Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/44/20240844

Exploring patterns and insights in a comprehensive diabetes dataset

Eric Zhuang[1,5,6,†], Yiyu Chen[2,7,†], Xima Ran[3,8], Jinfei Yi[4,9]

[1]Qingdao Academy, Qingdao, 266114, China
[2]Beijing Aidi School, Beijing, 100018, China
[3]University of Pavia, Pavia, 27100, Italy
[4]Northeastern University, Shenyang, 110819, China
[5]Corresponding author

[6]eric2007118@gmail.com
[7]2262917289@qq.com
[8]tigerhouston@hotmail.com
[9]82442300@qq.com
[†]These authors are co-first authors

Abstract. Diabetes, a pressing global health concern, imposes significant economic and healthcare burdens on millions worldwide. Understanding the multifaceted factors contributing to diabetes is pivotal for effective prevention strategies. In this study, we leverage a comprehensive dataset (952 instances, 17 predictor variables) and employ a multifaceted statistical approach to explore the intricate interplay among stress levels, blood pressure (BP), body mass index (BMI), age, sleep quality, and physical activity in relation to diabetes, with a focus on classification and predictive implications. Our research begins by establishing fundamental relationships between discrete variables using crosstabs and chi-square tests. We uncover close associations between stress levels and BP, heightened diabetes risk with increased BMI values, and the influence of age on sleep quality. Subsequent analysis, based on descriptive data, reveals a robust correlation between physical activity and stress levels, with the paradoxical observation that excessive exercise may increase stress levels. Factor analysis further elucidates the pivotal roles of sound sleep and regular exercise in diabetes prevention, supported by asymptotic significance levels below 0.05. To culminate our study, we construct a logistic regression model with an impressive 89.3% accuracy rate for predicting diabetes risk. Notably, age, family history of diabetes, and regular medication usage emerge as the most influential factors, with regular medication demonstrating significant potential for reducing diabetes risk. Our research underscores the intricate web of factors shaping individual health and offers valuable insights for a comprehensive understanding of health and well-being in the context of diabetes prevention. Moreover, it highlights the importance of considering multiple factors in health-related research. Future research could delve into the long-term effects of interventions targeting the identified risk factors, explore the impact of socio-economic factors on diabetes risk, and investigate the potential role of emerging technologies in personalized diabetes prevention strategies.

Keywords: Diabetes, Stress, Body Mass Index, Physically Active, Blood Pressure, Sound Sleep.

DOI: 10.54254/2753-8818/44/20240844

1. Introduction

The investigation of diabetes assumes paramount significance due to its classification as a chronic medical condition characterized by elevated blood glucose levels, stemming from the body's inability to effectively regulate them. Globally, approximately 422 million individuals are afflicted with diabetes, with the majority residing in low- and middle-income nations and 1.5 million fatalities can be directly attributed to diabetes annually. Both the incidence and prevalence of diabetes have demonstrated steady increments over the preceding few decades. Diabetes manifests in two primary forms: Type 1 Diabetes, an autoimmune ailment primarily diagnosed in childhood, necessitating lifelong insulin therapy, and Type 2 Diabetes, which predominantly manifests in adulthood and is frequently linked to factors such as obesity and genetics. Initially, Type 2 Diabetes can be managed through lifestyle modifications and oral medications. Gestational Diabetes arises during pregnancy and typically resolves postpartum but elevates the risk of Type 2 Diabetes later in life. Less prevalent forms include Monogenic Diabetes, caused by genetic mutations, and Secondary Diabetes, often stemming from underlying medical conditions or medication usage. Common symptoms encompass increased thirst, frequent urination, fatigue, and blurred vision. Effective management through lifestyle adjustments, pharmaceutical interventions, regular monitoring, and healthcare guidance is imperative to avert complications such as heart disease, renal dysfunction, and nerve damage associated with diabetes. Hence, the study of diabetes assumes pivotal significance in advancing our comprehension of its ethology, prophylaxis, and therapeutic strategies, ultimately enhancing the quality of life for millions of affected individuals worldwide.

We have employed a dataset provided by Neha Prerna Tigga and Dr. Shruti Garg from the Department of Computer Science and Engineering at BIT Mesra, Ranchi-835215. This dataset comprises a total of 952 instances, encompassing 17 independent predictor variables and one binary target or dependent variable, which is Diabetes. Our research endeavours to establish the importance of comprehending the interrelationships among stress, blood pressure, BMI, age, gender, physical activity, pregnancy, urination frequency, consumption of junk food, family history of diabetes, and sleep patterns. We underscore the relevance of these factors to overall health and well-being and subsequently delineate our research objectives.

The subsequent sections of this paper are structured as follows:

In Section 1, we employ crosstabs and chi-square tests for discrete variables. Firstly, we delve into the association between stress levels and blood pressure, statistically validate the observed disparities, elucidating a conspicuous correlation, we also refer to previous research to elucidate how stress can influence blood pressure through diverse pathways. Secondly, we scrutinize the relationship between BMI and the propensity for diabetes. Our findings reveal that as BMI values increase, the percentage of individuals with diabetes also rises. We employ statistical tests to affirm the significance of these discoveries and reference existing research to expound on the fact that elevated BMI values are associated with an augmented risk of diabetes due to factors such as insulin production and unhealthy dietary habits. Thirdly, we delve into the impact of age on the quality of sleep, presenting average sound sleep scores across different age groups and conducting significance testing to discern disparities. Our findings underscore age-related variations in sleep patterns, emphasizing the significance of age in the realm of sleep research.

