



Article

Spatiotemporal Analysis and Risk Assessment Model Research of Diabetes among People over 45 Years Old in China

Zhenyi Wang ^{1,2}, Wen Dong ^{1,2,*} and Kun Yang ^{1,2,*}

¹ Faculty of Geography, Yunnan Normal University, Kunming 650500, China

² GIS Technology Engineering Research Centre of West-China Resources and Environment of Educational Ministry, Yunnan Normal University, Kunming 650500, China

* Correspondence: kmynnu_gis@163.com (W.D.); kmdcynu@163.com (K.Y.);
Tel.: +86-0871-65941202 (W.D. & K.Y.)

Abstract: Diabetes, which is a chronic disease with a high prevalence in people over 45 years old in China, is a public health issue of global concern. In order to explore the spatiotemporal patterns of diabetes among people over 45 years old in China, to find out diabetes risk factors, and to assess its risk, we used spatial autocorrelation, spatiotemporal cluster analysis, binary logistic regression, and a random forest model in this study. The results of the spatial autocorrelation analysis and the spatiotemporal clustering analysis showed that diabetes patients are mainly clustered near the Beijing–Tianjin–Hebei region, and that the prevalence of diabetes clusters is waning. Age, hypertension, dyslipidemia, and smoking history were all diabetes risk factors ($p < 0.05$), but the spatial heterogeneity of these factors was weak. Compared with the binary logistic regression model, the random forest model showed better accuracy in assessing diabetes risk. According to the assessment risk map generated by the random forest model, the northeast region and the Beijing–Tianjin–Hebei region are high-risk areas for diabetes.



Citation: Wang, Z.; Dong, W.; Yang, K. Spatiotemporal Analysis and Risk Assessment Model Research of Diabetes among People over 45 Years Old in China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9861. <https://doi.org/10.3390/ijerph19169861>

Academic Editor: Paul B. Tchounwou

Received: 21 June 2022

Accepted: 8 August 2022

Published: 10 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: spatiotemporal analysis; risk factors; binary logistic regression; random forest model

1. Introduction

1.1. Background

In the past few decades, diabetes has become one of the most common chronic noncommunicable diseases in both developed and developing countries [1]. Diabetes is emerging as an epidemic all over the world, and it is a common chronic disease that seriously threatens human health [2]. It affects the quality of lives of many people around the world [3], and the quality of life for Chinese residents is also affected by diabetes. China has a large and rapidly growing elderly population. Studies have shown that diabetes may also lead to the occurrence of other diseases, such as metabolic-associated fatty liver disease [4–8]. Diabetes has become another serious health hazard, following cardiovascular and cerebrovascular diseases and tumors. Half (50.1%) of the population does not even know if they are diabetic, which greatly increases the global disease burden [9]. According to data published by the International Diabetes Federation (IDF), the prevalence of diabetes is increasing rapidly around the world. According to IDF estimates, the prevalence of diabetes in China has reached 10.6%, with the proportion of undiagnosed diabetics as high as 51.7% [10].

Disease mapping has been historically considered one of the most important public health issues, derived from an understanding of the relationship between health and location. Understanding this relationship has been the goal of scientists and researchers for decades [11]. Geographic information systems (GIS) are a type of computer software used for data capturing, thematic mapping, updating, retrieving, structured querying, and analyzing the distribution and differentiation of various phenomena, including communicable and non-communicable diseases across the world, with reference to various periods [12]. The most important characteristic of a geographic information system is its powerful spatial

analysis function. Nowadays, geographic information systems have played an irreplaceable role in many aspects of daily life. A GIS is, at its heart, a simple extension of statistical analyses that joins epidemiological, sociological, clinical, and economic data with references to space [13].

1.2. Research Status

The GIS approach has the potential for broader applications within public health program evaluation [14,15]. With the rapid development of GIS systems and related technologies, the advantages that GIS provides for the study of chronic diseases have been gradually recognized, and the application scope has also transitioned from infectious diseases to chronic diseases [16,17]. Some scholars have applied GIS to diabetes research and proposed that geospatial methods should be a part of diabetes research because many pathogenic pathways have inherent spatial properties [18]. GIS can be used to map the geographical distribution of disease prevalence, the trend of disease transmission, and the spatial modeling of environmental factors influencing disease occurrence [11].

Although diabetes is a health threat all over the world, its prevalence and trends in various countries and regions are heterogeneous [19]. Previous studies have showed that the prevalence of diabetes among middle-aged and elderly people in the central and eastern regions is higher than in the western regions, but the gap was closing [20]. At present, studies on diabetes in Chinese people over 45 years old are mostly regional or related to a single province, but the number of nationwide studies is lacking [18]. Moreover, GIS is seldom used to study the spatial patterns of diabetes [21]. Recent studies in the health field have adopted machine learning and deep learning algorithms. Since machine-learning approaches perform well in predicting diabetes, they are gaining traction in the health profession [22,23]. This research hoped to analyze the regional differences of diabetes among people over 45 years old in China, and to assess diabetes risk [24], thereby aiming to provide reference for the formulation of diabetes prevention and treatment programs.

2. Materials and Methods

2.1. Data Source

This study is based on the baseline data of the China Health and Retirement Longitudinal Study (CHARLS). The China Health and Retirement Longitudinal Study is part of a worldwide pension tracking survey. This database is one of the most commonly used databases in China to study the health of the middle-aged and older population, and provides high-quality microdata representing households and individuals aged ≥ 45 years in China. Many scholars have obtained many reliable research results based on CHARLS [25–28].

The China Health and Retirement Longitudinal Study (CHARLS) aims to collect a high-quality nationally representative sample of Chinese residents ages 45 and older to serve the needs of scientific research on the elderly. The baseline national wave of CHARLS was established in 2011 and includes about 10,000 households in 125 prefecture-level city and 450 villages/resident committees. CHARLS adopts multi-stage stratified probability-proportional-to-size sampling. CHARLS is based on the Health and Retirement Study (HRS) and on related aging surveys such as the English Longitudinal Study of Aging (ELSA) and the Survey of Health, Aging and Retirement in Europe (SHARE) [29].

2.2. Diabetes Definition

Prevalence refers to the proportion of the total number of people who have the disease at a specific point in time in a given place. Diabetes was defined as: fasting glucose level ≥ 126 mg/dL (7.0 mmol/L), or 2-h glucose level ≥ 200 mg/dL (11.1 mmol/L), or on medications for high blood sugar, or self-reported diagnosis of diabetes by a physician.

2.3. Methods

2.3.1. Spatial Autocorrelation

Global Spatial Autocorrelation statistics are often expressed as Moran's I (Equation (1)). According to the literature, the classical Moran's index of Spatial Autocorrelation has been widely used in many knowledge fields, such as epidemiology, ecology, and economics [30]. The index was used to explore the overall spatial pattern of disease prevalence. When the Moran index is between 0 and 1, it indicates that there is a positive correlation between geographical entities. The larger the value, the more obvious the spatial correlation. When the Moran index is between -1 and 0, there is a negative correlation. The smaller the Moran index, the greater the spatial difference. A value of 0 indicates no correlation. In addition, the value also needs to pass the hypothesis test, without which, the Moran index is meaningless.

$$I = \frac{\sum_i \sum_j W_{ij} Z_i Z_j / S_0}{\sum_i Z_i^2 / n} \quad (1)$$

where $Z_i = y_i - \bar{y}$, where $\bar{y}$ is the mean of the variable y representing the observations under study, W_{ij} is the spatial weight between feature i and j , and S_0 is the sum of all the elements in the spatial weights matrix ($S_0 = \sum_i \sum_j W_{ij}$) [31].

Getis and Ord's G^* assessed localized patterns of spatial association. Specifically, Getis and Ord's G^* can indicate regions where low values are clustered ($G^* > 0$) and regions where high values are clustered ($G^* < 0$) [32]. Local Spatial Autocorrelation can accurately indicate the aggregation mode of each spatial unit [33]. Generally, Local Spatial Autocorrelation analysis (LISA) is used. LISA had five results of "high-high" (H-H), "low-low" (L-L), "low-high" (L-H), "high-low" (H-L), and no statistical significance [34]. Respectively, the regions with high prevalence surround the regions with high prevalence, the regions with low prevalence surround the regions with low prevalence, the regions with low prevalence surround the regions with high prevalence and the regions with high prevalence surround the regions with low prevalence. In this study, Moran's I and LISA plots were calculated for the prevalence of diabetes in members of the Chinese population over 45 years old in 2011, 2013, 2015, and 2018, respectively. ArcGIS 10.4 software (ESRI Inc., Redlands, CA, USA) was used in this study.

2.3.2. Spatial Cluster Analysis

Temporal, spatial, and spatiotemporal scan statistics are now commonly used for disease cluster detection and assessment for a variety of diseases, including cancer, Creutzfeldt-Jakob disease, granulocytic ehrlichiosis, sclerosis, and diabetes. Spatial clustering analysis was performed using SaTScan software (Martin Kulldorff, Harvard Medical School, Boston and Information Management Services Inc, Calverton, MD, USA) to detect spatially clustered areas or high-risk areas of diabetes in members of the Chinese population over 45 years old. The "purely spatial analysis" and "space time analysis" were used to test whether the prevalence of diabetes was randomly distributed in space. To avoid preselection bias as described in the SaTScan User Guide (version 9.1) [35], a maximum spatial cluster size of 10% of the population at risk was used.

2.3.3. Binary Logistic Regression

Binary logistic regression is a linear regression analysis in which the dependent variable is a binary classification variable, requiring logit transformation of the target probability first, so as to ensure that when the probability is at (0, 1), the logit transformation value can be any real number, avoiding the structural defects of the linear probability model. The probability of each classification of a classification variable can be predicted by logistic regression. The dependent variable is a classification variable, and the independent variable can be an interval variable, a classification variable, or a mixture of the interval and the classification variable. Binary logistic regression model is a regression model established for binary variables, such as Equation (2) [36], which can capably meet the modeling require-

ments of classified data. It has become a commonly used modeling method for classifying variables and has been widely used in many fields, such as medicine. We used IBM SPSS Statistics 26 software (IBM Corp., Armonk, NY, USA) and the test level $\alpha = 0.05$ was used in this study.

