Res Care 11:e003080. 2023 https://doi.org/10.1136/bmjdrc-2022-003080

[86] Cornelis M.C., Chiuve S.E., Glymour M.M., Chang S.C., Tchetgen E.J.T., Liang L., Koenen K.C., Rimm E.B., Ichiro Kawachi I., & Kubzansky L. D., "Bachelors, divorces, and widowers: does marriage protect men from type 2 diabetes?" PLoS ONE 9(9):e106720. 2014 https://doi.org/10.1371/journal.pone.0106720

[87] Friedrich N., Schneider H.J., John U., Dörr M., Baumeister S.E., Homuth G., Völker U., Wallaschofski H., "Correlates of adverse outcomes in abdominally obese individuals: findings from the five-year followup of the population-based Study of Health in Pomerania." J Obes 2013:762012. 2013. https://doi.org/10.1155/2013/762012

[88] Huang J., Xiao L., Zhao H., Liu F., & Du L., "Living alone increases the risk of developing type 2 diabetes mellitus: a systematic review and meta-analysis based on longitudinal studies." Prim Care Diabetes 18:1–6. 2024. https://doi.org/10.1016/j.pcd.2023.11.011

[89] Kowall B. & Rathmann W., "Partnership and marriage and risk of type 2 diabetes: a narrative review. Diabetologia." 7. doi: 10.1007/s00125-025-06360-3. 2025.

[90] Lee L.Y., Hsieh C.J., & Lin Y.T., "Life satisfaction and emotional distress in people living with type 2 diabetes mellitus: The mediating effect of cognitive function." J Clin Nurs. 30(17-18):2673-2682. 2021. doi: 10.1111/jocn.15740.

[91] Knowler W.C., Barrett-Connor E., Fowler S.E., Hamman R.F., Lachin J.M., Walker E.A., & Nathan D.M., "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin." N. Engl. J. Med. 346, 393–403. 2002.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Ping Ouyang
co-first authors,
Western Michigan University,
Kalamazoo, Michigan U.S.A, 49008

Ya Zhang
co-first authors
corresponding author
Western Michigan University,
Kalamazoo, Michigan U.S.A, 49008

(This work is licensed under a Creative Commons Attribution 4.0 International License.)