In Section 2, we analysed descriptive data, reveal a robust correlation between the extent of physical activity and stress levels. We observe that higher physical activity is associated with diminished stress, with the intriguing observation that excessive exercise can paradoxically lead to increased stress. We recommend further research in this domain, highlighting a potential research gap about the impact of excessive physical activity on emotional well-being.

In Section 3, we conduct factor analysis to investigate the factors directly contributing to the development of diabetes. We ascertain the corresponding eigenvalues for each component, thus highlighting the significance of sound sleep and regular exercise in diabetes prevention. Moreover, an asymptotic significance level below 0.05 underscores the substantial importance of these factors.

DOI: 10.54254/2753-8818/44/20240844

In Section 4, we centre on the creation of a predictive model for diabetes using logistic regression, encompassing the overall accuracy, coefficients of various variables, and errors, all of which are meticulously presented in tabular format. Consequently, our model demonstrates an overall accuracy rate of 89.3%, with age, family history of diabetes, and regular medication usage emerging as the top three most influential factors within the model, where the regular medication could significantly reduce the possibility of getting diabetes.

2. Crosstabs and Chi-square Tests for Discrete Variables

2.1. Stress and BP level

Table 1. Crosstabs and Chi-Square Tests for stress and BP level.

Stress \ BP Level		0	1	2	Total
0	Count	0	116	20	136
	% of Stress	0.0%	85.3%	14.7%	100.0%
	% of BP Level	0.0%	16.4%	9.3%	14.3%
	% of Total	0.0%	12.2%	2.1%	14.3%
1	Count	28	440	96	564
	% of Stress	5.0%	78.0%	17.0%	100.0%
	% of BP Level	100.0%	62.1%	44.4%	59.2%
	% of Total	2.9%	46.2%	10.1%	59.2%
2	Count	0	108	56	164
	% of Stress	0.0%	65.9%	34.1%	100.0%
	% of BP Level	0.0%	15.3%	25.9%	17.2%
	% of Total	0.0%	11.3%	5.9%	17.2%
3	Count	0	44	44	88
	% of Stress	0.0%	50.0%	50.0%	100.0%
	% of BP Level	0.0%	6.2%	20.4%	9.2%
	% of Total	0.0%	4.6%	4.6%	9.2%
Total	Count	28	708	216	952
	% of Total	2.9%	74.4%	22.7%	100.0%

Table 2. Chi-Square Tests for stress and BP level.

Chi-Square Tests	Value	df	Asymptotic Significance
Pearson Chi-Square	81.302	6	<.001

The cross-tabulation and Chi-square analysis regarding the relationship between individuals' stress levels and blood pressure (BP) are presented in Table 1. Upon a thorough examination of the data, it becomes evident that a discernible correlation exists between individuals' BP levels and their stress levels.

Regarding individuals with stress level 0, which signifies the absence of stress, a noteworthy 85.3% of them exhibited normal BP levels. In contrast, for those with stress level 3, indicative of consistent high-stress levels, the proportion of the sample displaying normal BP levels declined to 50%, while the remaining half of the sample exhibited elevated BP levels. This statistical observation is further substantiated by the Chi-square Test, which indicates an asymptotic significance level of less than 0.05, signifying that the observed disparities hold statistical significance.

An earlier investigation [1] has provided comprehensive insights into this phenomenon. The study delineates four distinct pathways through which stress can trigger an elevation in blood pressure: via the

DOI: 10.54254/2753-8818/44/20240844

sympathetic nervous system, the hypothalamic-pituitary-adrenocortical axis, unhealthy lifestyle choices, and mental distress.

2.2. *Diabetic and BMI*

Table 3. Crosstabs Test for diabetic and BMI.

Diabetic \ BMI		15	17	18	19	20	21	22	23	24	25	26	27	28	29
0	Count	8	16	24	29	43	59	46	57	81	32	39	42	46	25
	% of Diabetic	1.2%	2.3%	3.5%	4.2%	6.3%	8.6%	6.7%	8.3%	11.9%	4.7%	5.7%	6.1%	6.7%	3.7%
	% of BMI	100.0%	100.0%	75.0%	80.6%	89.6%	67.0%	79.3%	75.0%	73.0%	94.1%	59.1%	66.7%	64.8%	89.3%
	% of Total	0.8%	1.7%	2.5%	3.1%	4.5%	6.2%	4.9%	6.0%	8.5%	3.4%	4.1%	4.4%	4.9%	2.6%
1	Count	0	0	8	7	5	29	12	19	30	2	27	21	25	3
	% of Diabetic	0.0%	0.0%	3.0%	2.6%	1.9%	10.9%	4.5%	7.2%	11.3%	0.8%	10.2%	7.9%	9.4%	1.1%
	% of BMI	0.0%	0.0%	25.0%	19.4%	10.4%	33.0%	20.7%	25.0%	27.0%	5.9%	40.9%	33.3%	35.2%	10.7%
	% of Total	0.0%	0.0%	0.8%	0.7%	0.5%	3.1%	1.3%	2.0%	3.2%	0.2%	2.8%	2.2%	2.6%	0.3%
Total	Count	0	16	32	36	48	88	58	76	111	34	66	63	71	28
	% of Total	0.0%	1.7%	3.4%	3.8%	5.1%	9.3%	6.1%	8.0%	11.7%	3.6%	7.0%	6.6%	7.5%	3.0%