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_i X_i \quad (2)$$

Suppose a survey of diabetes for conditional probability $P_i = P(Y_i = 1 | X_i)$, according to the type of binary logistic regression model assumes that the probability expression as shown in Equation (3).

$$P_i = \frac{1}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_i X_i)}} = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \quad (3)$$

2.3.4. Geographically Weighted Regression

The geographically weighted regression (GWR) (Equation (4)) is a statistical technique that is used to model heterogeneous spatial processes. It has high accuracy in analyzing location-affected relationships [37].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^n \beta_k(u_i, v_i) x_i k + \varepsilon_i \quad (4)$$

where (u_i, v_i) denotes the coordinates of the i -th point in space, $\beta_k(u_i, v_i)$ is the regression coefficient of each variable at point i , $\beta_0(u_i, v_i)$ is a constant term, ε_i is the random error term at point i , and n is the number of independent variables.

GWR is a local modeling tool based on the optimization of global regression models, which complements the global model by providing a set of coefficients for each geographic unit to determine the spatial variability of the observations [38]. GWR was used to explore the spatial heterogeneity of risk factors in this study.

2.3.5. Random Forest Model

The random forest algorithm can deal with nonlinear problems, has good anti-noise ability, and tends to avoid overfitting. Compared with the traditional multiple linear regression model, the random forest algorithm does not need to set the function form in advance and overcome the complex interaction between covariables [39]. The building blocks of the decision tree-based modeling approach, the random forest model, are bootstrapped and are called bagged aggregates. Random forest models randomly use bagging to identify features, thereby separating each node by selecting the most critical possible to assess or predict variables, which will improve the model's accuracy without causing overfitting. At present, the random forest model has been widely applied to predict and assess soil moisture, shallow water level, hydrology, and environmental management. In a random forest, factors with a significant influence on logistic regression are included as independent variables into random forest modeling [40], and the presence of diabetes is set as the dependent variable. The total data are divided into a training set and test set according to 7:3. The model parameters are trained through the training set for the assessment of the test set.

3. Results

3.1. Statistical Analysis and Spatial Distribution

In 2011, a total of 20,525 samples were included, including 1088 cases, with a prevalence of 5.30%. In 2013, a total of 20,525 samples were included, including 1333 cases, with a prevalence of 6.49%; In 2015, a total of 20,525 samples were included, including 1766 cases, with a prevalence of 8.60%. In 2018, a total of 18,174 samples were included, including 1032 cases, with a prevalence of 5.68%.

As shown in Figure 1, the highest prevalence of diabetes was in 2015. The overall prevalence of the respondents was 8.60%, of which, the prevalence of male respondents was 7.44% and the prevalence of female respondents was 9.74%; the lowest prevalence of diabetes was in 2011, when the overall prevalence of the respondents was 5.30%, of which, the prevalence of male respondents was 4.68% and the prevalence of female respondents was 5.91%. In addition, the survey data showed that the prevalence of female respondents was higher than that of male respondents.

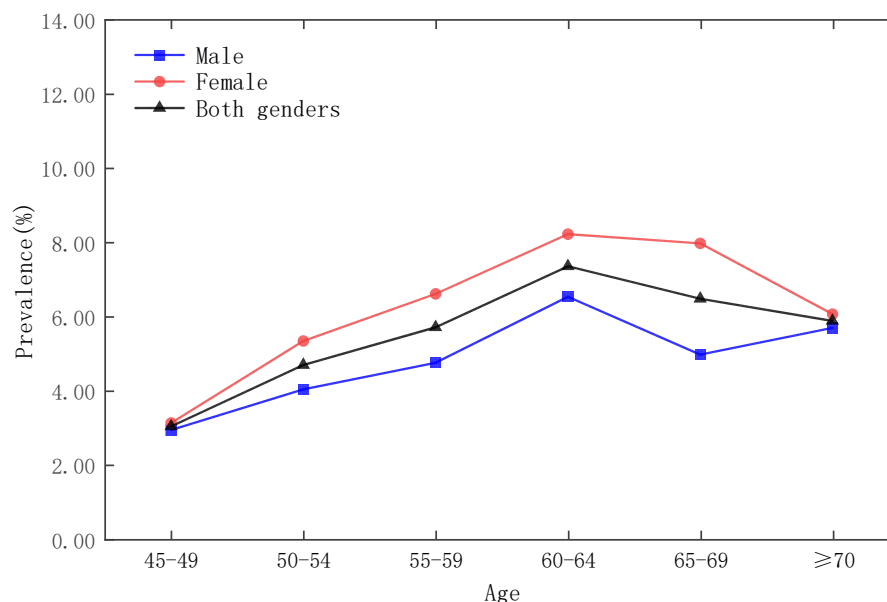


Figure 1. Prevalence of diabetes by gender in 5-year age groups in the CHARLS 2011 national survey.

The survey respondents are stratified according to age groups, as shown in Figures 2–5, which show that the age group with the lowest prevalence of respondents was 45 to 49 years old, the age groups with the highest prevalence of respondents were 60 to 64 years old and 65–69 years old, and the prevalence of female respondents was higher than that of male respondents in almost any age group.

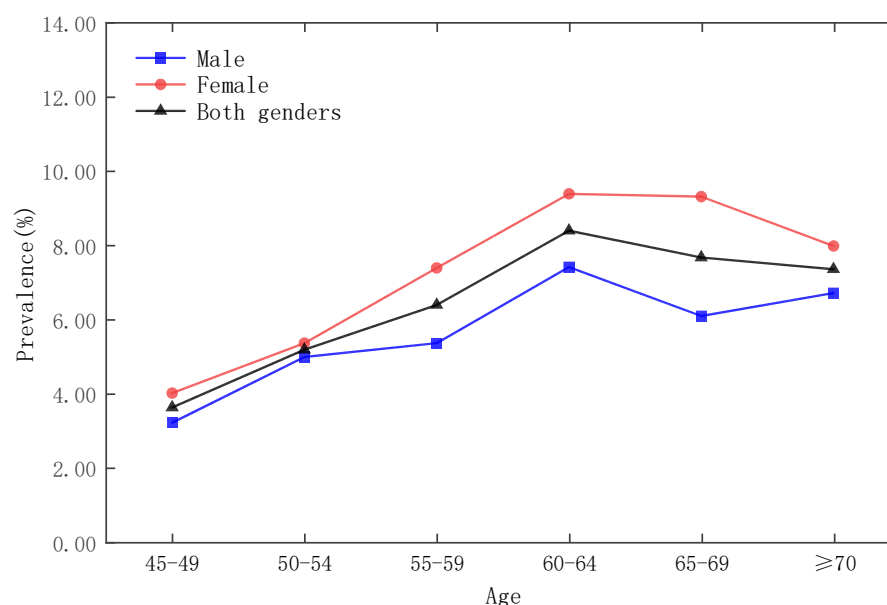


Figure 2. Prevalence of diabetes by gender in five-year age groups in the CHARLS 2013 national survey.

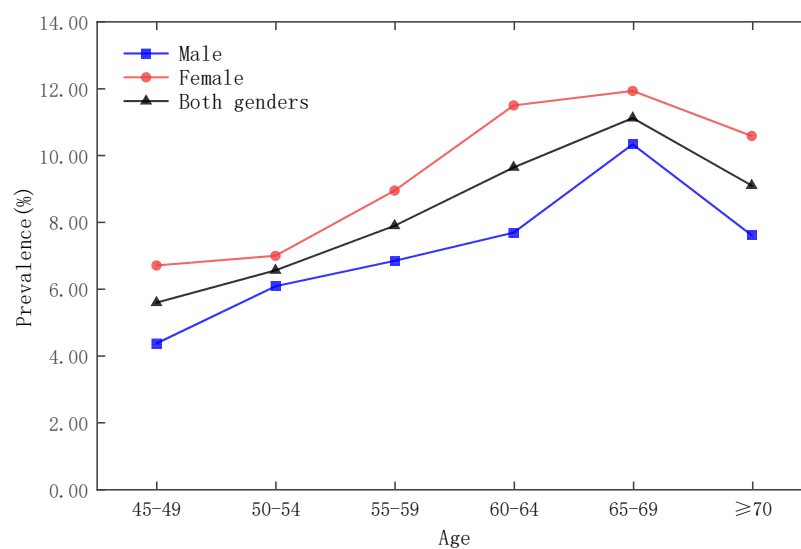


Figure 3. Prevalence of diabetes by gender in five-year age groups in the CHARLS 2015 national survey.

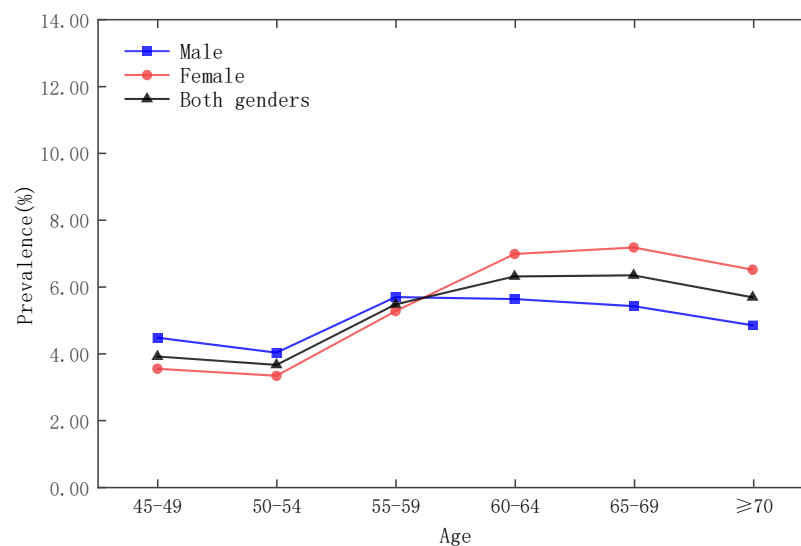


Figure 4. Prevalence of diabetes by gender in five-year age groups in the CHARLS 2015 national survey.

