Article

Machine learning-based diagnosis of Type 2 Diabetes Mellitus using Social Determinants of Health

Supplementary material 3

I. SDoH variables

Awa's study [68] showed that advanced age, obesity, hypertension, dyslipidemia, and male gender are important risk factors for T2DM.

Research [69,70] has shown that individual-level social factors, such as education and annual income, are consistently associated with the risk of T2DM. and

Wang's study [71] showed that age, smoking, alcohol consumption, regular exercise, hypertension, gender, BMI, and the incidence of T2DM are closely related.

Research has shown that urbanization [72,73] impacts the risk of developing type 2 diabetes. -Residence

Some studies [74,75] have shown that, among middle-aged and older adults, unmarried individuals are more likely to develop diabetes compared to their married counterparts. -Marital status

Evidence from numerous laboratory and epidemiological studies [76,77] suggests that reduced sleep duration or poor sleep quality may increase the risk of developing diabetes.

Sedentary behavior is significantly associated with an increased risk of diabetes, while higher levels of physical activity are linked to a lower incidence of the disease [78,79].

Research has shown that self-rated health scores are independently positively correlated with the development of T2DM [80] and can be used to predict the risk of T2DM [81].

II. Data preprocessed

Table S1. Data preprocessed.

Variable name	Variable assignment	Missing Data imputed	Outliers excluded
Dependent variable			
T2DM	No T2DM = 0, T2DM = 1	89	0
Demographic variable			
Age	Continuous variables	0	0
BMI	Continuous variables	59	86
Residence	Rural area = 0, Urban area = 1	10	0
Gender	Female = 0, Male = 1	9	0
Education	Below primary school = 0, Primary school = 1, Junior high school = 2, Junior college and above = 3	6	1
Annual income	Lower income = 0 (under 20,442 yuan); Middle income = 1(20,442 ≤ income ≤ 50,220); Higher income = 2(over 50,220 yuan)	43	54
Marital status	Solitary = 0, Cohabitation = 1	19	0
Medical insurance	None = 0, Yes = 1	8	0
Lifestyle variable			
Daily sleep duration	Continuous variables	36	21
Sleep quality	Bad = 0, Average = 1, Good = 2	5	0

Sedentary time	Continuous variables	10	35
PA	Low PA = 0, Moderate PA = 1, High PA = 2	16	2
Drinking status	<1/month, 1–4/month, Every week	0	0
Smoking status	Never = 0, Former = 1, Current = 2	1	0
Physiological health variable			
Chronic disease	None = 0, Yes = 1	0	1
Hypertension	None = 0, Yes = 1	0	0
Self-rated health	Bad = 0, Average = 1, Good = 2	0	0
Dyslipidemia	None = 0, Yes = 1	0	0
Central obesity	None = 0, Yes = 1	5	0