Diabetic \ BMI		30	31	32	33	34	35	36	38	39	40	42	45	Total
0	Count	20	8	11	46	8	0	16	20	4	0	3	0	683
	% of Diabetic	2.9%	1.2%	1.6%	6.7%	1.2%	0.0%	2.3%	29.0%	0.6%	0.0%	0.4%	0.0%	100.0%
	% of BMI	60.6%	50.0%	68.8%	71.9%	66.7%	0.0%	80.0%	71.4%	100.0%	0.0%	100.0%	0.0%	72.0%
	% of Total	2.1%	0.8%	1.2%	4.9%	0.8%	0.0%	1.7%	2.1%	0.4%	0.0%	0.3%	0.0%	72.0%
1	Count	13	8	5	18	4	12	4	8	0	4	0	1	265
	% of Diabetic	4.9%	3.0%	1.9%	6.8%	1.5%	4.5%	1.5%	3.0%	0.0%	1.5%	0.0%	0.4%	100.0%
	% of BMI	39.4%	50.0%	31.3%	28.1%	33.3%	100.0%	20.0%	28.6%	0.0%	100.0%	0.0%	100.0%	28.0%
	% of Total	1.4%	0.8%	0.5%	1.9%	0.4%	1.3%	0.4%	0.8%	0.0%	0.4%	0.0%	0.1%	28.0%
Total	Count	33	16	16	64	12	12	20	28	4	4	3	1	984
	% of Total	3.5%	1.7%	1.7%	6.8%	1.3%	1.3%	2.1%	3.0%	0.4%	0.4%	0.3%	0.1%	100.0%

Table 4. Chi-Square Test for diabetic and BMI

Chi-Square Tests	Value	df	Asymptotic significance (2-sided)
Pearson Chi-Square	95.112	25	<.001

The presented tables 3 and 4 showcase the results of the crosstabulation and Chi-square analysis, elucidating the association between diabetes and the Body Mass Index (BMI) within the sample. It is evident from the graphical representations that a correlation exists between these two variables at the individual level.

DOI: 10.54254/2753-8818/44/20240844

According to Table 3, The BMI values span a range from 15 to 45 among the sampled individuals. Notably, the proportion of individuals with diabetes displays an ascending trend as BMI values increase within the sample. To illustrate, when examining individuals with a BMI of 20 among the total 48 individuals, only 10% exhibit diabetes. Conversely, as the BMI escalates to higher levels, such as 38, the prevalence of diabetes among the sample also increases to 28.6%. Furthermore, the application of the Chi-square test as illustrated in Table 4 aids in substantiating the statistical significance of these observed disparities.

This phenomenon finds its rationale in prior research [2]. Diabetes mellitus is a medical condition arising from the inadequate production of insulin within the body—a hormone responsible for regulating blood sugar levels. The insufficiency of insulin stands as the principal causative factor behind diabetes mellitus. High BMI values, often indicative of a substantial body fat percentage, illuminate this correlation within the dataset. Individuals with obesity, characterized by an elevated BMI, frequently maintain an unhealthy dietary regimen, marked by excessive sugar consumption. A fraction of this sugar undergoes conversion into fat, while the remainder lingers in the bloodstream. Consequently, individuals with a high BMI are more susceptible to the development of diabetes.

2.3. Physically active and stress

Table 5. Crosstabs Test for Physically Active and Stress

Physically Active \ Stress		0	1	2	3	Total
0	Count	20	72	20	20	132
	% within PhysicallyActive	15.2%	54.5%	15.2%	15.2%	100.0%
	% within Stress	14.7%	12.8%	12.2%	22.7%	13.9%
	% of Total	2.1%	7.6%	7.6%	2.1%	13.9%
1	Count	48	188	76	24	336
	% within PhysicallyActive	14.3%	56.0%	22.6%	7.1%	100.0%
	% within Stress	35.3%	33.3%	46.3%	27.3%	35.3%
	% of Total	5.0%	19.7%	8.0%	2.5%	35.3%
2	Count	32	188	36	16	272
	% within PhysicallyActive	11.8%	69.1%	13.2%	5.9%	100.0%
	% within Stress	23.5%	33.3%	22.0%	18.2%	28.6%
	% of Total	3.4%	19.7%	3.8%	1.7%	28.6%
3	Count	36	116	32	28	212
	% within PhysicallyActive	17.0%	54.7%	15.1%	13.2%	100.0%
	% within Stress	26.5%	20.6%	19.5%	31.8%	22.3%
	% of Total	3.8%	12.2%	3.4%	2.9%	22.3%
Total	Count	136	564	164	88	952
	% within PhysicallyActive	14.3%	59.2%	17.2%	9.2%	100.0%

Table 6. Chi-Square Test for Physically Active and Stress

Chi-Square Tests	Value	df	Asymptotic Significance
Pearson Chi-Square	31.226	9	<.001

The tables presented above depict cross-tabulations and the results of a Chi-square test examining the relationship between the degree of physical activity and stress. An examination of stress level distribution across various groups of participants with differing levels of physical activity reveals a vigorous correlation between these two variables.