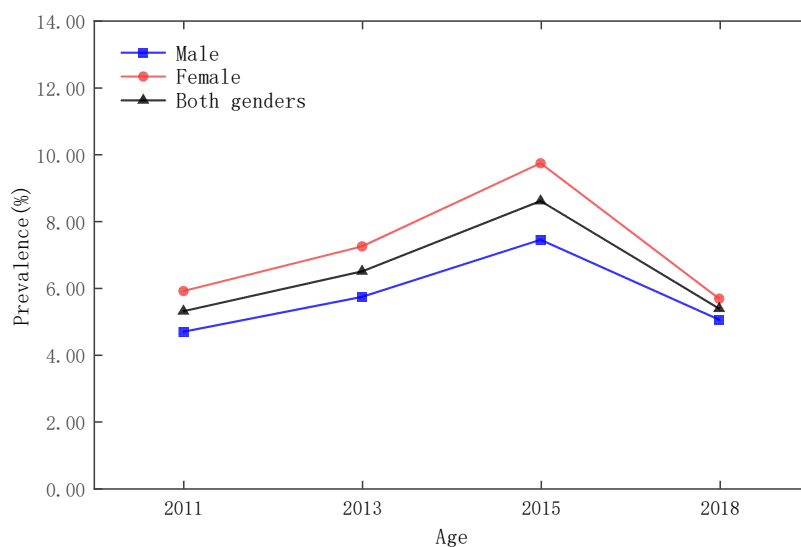
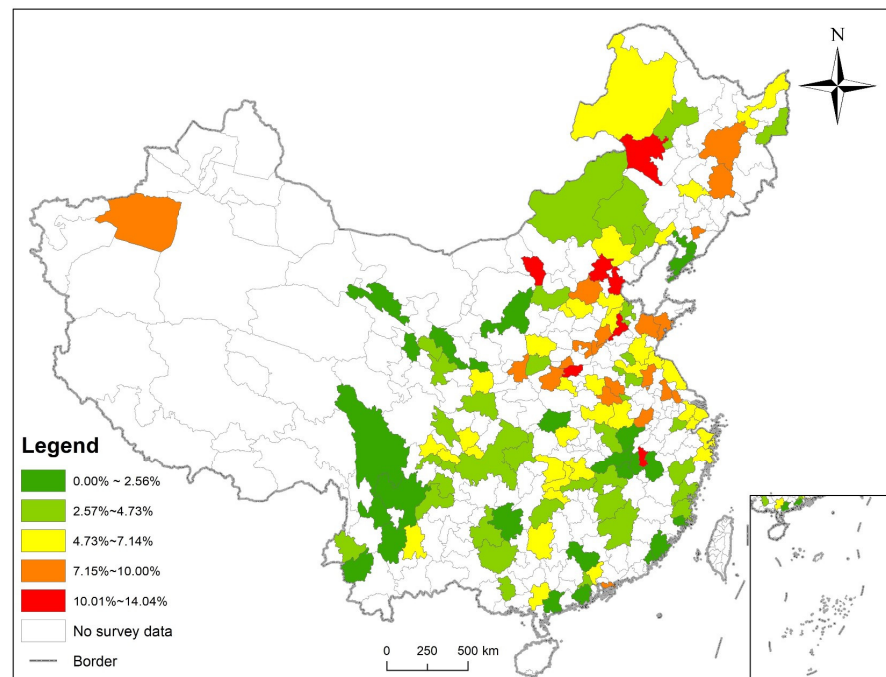
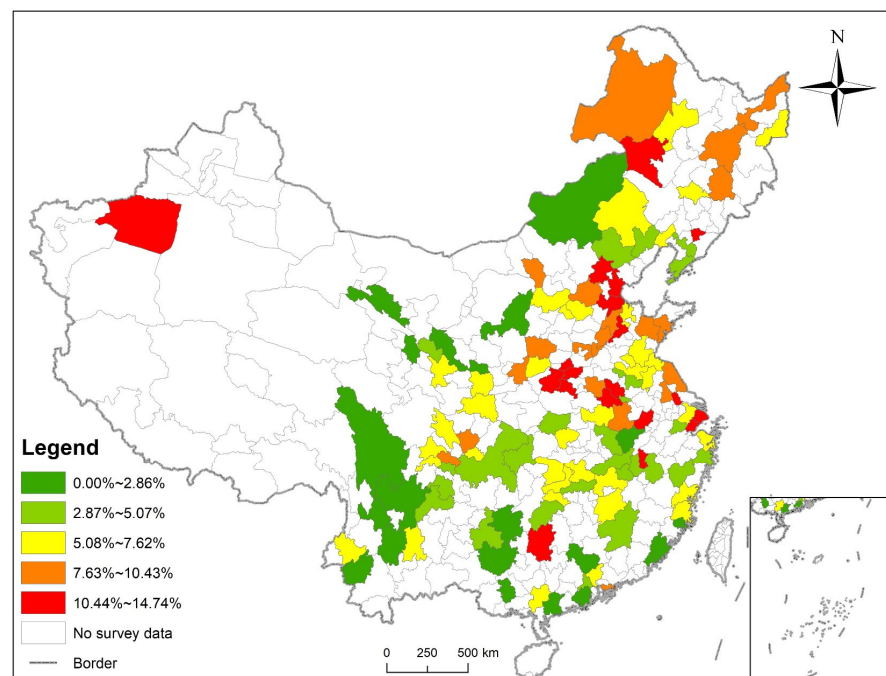


Figure 5. Prevalence of diabetes by age in 2011, 2013, 2015, and 2018.

The prevalence of diabetes in 2011, 2013, 2015, and 2018 were calculated according to the sampled 125 prefecture-level administrative regions, and visualized using ArcGIS 10.2. The results are shown in Figure 6.

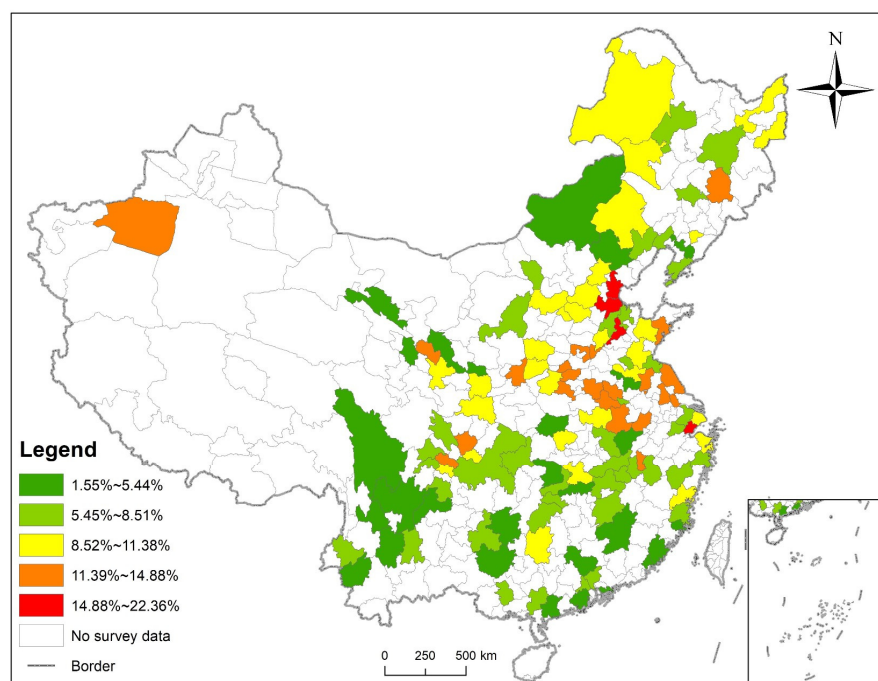


(a)

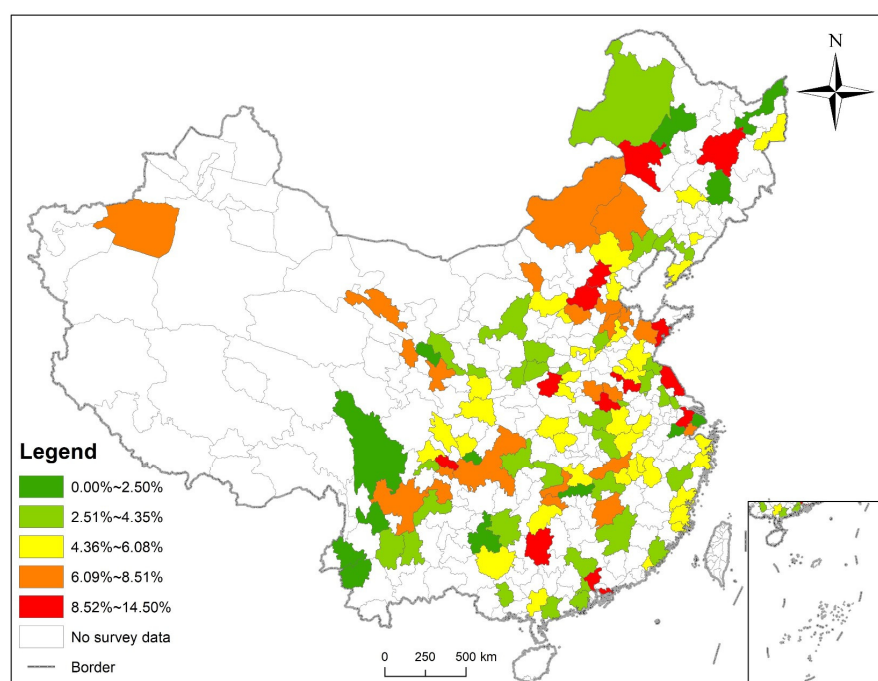


(b)

Figure 6. Cont.



(c)



(d)

Figure 6. Prevalence of diabetes visualized. (a) The prevalence of diabetes in 2011 was divided into five classifications according to the natural breaks method; (b) The prevalence of diabetes in 2013 was divided into five classifications according to the natural breaks method; (c) The prevalence of diabetes in 2015 was divided into five classifications according to the natural breaks method; (d) The prevalence of diabetes in 2018 was divided into five classifications according to the natural breaks method.

In 2011, the prevalence of diabetes in the respondents was between 0.00% and 14.04%, and the prefecture-level cities with higher prevalence were mainly located in the northeast region and Beijing–Tianjin–Hebei region. In 2013, the prevalence of diabetes in the respondents was between 0.00% and 14.74%, and the prefecture-level cities with higher prevalence were mainly located in the central region, the northeast region and Beijing–Tianjin–Hebei region. In 2015, the prevalence of diabetes in the respondents was between 1.55% and 22.36%, and the prefecture-level cities with high prevalence were mainly located in the Beijing–Tianjin–Hebei region. In 2018, the prevalence of diabetes in the respondents was between 0.00% and 14.50%, and prefecture-level cities with high prevalence were distributed in the central region and the northeast region. The prevalence of diabetes is generally higher in the north than in the south, and in the coastal areas than in the inland [18].

3.2. Spatial Autocorrelation Analysis

Hotspot analysis was performed on the prevalence of diabetes of respondents in prefecture-level cities in 2011, 2013, 2015, and 2018, and their LISA maps were also made. The results are shown in Figures 7–10, combined with global spatial autocorrelation (Table 1), showing that in 2011, 2013, 2015 and 2018, the prevalence of diabetes was clustered in China. The four-year prevalence hotspots appeared near the Beijing–Tianjin–Hebei region, and the Beijing–Tianjin–Hebei region has experienced high-value clusters of diabetes prevalence for four years according to LISA. However, Moran's index decreased after 2013. Many hot and cold spots became not significant after 2013. High-High or Low-Low distribution areas also decreased slowly.