As shown in Table 5, Among individuals with low physical activity, approximately 30% exhibit stress levels of 2 and 3, indicative of relatively high stress. However, as the degree of physical activity increases, the percentage of individuals experiencing high-level stress significantly decreases. This trend is most pronounced at the second level of physical activity, where only 20% of participants fall within the stress level range of 2 to 3. A vertical analysis reinforces these findings, with low-stress levels (0-1) predominantly associated with high levels of physical activity. Contrarily, high stress levels (2-3) are most frequently observed among those with low physical activity levels. Furthermore, an intriguing observation emerges excessive exercise may also lead to heightened stress levels. Among individuals with the highest level of physical activity, stress level 3 is the most prevalent, signifying the highest degree of stress.

In light of these observed phenomena, the previous research [3] holds promise in providing valuable insights. It is worth noting that a substantial body of research consistently yields congruent findings regarding the association between stress and physical activity. These studies consistently highlight the beneficial role of moderate exercise in mitigating depression and managing stress. Notably, however, none of these investigations have delved into the potential consequences of excessive physical activity on an individual's emotional well-being—an area ripe for further exploration.

The findings of the study [4], coupled with their previous research [5-8], affirm that exposure to stress (referred to as IS) results in a lasting decrease over a four-week period in spontaneous wheel-running activity and a decrease in anti-KLH Ig levels. Importantly, engaging in voluntary freewheel running prior to experiencing stress (IS) effectively prevents both the stress-induced decline in wheel-running activity and the suppression of the anti-KLH antibody response. These findings suggest that engaging in physical activity prior to stress exposure mitigates the behavioral downturn and immune suppression caused by stress. However, the specific mechanisms responsible for this effect remain unidentified.

The stress-buffering effect of freewheel running can be explained through various potential immunological mechanisms. To understand this, we need to look at the immune response triggered by the KLH antibody. This response involves the interaction of different immune cells, such as antigen-presenting cells, T helper cells (specifically Th1 and Th2), and B cells.

A study conducted by Fleshner and colleagues [9] revealed that stress, in this context referred to as "IS" (immobilization stress), has an adverse impact on the KLH-stimulated immune response. Specifically, it was found that stress reduced the number of Th1 splenocytes and interferon-gamma (IFN-γ) production, which is a cytokine associated with Th1 cells. IFN-γ plays a crucial role in the transition of antibodies from the IgM class to the IgG2a class.

The selective suppression of anti-KLH IgG2a observed in response to stress suggests that Th1 cells are particularly sensitive to stressors. In simpler terms, stress during KLH immunization leads to a decrease in Th1 cells, resulting in reduced IFN-γ production and subsequently lower levels of anti-KLH IgG2a antibodies.

However, there is an interesting observation related to physical activity. It appears that engaging in physical activity prevents the stress-induced reduction in anti-KLH IgG2a antibodies without significantly affecting anti-KLH IgG1. This leads us to speculate that physical activity may act as a protective factor by preventing the stress-induced decrease in anti-KLH Th1 cells, ultimately contributing to the maintenance of a robust immune response against KLH.

DOI: 10.54254/2753-8818/44/20240844

3. Descriptives Data

3.1. SoundSleep and Age

Table 7. Case Processing Summary for Descriptive Data SoundSleep and Age

Case Processing Summary							
Age		Cases					
		Effective		Loss		Total	
		N	Percentage	N	Percentage	N	Percentage
SoundSleep	40-49	164	100.0%	0	0.0%	164	100.0%
	50-59	156	100.0%	0	0.0%	156	100.0%
	> 60	144	100.0%	0	0.0%	144	100.0%
	< 40	488	100.0%	0	0.0%	488	100.0%

Descriptives				
Age			Statistic	Std.Error
SoundSleep	< 40	Mean	5.54	0.087
	40-49	Mean	5.73	0.160
	50-59	Mean	5.10	0.113
	> 60	Mean	5.50	0.152

The data depicted in Table 7 indicates that individuals in the age group of 40-49 exhibit an average duration of restful sleep amounting to 5.73 hours, while individuals aged 50-59 experience an average restful sleep period of 5.10 hours. On the contrary, those in the age bracket of 60 and above demonstrate an average duration of restful sleep at 5.50 hours, whereas individuals below the age of 40 manifest an average restful sleep duration of 5.54 hours, as illustrated in Table 7.

Table 8. Test of Normality for the distribution of SoundSleep and Age

Test of Normality							
Age		Kolmogorov-Smirnov(V)[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
SoundSleep	< 40	0.124	488	0.000	0.962	488	0.000
	40-49	0.155	164	0.000	0.947	164	0.000
	50-59	0.147	156	0.000	0.949	156	0.000
	> 60	0.136	144	0.000	0.955	144	0.000

Considering that all categories of payment users exhibit significance levels below 0.05, signifying a departure from a Gaussian distribution, it is advisable to pursue non-parametric testing, as depicted in Table 8.