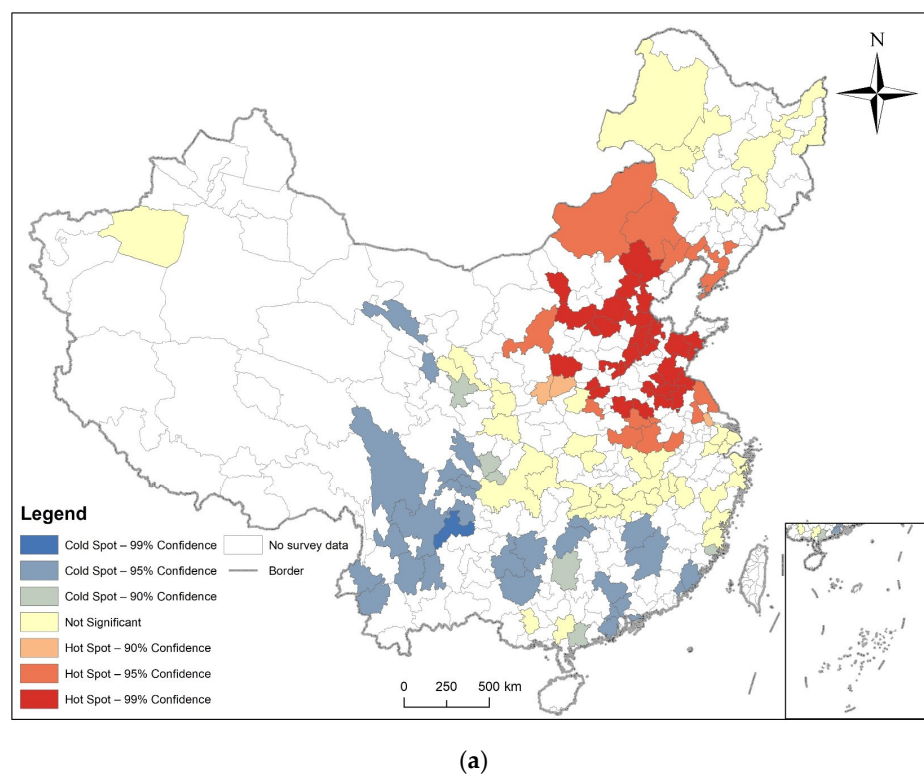
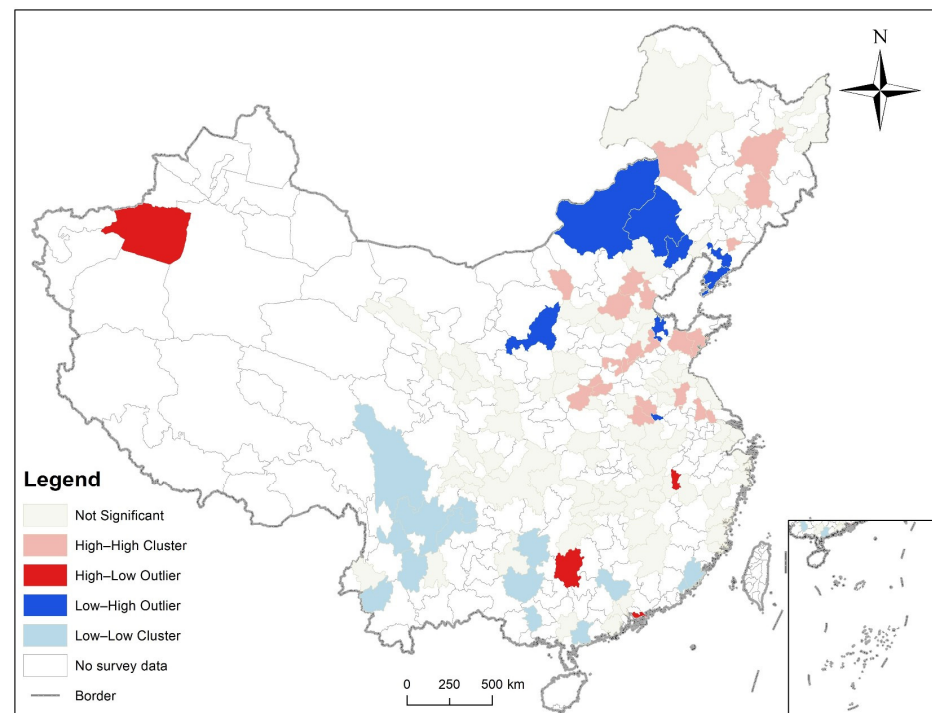
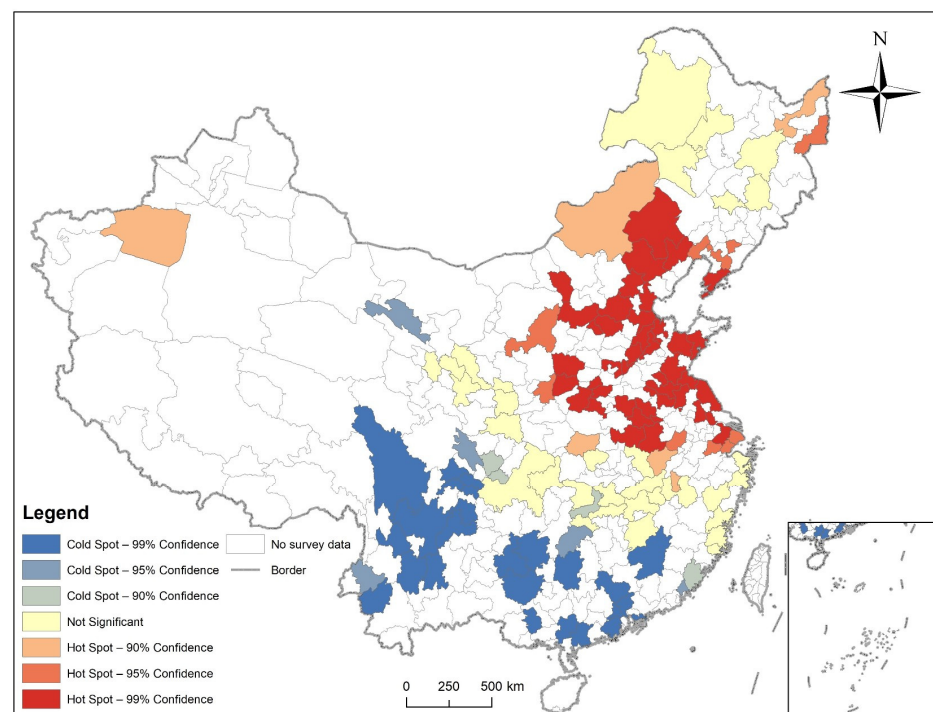


Figure 7. Cont.



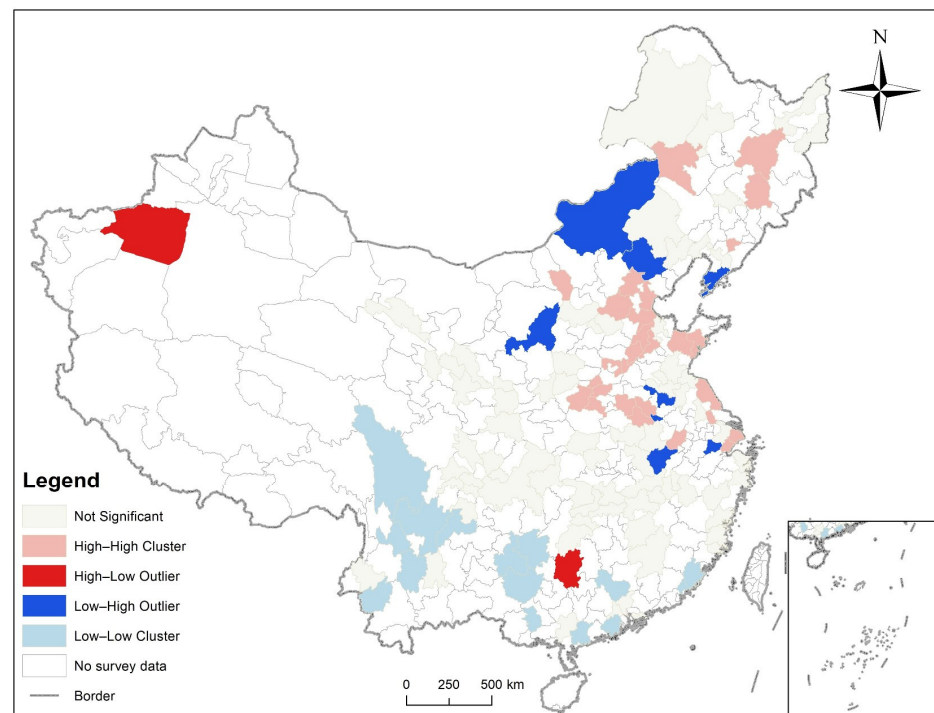
(b)

Figure 7. Spatial autocorrelation analysis of diabetes prevalence in 2011. (a) The result of Getis and Ord's G^* ; (b) The result of local spatial autocorrelation analysis.



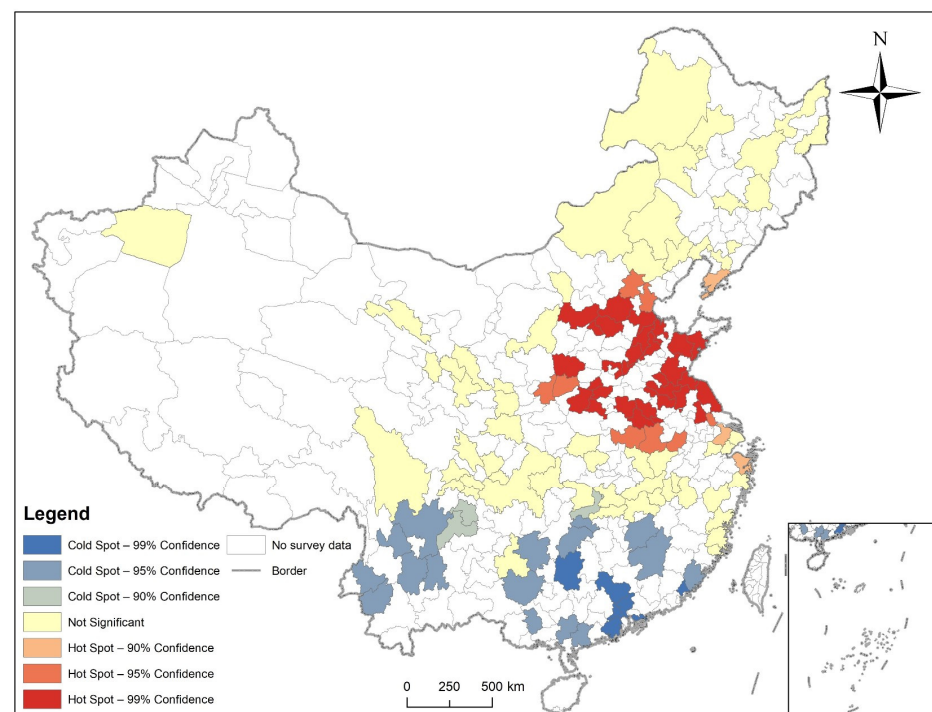
(a)

Figure 8. Cont.



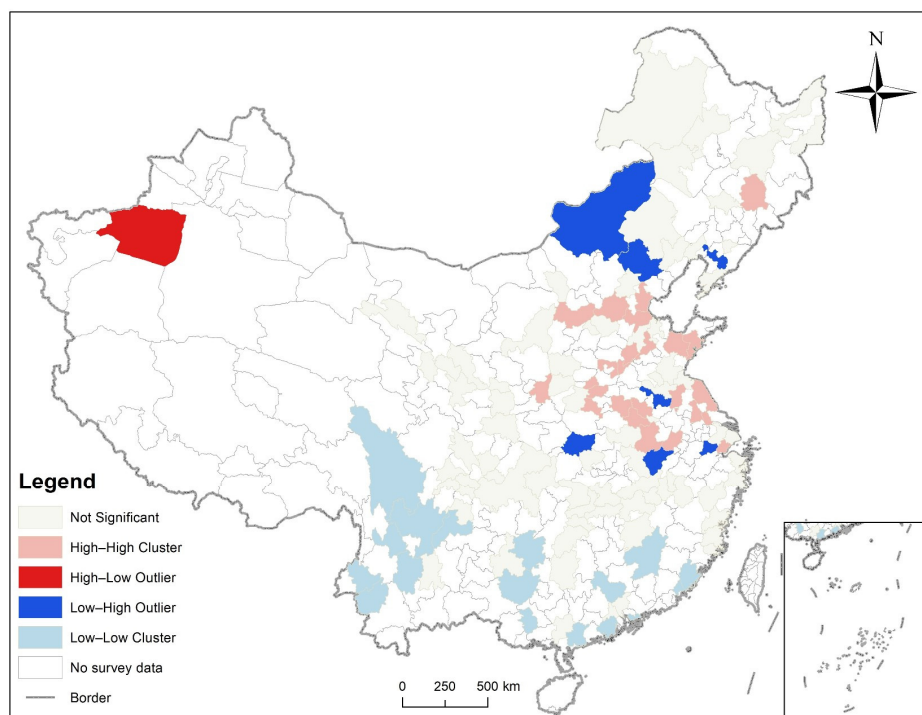
(b)

Figure 8. Spatial autocorrelation analysis of diabetes prevalence in 2013. (a) The result of Getis and Ord's G^* ; (b) The result of local spatial autocorrelation analysis.



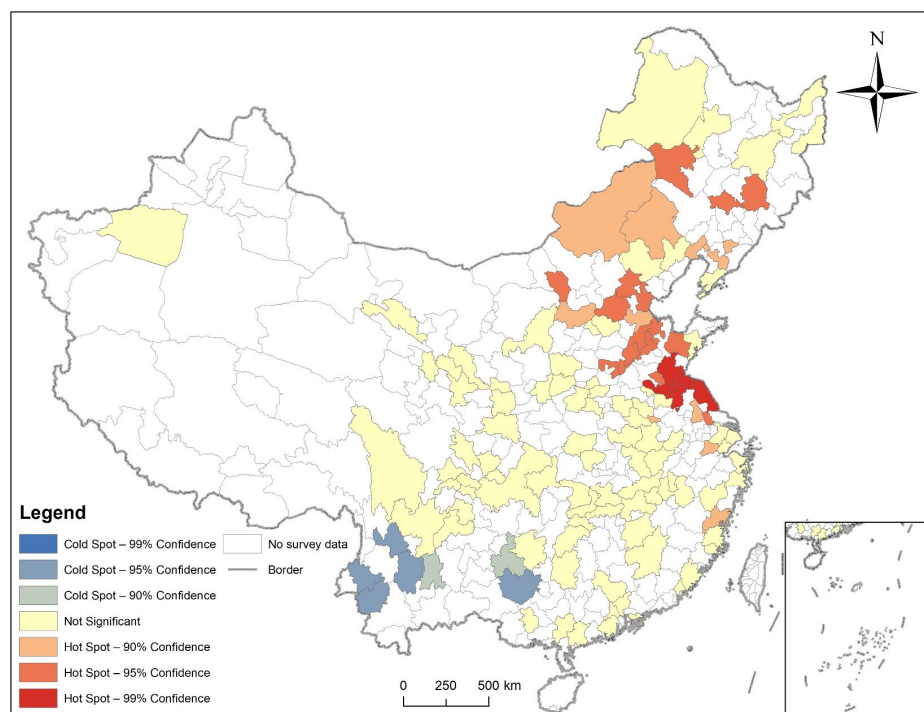
(a)

Figure 9. Cont.



(b)

Figure 9. Spatial autocorrelation analysis of diabetes prevalence in 2015. (a) The result of Getis and Ord's G^* ; (b) The result of local spatial autocorrelation analysis.



(a)

Figure 10. Cont.

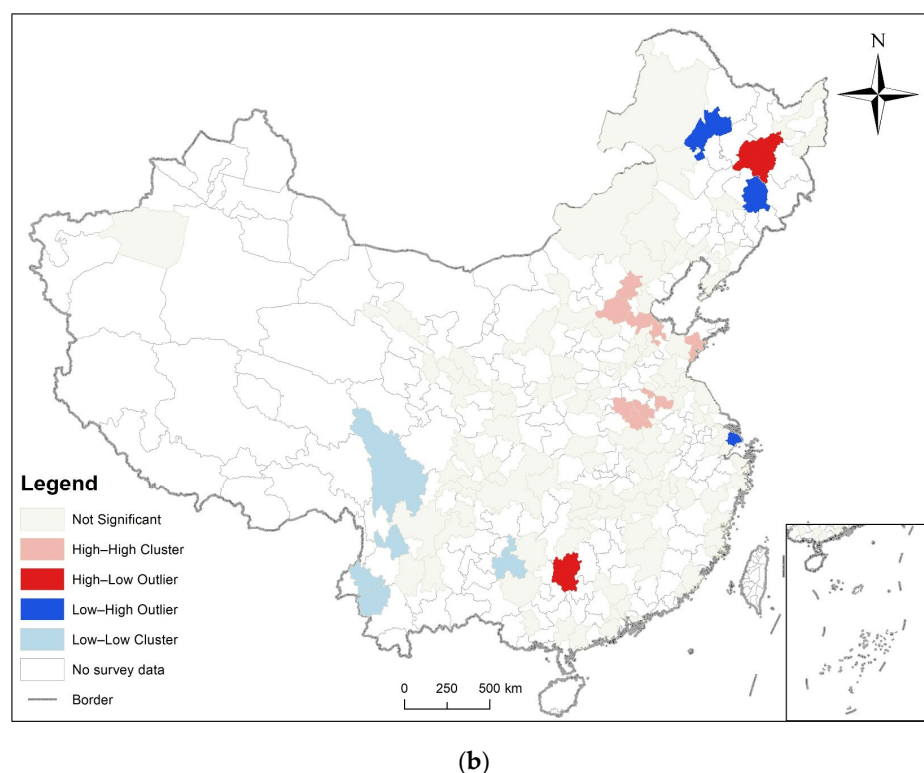


Figure 10. Spatial autocorrelation analysis of diabetes prevalence in 2018. (a) The result of Getis and Ord's G^* ; (b) The result of local spatial autocorrelation analysis.

Table 1. Global spatial autocorrelation.

Date	Moran's Index	p -Value	Z-Score	Spatial Distribution Model
2011	0.103458	<0.001	7.139808	Cluster
2013	0.104485	<0.001	7.205062	Cluster
2015	0.067174	<0.001	4.835403	Cluster
2018	0.025585	<0.007	2.652944	Cluster

3.3. Analysis of Time and Space

Using SaTScan software to conduct a purely spatial analysis of the respondents in 2018 to accurately locate the spatial clustering area of diabetes, a Poisson distribution was used, and we set a maximum of 10% of the population in the at risk group. The results showed that the most likely clustering center appears in Cangzhou, Hebei Province. There were ten cities are in the dangerous areas (Cangzhou, Tianjin, Dezhou, Baoding, Binzhou, Beijing, Jinan, Shijiazhuang, Liaocheng, Weifang) (Table 2 and Figure 11), and 1899 respondents at risk.

Table 2. Purely spatial analysis results by using SaTScan software.

Cluster Center	Radius (km)	Region	Logarithmic Likelihood Ratio	Relative Risk Level	p -Value
Cangzhou, Hebei Province	270.98	Cangzhou, Tianjin, Dezhou, Baoding, Binzhou, Beijing, Jinan, Shijiazhuang, Liaocheng, Weifang	52.819422	1.54	<0.001
Tianjin	153.02	Tianjin, Cangzhou, Beijing, Baoding	41.161335	1.78	<0.001
Zhengzhou, Henan Province	221.64	Zhengzhou, Jiaozuo, Luoyang, Pingdingshan, Zhoukou, Anyang, Puyang, Bozhou	39.852687	1.54	<0.001

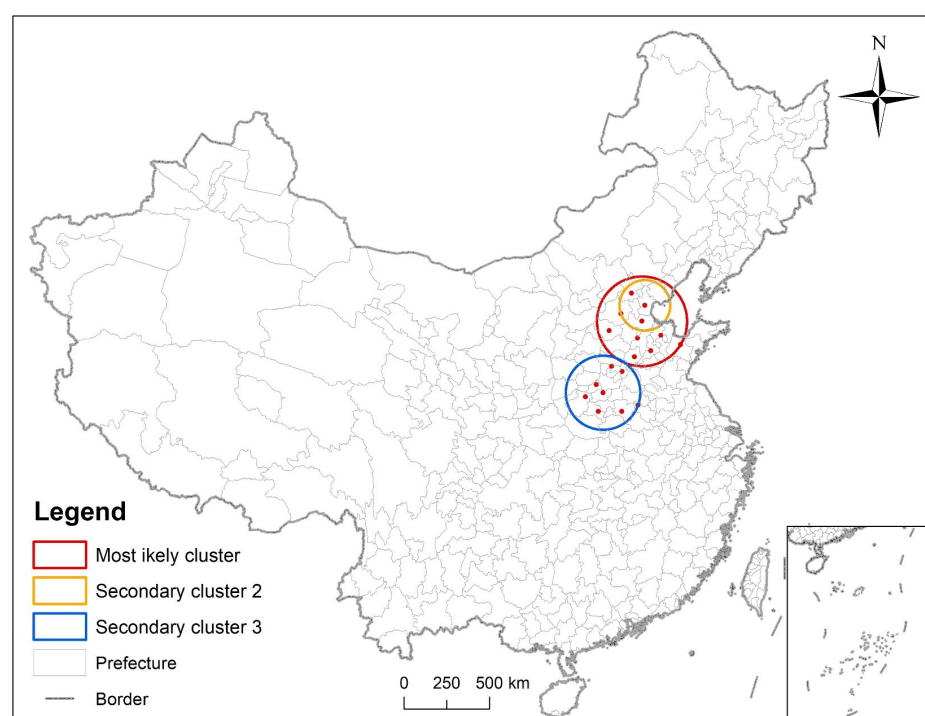


Figure 11. Three clusters were detected by purely spatial analysis.