Table 9. The Hypothesis Test for SoundSleep and Age

Hypothesis Test Summary			
Null Hypothesis	Test	Sig.	Decision
The distribution of SoundSleep is the same across categories of Age.	Independent-Samples Kruskal-Wallis Test	0.023	Reject the null hypothesis.

Table 10. The Independent-Samples Kruskal-Wallis Test Summary for SoundSleep and Age

Independent-Samples Kruskal-Wallis Test Summary	
Total N	952
Test Statistic	9.513[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	0.023

The null hypothesis rejection, as determined through the decision-making process, signifies a discernible variance in the distribution of sound sleep among various age categories. Consequently, the alternative hypothesis posits that the distribution of sound sleep exhibits variability across distinct age groups, as visually represented in Table 9.

Moreover, the attainment of a two-sided asymptotic significance level below 0.05 substantiates the statistical significance of the observed distinctions, as elucidated in Table 10.

The phenomenon showcased above has been consistently discerned in a plethora of research studies [10], revealing a notable and inverse correlation between these two variables. In more precise terms, as individuals progress in age, there is a tendency for their sleep quality to deteriorate.

This empirical observation aligns with antecedent scholarly investigations, as exemplified by the research conducted by Uhlig et al. [11], who documented a heightened prevalence of insomnia among individuals aged over 50. Furthermore, a multitude of additional studies has lent credence to the nexus between age and sleep quality [12-13], thereby accentuating the robustness of this association.

The fundamental mechanism underpinning this phenomenon seems to be intricately linked to the natural physiological changes that manifest in adults as they advance in years. These alterations entail an escalating susceptibility of the body's regulatory system responsible for orchestrating the sleep-wake cycle [14]. Consequently, these modifications can exert a direct influence on the caliber of an individual's sleep.

It is noteworthy to mention that, despite the well-documented nature of age-related alterations in sleep patterns, such as disruptions in sleep continuity or shifts in sleep phases, an ongoing discourse persists regarding the appropriate classification of these changes as a distinct sleep disorder [15]. This ongoing dialogue underscores the intricate nature of the interrelationship between age and sleep quality and engenders inquiries about the categorization and management of these age-associated variations in sleep patterns.

4. Factor Analysis

Table 11. KMO and Barlett's Test for factors in dataset

KMO and Barlett's Test		
Kaiser - Meyer - Olkin Measure of Sampling Adequacy.		0.523
Bartlett's Test of Sphericity	**Approx. Chi-Square**	337.864
	df	28
	Sig.	<.001

DOI: 10.54254/2753-8818/44/20240844

Table 11. (continued).

Communalities		
	Initial	**Extraction**
Age	1.000	0.619
PhysicallyActive	1.000	0.890
BMI	1.000	0.543
Sleep	1.000	0.751
SoundSleep	1.000	0.833
JunkFood	1.000	0.803
Stress	1.000	0.591
BPLevel	1.000	0.512

Table 12. Eigenvalues for factors in dataset

Total Variance Explained						
	Initial Eigenvalues			**Extraction Sums of Squared Loadings**		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	1.977	24.718	24.718	1.977	24.718	24.718
2	1.498	18.723	43.441	1.498	18.723	43.441
3	1.056	13.206	56.447	1.056	13.206	56.447
4	1.011	12.636	69.284	1.011	12.636	69.284
5	0.85	10.623	79.907			
6	0.693	8.668	88.575			
7	0.617	7.714	96.289			
8	0.297	3.711	100			

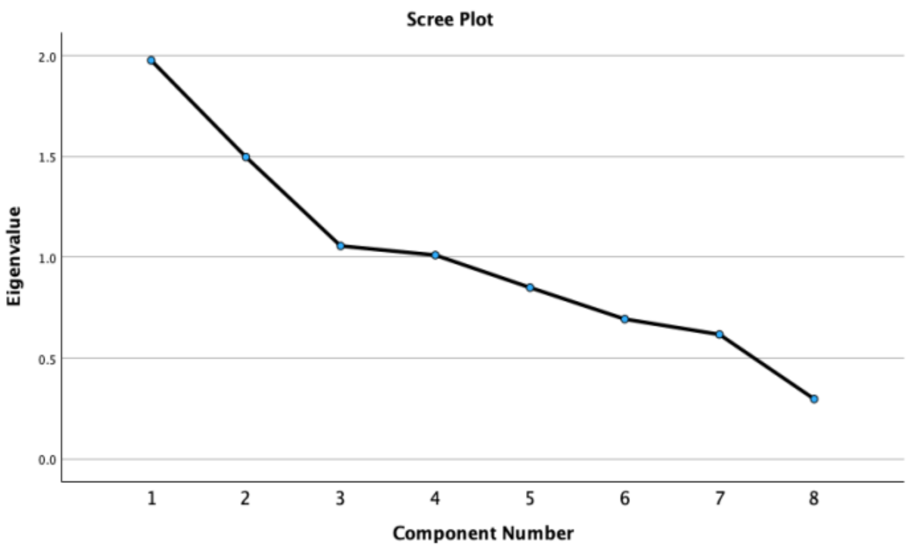

Figure 1. Eigenvalues for factors in dataset

DOI: 10.54254/2753-8818/44/20240844

Table 13. Component matrix for factors in dataset

Component Matrix				
	Components			
	1	**2**	**3**	**4**
Age	0.375	-0.647	0.022	0.244
PhysicallyActive	-0.177	-0.211	0.642	-0.634
BMI	-0.639	-0.088	0.087	0.345
Sleep	0.607	0.455	0.366	0.206
SoundSleep	0.861	0.213	0.213	0.011
JunkFood	-0.314	-0.156	0.669	0.483
Stress	-0.245	0.647	-0.005	0.276
BPLevel	-0.358	0.544	0.097	-0.281