In order to explore if diabetes had clustering characteristics in space and time, a spatiotemporal analysis of respondents in 2011, 2013, 2015, and 2018 was performed using SaTSca, with a maximum of 10% of the population at risk. The results showed that the most likely agglomeration center appears in Dezhou, Shandong Province. There are ten cities in the danger zone (Dezhou, Cangzhou, Jinan, Liaocheng, Binzhou, Shijiazhuang, Baoding, Tianjin, Puyang, Anyang) (Table 3 and Figure 12), and 1931 respondents at risk.

Table 3. Spatiotemporal analysis results by using SaTScan software.

Cluster Center	Radius (km)	Region	Logarithmic Likelihood Ratio	Relative Risk Level	<i>p</i> -Value
Dezhou, Shandong Province	229.44	Dezhou, Cangzhou, Jinan, Liaocheng, Binzhou, Shijiazhuang, Baoding, Tianjin, Puyang, Anyang	163.632756	4.16	<0.001
Suqian, Jiangsu Province	264.81	Suqian, Xuzhou, Lianyungang, Suzhou, Linyi, Zaozhuang, Yancheng, Huainan, Yangzhou, Taizhou, Bozhou, Fuyang, Hefei	109.037860	3.39	<0.001
Weinan, Shanxi Province	377.23	Weinan, Yuncheng, Baoji, Linfen, Luoyang, Hanzhong, Pingliang, Pingdingshan, Jiaozuo, Xiangfan, Zhengzhou	94.209061	3.20	<0.001

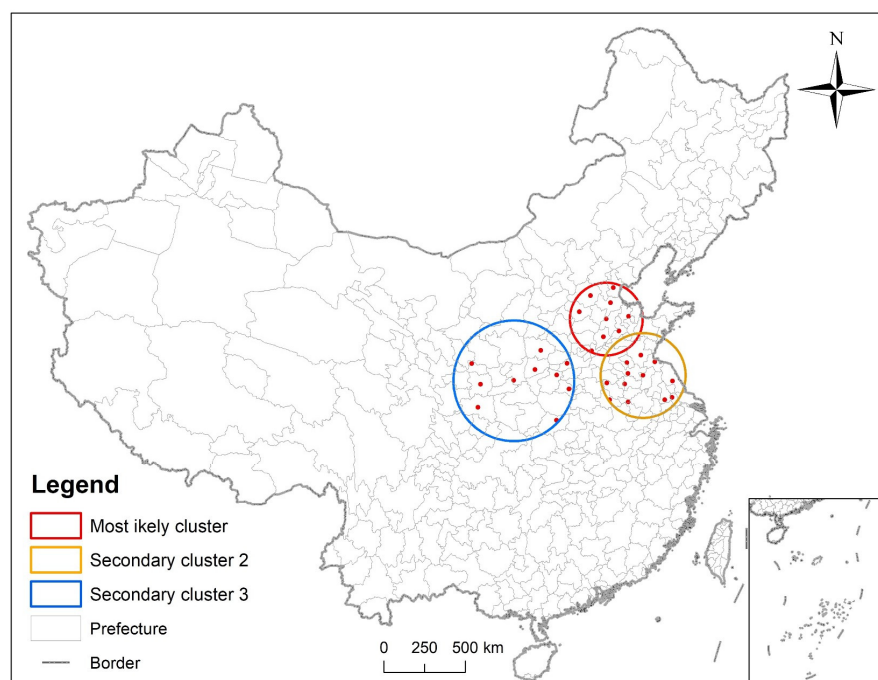


Figure 12. Three clusters were detected by spatiotemporal analysis.

3.4. Binary Logistic Regression

In order to explore the factors that affect the occurrence of diabetes and assess the risk of diabetes, binary logistic regression was used for exploration based on the baseline data of 2018. The initial assignment of variables is shown in Table 4.

Table 4. Variables and assignments.

Variables	Type	Assignments
Gender	Integer	0 = Male; 1 = Female
Age	Integer	0 = 45–49; 1 = 50–54; 2 = 55–59; 3 = 60–64; 4 = 65–69; 5 = 70 or more
Location of Residential Address	Integer	0 = Central of City/Town; 1 = Urban-Rural Integration Zone; 2 = Rural; 3 = Special Zone
Education	Integer	0 = Illiterate; 1 = Did not Finish Primary School; 2 = Sishu/Home School; 3 = Elementary School; 4 = Middle School; 5 = High School; 6 = Vocational School; 7 = Two-/Three-Year College/Associate Degree; 8 = Four-Year College/Bachelor's Degree or more
Marital Status	Integer	0 = Married with Spouse Present; 1 = Married but Not Living with Spouse Temporarily for Reasons Such as Work; 2 = Separated; 3 = Divorced; 4 = Widowed; 5 = Never Married
Nation	Integer	0 = Han Nationality; 1 = Zhuang Nationality; 2 = Manchu; 3 = Hui Nationality; 4 = Miao Nationality; 5 = Uyghur Nationality; 6 = Tujia Nationality; 7 = Yi Nationality; 8 = Other Nationality
Hypertension	Integer	0 = No; 1 = Yes
Dyslipidemia	Integer	0 = No; 1 = Yes
Diabetes	Integer	0 = No; 1 = Yes
Cancer	Integer	0 = No; 1 = Yes
Liver Disease	Integer	0 = No; 1 = Yes
Emotional Problems	Integer	0 = No; 1 = Yes
Smoking History	Integer	0 = No; 1 = Yes
Alcohol Use	Integer	0 = No; 1 = Yes

Table 5 shows the results of the chi-square test for single factors: age, location of residential address, education, hypertension, dyslipidemia, cancer, liver disease, smoking history, and alcohol use. A total of nine factors passed the chi-square test ($p < 0.05$) and could be included in binary logistic regression.

Table 5. Chi-square test result.

Factors		The Total Number of Samples	Number of Cases		X2	p-Value
Gender	Male	8715	463	5.31%	3.734	0.053
	Female	9459	569	6.02%		
Age	45–49	1307	53	4.06%	37.133	<0.001
	50–54	3173	120	3.78%		
	55–59	3135	181	5.77%		
	60–64	2856	192	6.72%		
	65–69	3063	207	6.76%		
	≥70	4640	279	6.01%		
Location of Residential Address	Central of City/Town	3486	232	6.66%	13.003	0.005
	Urban-Rural Integration Zone	1270	90	7.09%		
	Rural	13,346	706	5.29%		
	Special Zone	72	4	5.56%		
Education	Illiterate	4022	252	6.27%	17.018	0.03
	Did not Finish Primary School	3764	208	5.53%		
	Sishu/Home School	41	2	4.88%		
	Elementary School	4030	221	5.48%		
	Middle School	4023	201	5.00%		
	High School	1503	81	5.39%		
	Vocational School	420	35	8.33%		
	Two-/Three-Year College/Associate Degree	229	22	9.61%		
	Four-Year College/Bachelor's Degree or more	142	10	7.04%		
Marital Status	Married with Spouse Present	14,281	820	5.74%	3.355	0.645
Nation	Married But Not Living with Spouse Temporarily for Reasons Such as Work	1214	63	5.19%	10.489	0.232
	Separated	65	4	6.15%		
	Divorced	226	9	3.98%		
	Widowed	2280	133	5.83%		
	Never Married	108	3	2.78%		
	Han Nationality	17,077	975	5.71%		
	Zhuang Nationality	177	8	4.52%		
	Manchu	301	12	3.99%		
	Hui Nationality	107	12	11.21%		
	Miao Nationality	112	3	2.68%		
	Uyghur Nationality	81	6	7.41%		
	Tujia Nationality	25	1	4.00%		
	Yi Nationality	97	3	3.09%		
	Other Nationality	197	12	6.09%		

Table 5. Cont.

Factors		The Total Number of Samples	Number of Cases		X2	p-Value
Hypertension	No	16,273	792	4.87%	161.428	<0.001
	Yes	1901	240	12.62%		
Dyslipidemia	No	16,601	739	4.45%	433.646	<0.001
	Yes	1573	293	18.63%		
Cancer	No	17,946	1008	5.62%	8.651	0.003
	Yes	228	24	10.53%		
Liver Disease	No	17,603	979	5.56%	12.350	<0.001
	Yes	571	53	9.28%		
Emotional Problems	No	17,968	1015	5.65%	2.246	0.134
	Yes	206	17	8.25%		
Smoking History	No	17,359	955	5.50%	19.540	<0.001
	Yes	815	77	9.45%		
Alcohol Use	No	11,936	731	6.12%	11.566	0.001
	Yes	6238	301	4.83%		

Binary logistic regression took diabetes as the dependent variable, age, location of residential address, education, hypertension, dyslipidemia, cancer, liver disease, kidney disease, smoking history, and alcohol use as independent variables. The Hosmer–Lemeshow test of the model was greater than 0.05 (0.889), indicating that the model had fully utilized the data and there was no very significant difference between the predicted value and the true value. Meanwhile, the result of the Omnibus test indicated that the model was statistically significant ($p < 0.05$). The established binary logistic regression can be expressed as Equation (5), according to Table 6.

$$\ln \frac{p}{1-p} = -3.549 - 0.062 * Age(50 - 54) + 0.348 * Age(55 - 59) + 0.488 * Age(60 - 64) + 0.475 * Age(65 - 69) + 0.389 * Age(\geq 70) + 0.703 * Hypertension + 1.302 * Dyslipidemia + 0.373 * Kidney Disease \quad (5)$$

Table 6. Binary logistic regression analysis.

Variables	B	SE	Wald	Df	p-Value	OR	95% CI	
							Lower	Upper
Age (45–49)			31.808	5	0.000			
50–54	−0.062	0.170	0.134	1	0.714	0.939	0.673	1.311
55–59	0.348	0.162	4.629	1	0.031	1.416	1.031	1.944
60–64	0.488	0.161	9.193	1	0.002	1.629	1.188	2.232
65–69	0.475	0.160	8.843	1	0.003	1.607	1.176	2.198
≥70	0.389	0.155	6.291	1	0.012	1.475	1.089	2.000
Hypertension	0.703	0.081	75.339	1	0.000	2.020	1.723	2.367
Dyslipidemia	1.302	0.076	295.059	1	0.000	3.676	3.169	4.265
Smoking history	0.373	0.128	8.446	1	0.004	1.452	1.129	1.867
Constant	−3.549	0.144	606.838	1	0.000	0.029		

The results showed that the occurrence of diabetes was significantly correlated with age, hypertension, dyslipidemia, kidney disease, and smoking history. The risk was higher in the 60–64 age group than in other age groups ($OR = 1.635$, $p < 0.001$). Patients with hypertension had a significantly higher risk of diabetes than those with other chronic diseases ($OR = 2.004$, $p < 0.001$). The highest risk was associated with dyslipidemia ($OR = 3.598$, $p < 0.001$).

3.5. Geographically Weighted Regression

Figure 13 showed the local R^2 by using GWR ($AIC_c = 640.402523$, $R^2 = 0.621877$, Adjusted $R^2 = 0.609018$). The distribution of residuals of GWR in space was randomized using Global Spatial Autocorrelation ($p = 0.233661$, spatial distribution model was random). Table 7 shows the statistics of local coefficient variables, illustrating that none of the factors exhibited significant spatial heterogeneity.

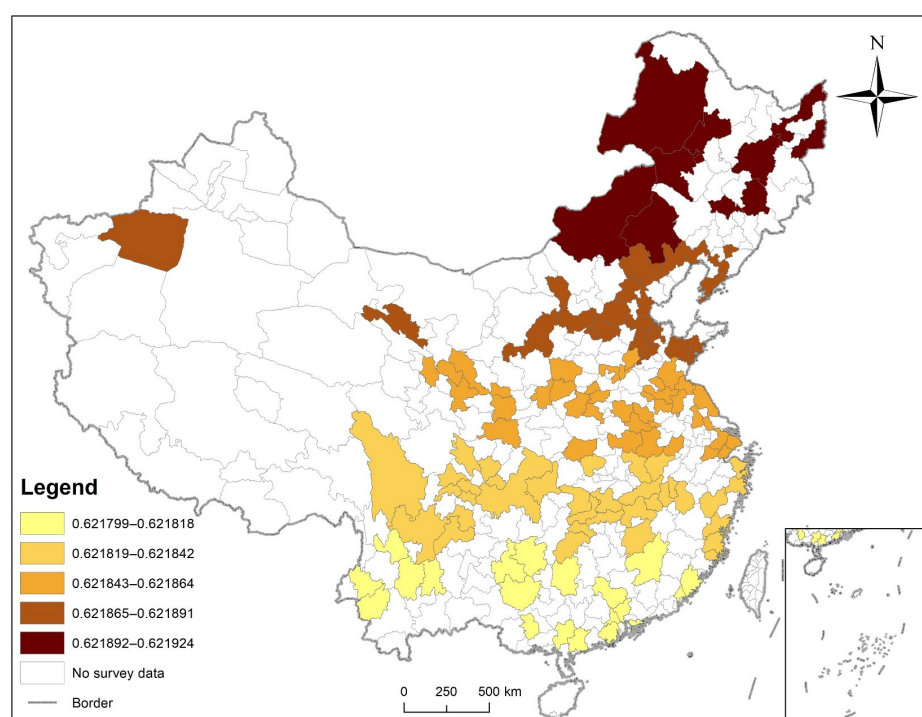


Figure 13. The local R^2 of GWR.

Table 7. Coefficient of risk factors.

Variables	Mean	Max	Min
Age	0.0571	0.057091	0.05713
Hypertension	0.007537	0.007361	0.00765
Dyslipidemia	0.265775	0.26573	0.265843
Smoking History	0.00879	0.008752	0.008823

3.6. Disease Risk Assessment

Through binary logistic regression, we chose age, hypertension, dyslipidemia, cancer, heart attack, stroke, kidney disease, smoking history, and alcohol use as independent variables. We chose diabetes as the dependent variable to establish the binary logistic model and random forest model. AUC (area under the ROC curve) was used to evaluate the assessment model in this study. To verify whether the model's expected risk result is consistent with the actual prevalence of diabetes, ArcGIS 10.4 was used to visualize the actual diabetes prevalence map and the diabetes risk assessment map (Figure 14), the high-risk assessment areas are mainly located in the Beijing–Tianjin–Hebei region and the northeast region. The random forest model's assessment results are consistent with the actual prevalence, while the binary logistic regression model's assessment results are far from the real incidence rate. Meanwhile, according to the ROC curve (Figures 15 and 16), the accuracy of the random forest model (AUC = 0.7745) was higher than the binary logistic model (AUC = 0.6677). However, the random forest model cannot explain the function direction of independent variables and the relative risk degree of influencing factors, but binary logistic regression analysis can define the model and variables well.

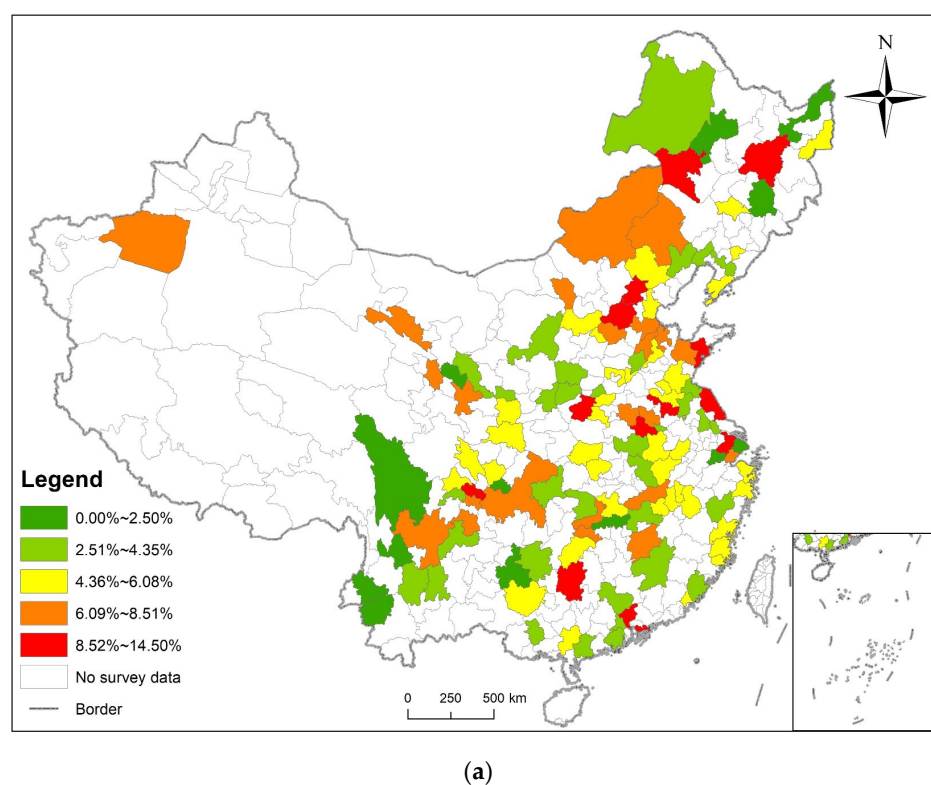
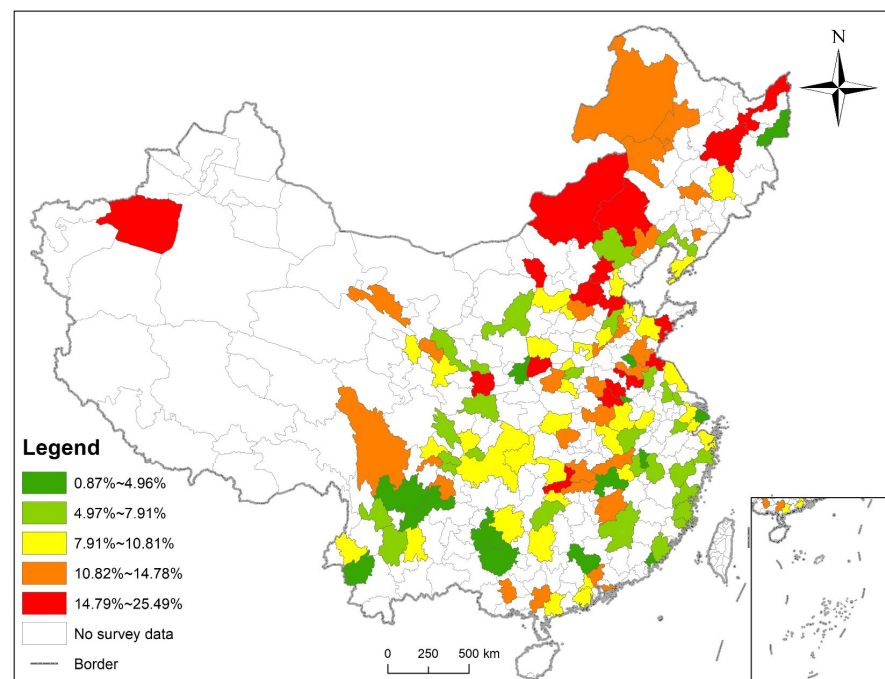
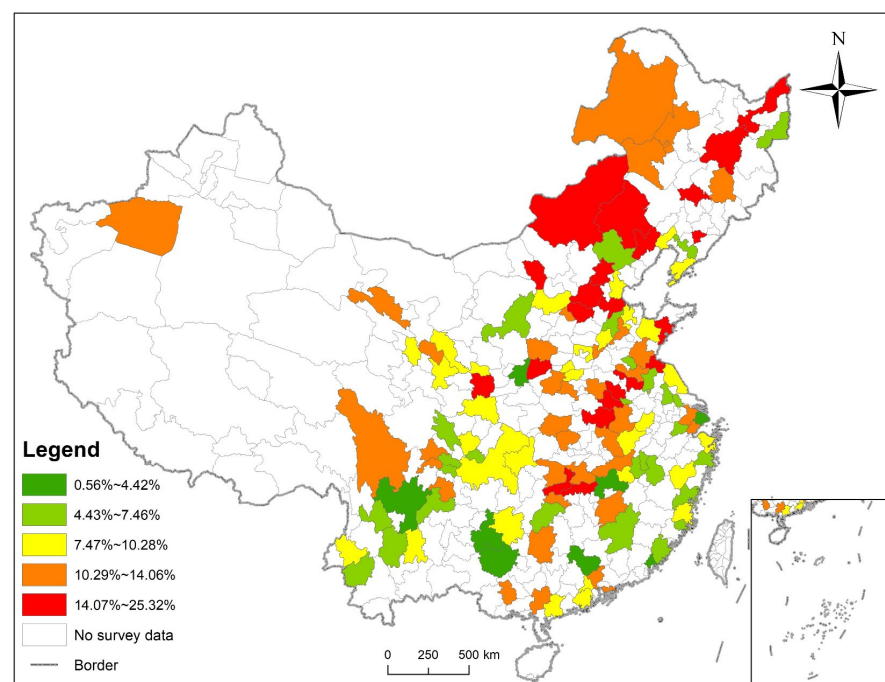


Figure 14. Cont.