Tables and figure above delineate the variability attributed to a set of meticulously chosen variables: Age, Physical Activity, BMI, Sleep, Quality of Sleep, Dietary Patterns, Stress, and Blood Pressure Levels. A comprehensive analysis employing the Kaiser-Meyer-Olkin (KMO) metric and Bartlett's Test has unveiled a statistically noteworthy similarity in the presented dataset, with physical activity, sleep patterns, and dietary behaviors emerging as the three foremost influential factors contributing to the development of diabetes.

Table 12 and Figure 1 highlights the eigenvalues of the individual components. Specifically, Component 1 exhibits an eigenvalue of 1.977, Component 2 an eigenvalue of 1.498, Component 3 an eigenvalue of 1.056, and Component 4 an eigenvalue of 1.011, among others. Components with eigenvalues exceeding 1 have been selected for further examination and have been visualized on the Scree Plot, illustrating a gradual descending trend. Particularly, after the fourth component, there is a noticeable sharp decline in eigenvalues, indicating a significant difference.

Additionally, based on Table 13, the Component Matrix table presents distinct coefficients for variables corresponding to Components 1 through 4, all of which possess eigenvalues surpassing 1. As we are selecting four components from the initial pool of eight variables, the contribution of each variable is notably reduced, with the remaining percentages outlined in the Communalities table.

Laboratory studies [16-22] conducted on healthy volunteers have revealed that experimental sleep restriction can have detrimental effects on glucose regulation in the body. Specifically, it leads to a rapid and substantial decrease in insulin sensitivity, without sufficient compensation in the functioning of beta cells. This imbalance in glucose control significantly elevates the risk of developing diabetes.

Prospective epidemiological studies, conducted in both children and adults, consistently point towards a causal relationship between insufficient sleep and the heightened risk of diabetes. Moreover, sleep curtailment disrupts the neuroendocrine regulation of appetite. It reduces the levels of the satiety hormone leptin while increasing the levels of the hunger-promoting hormone ghrelin. As a result, sleep loss may disrupt the accurate signalling of caloric needs by leptin and ghrelin, leading to an internal misperception of insufficient energy availability.

The adverse effects of sleep deprivation on appetite regulation are thought to be driven by increased activity in certain neuronal populations expressing excitatory peptides known as orexins, which promote both wakefulness and feeding. These findings align with multiple epidemiological studies that have consistently shown an association between shorter sleep durations and higher body mass index, even after accounting for various potential confounding factors. In summary, the accumulating evidence suggests that inadequate sleep may play a significant role in the development of obesity and diabetes, affecting both glucose regulation and appetite control.

5. Logistic Regression

Table 14. Model Summary for Logistic Regression

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	532.986[a]	0.460	0.657

Table 15. Classification Table for Logistic Regression

Classification Table		Diabetic (Predicted)		Percentage correct
		0	1	
Diabetic (Observed)	0	604	39	93.9%
	1	58	205	77.9%
Overall Percentage		91.24%	84.02%	89.3%

Table 16. Variables included in the Logistic Regression

Variables in the Equation					
Variable	B	Sig.	Variable	B	Sig.
High BP(1)	0.423	0.258	Junk Food(1)	-1.079	0.093
Age		0.000	Junk Food(2)	-0.442	0.482
Age(1)	1.826	0.000	Junk Food(3)	-0.258	0.750
Age(2)	2.110	0.000	Stress		0.030
Age(3)	3.467	0.000	Stress(1)	-0.250	0.632
Gender(1)	0.606	0.071	Stress(2)	-1.032	0.015
Family Diabetes(1)	-1.319	0.000	Stress(3)	-0.881	0.057
BMI	0.019	0.374	BP Level		0.006
Smoking(1)	-0.975	0.033	BP Level(1)	-19.432	0.998
Alcohol(1)	0.478	0.175	BP Level(2)	-1.140	0.001
Sleep	0.027	0.807	Pregnancies	0.354	0.024
Sound Sleep	0.243	0.006	UriationFreq(1)	-0.035	0.902
RegularMedicine(1)	-2.428	0.000	Constant	-0.494	0.659
Junk Food		0.087			

The tables presented above underwent logistic regression analysis, encompassing a range of factors including age, gender, family history of diabetes, blood pressure, BMI, smoking habits, alcohol consumption, sleep patterns, quality of sleep, regular medication usage, consumption of junk food, stress levels, history of pregnancies, and urination frequency. Table 15 reveals remarkable results, with 39 non-diabetic individuals being incorrectly classified as diabetic, yielding a correct classification rate of 93.9%. Similarly, 58 diabetic individuals were erroneously classified as non-diabetic, resulting in a correct classification rate of 77.9%. Consequently, the overall correct classification rate stands at 89.3%, suggesting that this model exhibits a strong predictive capability for diabetes.

Examining Table 16, we observe distinct coefficients and variations associated with each factor. This analysis highlights that certain factors significantly influence the ultimate prediction. For instance, all age groups exhibit an asymptotic significance below 0.05, signifying that age is one of the most influential factors in the model's predictions. Conversely, BMI, with an asymptotic significance of 0.374, does not exert a deterministic influence and may not be considered a causative factor in the development of diabetes.

Previous research [23] has indicated that when it comes to long-term weight loss maintenance and improved glycaemic control, combining diet and physical activity interventions is more effective than implementing either one in isolation [24]. Furthermore, engaging in physical activity can trigger positive changes in other behaviors; individuals who lead more active lifestyles tend to adopt healthier dietary habits and are less likely to smoke [25].

This favourable outcome of combining diet and exercise stands in contrast to certain scenarios where pursuing two behavioral goals concurrently leads to diminished results in each behaviour. For example, in the treatment of hypertension, patients instructed to simultaneously adhere to a low-sodium diet and lose weight tend to have lower compliance compared to when these changes are introduced separately [26].

In the context of smoking cessation programs, efforts have been made to address concerns about weight gain following smoking cessation by incorporating other lifestyle changes into the program. However, studies attempting to combine smoking cessation with weight loss interventions have not yielded significant success [27]. In contrast, a recent study that combined smoking cessation with physical activity has shown promise [28]. Additionally, there is evidence that targeting both alcohol consumption and smoking can enhance abstinence rates for both behaviors [29].

Surprisingly, there has been limited research on how to effectively combine lifestyle modifications with pharmacological treatments to maximize compliance with both regimens and enhance overall effectiveness [30]. Given the growing interest in drug treatments for obesity, such research holds significant importance and warrants further exploration.

6. Conclusion

The investigation of diabetes holds paramount importance in advancing our understanding of its etiology, prevention, and therapeutic approaches, thereby positively impacting the quality of life for millions of individuals worldwide. Our comprehensive analysis, employing a diverse array of methodologies, including comparative analysis, descriptive statistics, factor analysis, and logistic regression, has yielded invaluable insights.

By analyzing crosstabs and applying Chi-Square Tests for different variables on individuals in section 2, we discerned a direct correlation between blood pressure levels and stress, while also establishing a proportional relationship between BMI and the likelihood of diabetes. Additionally, non-parametric test on the distribution of sound sleep under various age level, as shown in section 3, has also illuminated an inverse correlation between age and sleep quality, with individuals aged around 50 experiencing the most diminished sleep quality.

The factor analysis emphasized the roles of physical activity, dietary habits, especially the consumption of junk food, and sleep quality in the context of diabetes. Finally, our logistic regression model, boasting an impressive 89.3% accuracy rate, pinpointed age, family history of diabetes, blood pressure, and medication adherence as pivotal factors in diabetes development. Nonetheless, the imperative for further extensive analysis, involving a larger and more diverse dataset, remains pronounced to fortify the foundation for more robust recommendations in this globally significant realm.

Acknowledgement

Eric Zhuang and Yiyu Chen contributed equally to this work and should be considered co-first authors.

References

[1] Munakata, M. Clinical significance of stress-related increase in blood pressure: current evidence in office and out-of-office settings. Hypertens Res 41, 553–569 (2018). https://www.nature.com/articles/s41440-018-0053-1

[2] Gray N, Picone G, Sloan F, Yashkin A. Relation between BMI and diabetes mellitus and its complications among US older adults. South Med J. 2015 Jan;108(1):29-36. doi: 10.14423/SMJ.0000000000000214. PMID: 25580754; PMCID: PMC4457375.

[3] Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med. 2014 Jan;44(1):81-121. doi: 10.1007/s40279-013-0090-5. PMID: 24030837; PMCID: PMC3894304.

[4] Moraska, A., & Fleshner, M. (2001). Voluntary physical activity prevents stress-induced behavioral depression and anti-KLH antibody suppression. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. https://doi.org/10.1152/ajpregu.2001.281.2.R484

[5] Fleshner M.Exercise and neuroendocrine regulation of antibody production: protective effect of physical activity on stress-induced suppression of the specific antibody response. Int J Sports Med21, Suppl12000S14S19

[6] Fleshner M, Brennan FX, Nguyen K, Watkins LR, Maier SF.RU-486 blocks differentially suppressive effect of stress on in vivo anti-KLH immunoglobulin response. Am J Physiol Regulatory Integrative Comp Physiol2711996R1344R1352

[7] Fleshner M, Hermann J, Lockwood LL, Laudenslager ML, Watkins LR, Maier SF.Stressed rats fail to expand the CD45RC+CD4+ (Th1-like) T cell subset in response to KLH: possible involvement of IFN-gamma. Brain Behav Immun91995101112

[8] Maier SF, Silbert LH, Woodmansee WW, Desan PH.Adinazolam both prevents and reverses the long-term reduction of daily activity produced by inescapable shock. Pharmacol Biochem Behav361990767773

[9] Feng N, Pagniano R, Tovar CA, Bonneau RH, Glaser R, Sheridan JF.The effect of restraint stress on the kinetics, magnitude, and isotype of the humoral immune response to influenza virus infection. Brain Behav Immun51991370382

[10] Madrid-Valero, J. J., Martínez-Selva, J. M., Ribeiro do Couto, B., Sánchez-Romera, J. F., & Ordoñana, J. R. (2017). Age and gender effects on the prevalence of poor sleep quality in the adult population. Science Direct, 31(1), 18-22. Retrieved from https://www.sciencedirect.com/science/article/pii/S0213911116301182?via%3Dihub

[11] Uhlig BL, Sand T, Odegård SS, et al. Prevalence and associated factors of DSM-V insomnia in Norway: the Nord-Trøndelag Health Study (HUNT 3). Sleep Med. 2014;15:708-13.

[12] Xu M, Bélanger L, Ivers H, et al. Comparison of subjective and objective sleep quality in menopausal and non-menopausal women with insomnia. Sleep Med. 2011;12:65-9.

[13] Ohayon MM, Carskadon MA, Guilleminault C, et al. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep. 2004;27: 1255-73.

[14] Carrier J, Paquet J, Morettini J, et al. Phase advance of sleep and temperature circadian rhythms in the middle years of life in humans. Neurosci Lett. 2002;320:1-4.

[15] Grupo de Trabajo de la Guía de Práctica Clínica para el Manejo de Pacientes con Insomnio en Atención Primaria. Guía de práctica clínica para el manejo de pacientes con insomnio en atención primaria. Plan de Calidad para el Sistema Nacional de Salud del Ministerio de Sanidad y Política Social. Unidad de Evaluación de Tecnologías Sanitarias. Agencia Laín Entralgo. Comunidad de Madrid; 2009. Guías de Práctica Clínica en el SNS: UETS Nº 2007/5-1.

[16] Knutson, Kristen L., and Eve Van Cauter. “Associations between Sleep Loss and Increased Risk of Obesity and Diabetes.” Annals of the New York Academy of Sciences, vol. 1129, no. 1, 2008, pp. 287-304. https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1196/annals.1417.033

[17] Kelly MR, Zeineddine S, Mitchell MN, et al. Insomnia severity predicts depression, anxiety, and posttraumatic stress disorder in veterans with spinal cord injury or disease: a cross-sectional observational study. J Clin Sleep Med. 2023;19(4):695–701

[18] Flegal, K.M.; Kit, B.K.; Orpana, H.; Graubard, B.I. Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis. JAMA 2013, 309, 71–82
[19] Qian, J.; Scheer, F.A. Circadian system and glucose metabolism: Implications for physiology and disease. Trends Endocrinol. Metab. 2016, 27, 282–293
[20] Arora, T.; Taheri, S. Associations among late chronotype, body mass index and dietary behaviors in young adolescents. Int. J. Obes. 2015, 39, 39–44
[21] Krejcie, R.V.; Morgan, D.W. Determining sample size for research activities. Educ. Psychol. Meas. 1970, 30, 607–610
[22] Buysse, D.J.; Reynolds, I.I.I.C.F.; Monk, T.H.; Berman, S.R.; Kupfer, D.J. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res. 1989, 28, 193–213
[23] Rena R. Wing, Michael G. Goldstein, Kelly J. Acton, Leann L. Birch, John M. Jakicic, James F. Sallis, Delia Smith-West, Robert W. Jeffery, Richard S. Surwit; Behavioral Science Research in Diabetes: Lifestyle changes related to obesity, eating behavior, and physical activity. Diabetes Care 1 January 2001; 24 (1): 117 123. https://doi.org/10.2337/diacare.24.1.117
[24] NHLBI Obesity Education Initiative Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults:Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report. Obes Res 6 (Suppl. 2): 51S-210S, 1998
[25] Wankel L, Sefton J: Physical activity and other lifestyle behaviors. In Physical Activity, Fitness, and Health Campaign. Bouchard C, Shepard R, Stephens T, Eds. Champaign, IL,Human Kinetics, 1994, p. 530-550
[26] The Hypertension Prevention Trial Research Group: The Hypertension Prevention Trial (HPT): three-year effects of dietary changes on blood pressure. Arch Intern Med 150: 153-162, 1990
[27] Hall SM, Tunstall CD, Vila KL, Duffy J: Weight gain prevention and smoking cessation: cautionary findings. Am J Public Health 82: 799-803, 1992Close Modal
[28] Marcus BH, Albrecht AE, King TK, Parisi AF, Pinto BM, Roberts M,Niaura RS, Abrams DB: The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial. Arch Intern Med 159: 1229-1234, 1999
[29] Hughes J: Clinical implications of the association between smoking and alcoholism. In Alcohol and Tobacco: From Basic Science to Policy (NIAAA Research Monograph 30). Fertig J, Fuller R, Eds. Washington, DC, U.S. Government Printing Office, 1995, p. 171-181
[30] adden TA, Berkowitz R, Vogt RA, Steen SN, Stunkard AJ, Foster GD:Lifestyle modification in the pharmacologic treatment of obesity: a pilot investigation of a potential primary care approach. Obes Res 5: 218-226, 1997