(b)



(c)

Figure 14. Assessment result. (a) The prevalence of diabetes in 2018 was divided into five classifications according to the natural breaks method. (b) The disease risk assessment result of binary logistic regression model was divided into five classifications according to the natural breaks method; (c) The disease risk assessment result of random forest model was divided into five classifications according to the natural breaks method.

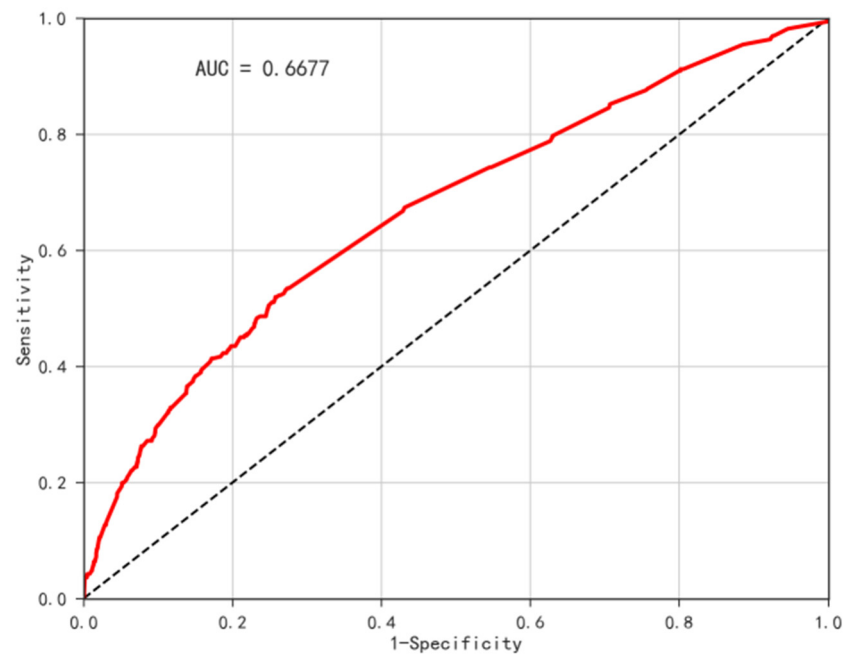


Figure 15. ROC Curve of binary logistic regression model.

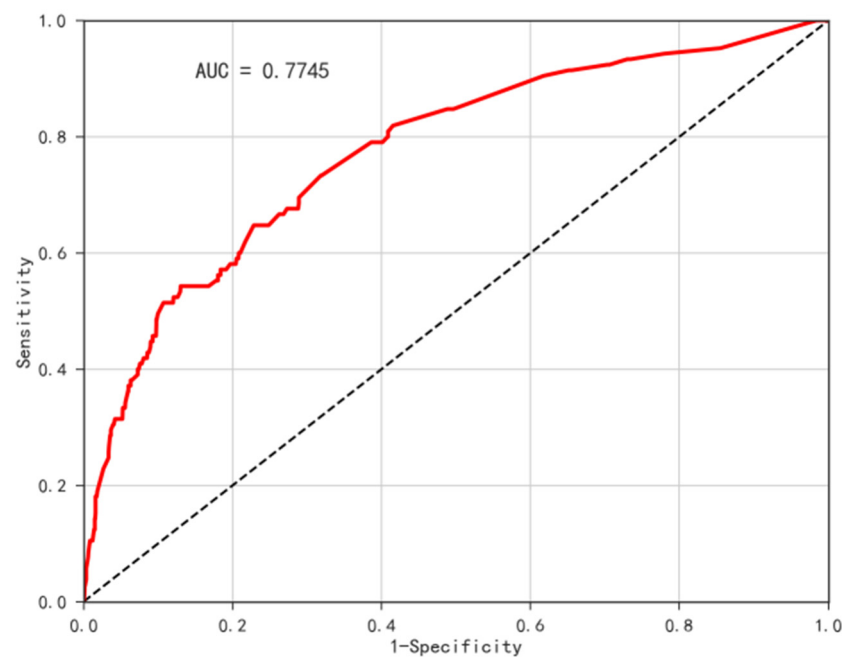


Figure 16. ROC Curve of random forest model.

4. Discussion

4.1. Innovation in This Study

Because the traditional data analysis method does not easily avoid interactions between the independent variables, as an emerging machine learning algorithm, the random forest algorithm performs well in avoiding multicollinearity. Therefore, it is widely used in the assessment of disease risk. The use of a random forest model to establish a concise and accurate diabetes risk assessment model is an innovative way to assess the risk of diabetes among people over 45 years old in China. Because the dataset does not always contain complete information, the distribution between positive and negative classes is mostly imbalanced, and some parameters are of low importance for the decision class, the random forest model performed better in this situation. We used the random forest

model to make our diabetes risk assessment map, compared it with the assessment results of logistic regression, and noted that the assessment result was consistent with the actual prevalence. Thus, we conclude that the random forest model can achieve greater accuracy in assessing diabetes risk [41]. However, binary logistic regression analysis can intuitively explain diabetes risk factors, which is a disadvantage of the random forest model. The advantages of the two models should be combined in practical applications to allow them to jointly play a valuable role in disease risk assessment.

4.2. Scale Effect

The selection of different observation and analysis scales will result in the detection of different phenomena. This is known as the scale effect [42]. We took this into consideration when conducting our research. Our preliminary experiments showed that the spatial patterns obtained from the study at the prefecture-level city scale and the provincial scale are basically the same. Therefore, in order to get more detailed spatial patterns, our spatiotemporal analysis was based on the city-level prefecture scale.

4.3. Spatiotemporal Characteristic of Diabetes Prevalence

Diabetes prevalence remains high in China. According to the report from the International Diabetes Federation, diabetes prevalence in China had increased from 8.8% in 2011 to 10.9% in 2018 in adults 20–79 years. The prevalence of diabetes among people over 45 years old increased from 0.00% to 14.04% in 2011 to 0.00% to 14.50% in 2018 in the study area where the sample is located.

A significant Moran's I test indicates that there is a presence of spatial autocorrelation, Getis and Ord's G^* could identify the hot or cold spot areas. Identifying hot spots for diseases is important for public health authorities who should adopt them for better-targeted interventions [43]. To determine the spatial patterns of a disease, local indicators of spatial association (LISA) in the environmental GIS are very helpful. This model is a set of methods used to describe and visualize spatial distributions, identify atypical locations or spatial outliers, determine patterns of spatial association, clusters, or hot-spots, and propose spatial regimes or other shapes of spatial heterogeneity [44].

In 2011, 2013, 2015, and 2018, the Moran's I coefficient of diabetes prevalence in China was between 0.025585 and 0.104485, and showed non-random spatial distribution. Getis and Ord's G^* showed that hot spots are mostly found in the eastern and central regions, while cold spots are more common in southern regions. Local Spatial Autocorrelation analysis found that the High-High distribution pattern of diabetes is mainly found in cities close to the Beijing–Tianjin–Hebei region.

We also found that the spatial distribution model of diabetes was clustered, but that the tendency to cluster is waning, as the Moran's I decreased from 0.103458 in 2011 to 0.025585 in 2018, and the hot and cold spot areas were also conspicuously decreased. Many areas also showed not significant High-High or Low-Low distributions.

The spatial scan statistic is a useful and widely used tool for detecting spatial or space-time clusters in disease surveillance. The software SaTScan, available for free, enhances this method's ease-of-access for researchers [45]. We used SaTScan to accurately locate the spatial clustering areas of diabetes and to explore if diabetes had clustering characteristics in space and time.

Spatiotemporal clustering areas were detected by SaTScan software and they were located near the Beijing–Tianjin–Hebei region.

Therefore, diabetes prevalence has obvious spatial distribution characteristics in the population over 45 years old in China, that is, the north is higher than the south, the coast is higher than the inland, and economically developed areas are higher than economically underdeveloped areas. The specific reasons for the patterns need further research, but should be related to differences in eating habits and lifestyle changes caused by economic development, and by glycemic control, which varied greatly across geographic regions [46,47].

4.4. Diabetes Risk Factors

Binary logistic regression is often used to explore diabetes risk factors [48,49]. Binary logistic regression analysis showed that age, hypertension, dyslipidemia, and smoking history were all diabetes risk factors in this study.

In China, diabetes poses a severe threat to the population. Age is a main factor for diabetes [50]. In this study, especially after the age of 55, diabetes risk increased significantly with age. Therefore, middle-aged and elderly residents in China should always pay attention to their health, so as not to miss the best treatment time.

Besides, compared with other chronic diseases, hypertension and dyslipidemia are more likely to lead to diabetes, and diabetes also likely leads to the occurrence of hypertension or dyslipidemia [51–53]. As the main component of metabolic syndrome, diabetes, hyperglycemia, and hyperlipidemia interconnect and influence each other, forming a complex framework of chronic diseases [54]. With the prolongation of the disease's course, the patient's body's immune function becomes increasingly abnormal, the function of many systems is weakened, and multiple diseases are prone to occur. With the prolongation of the disease's course, the function of many systems in the patient's body is weakened, which always leads to multiple diseases [55–58].

More and more studies show that smoking significantly increases the risk of diabetes [59]. Thus, diabetes patients with a history of smoking are reported to be at especially increased risk of incidence and poor outcomes from severe acute respiratory syndrome coronavirus [60]. China is one of the countries with the largest number of tobacco consumers in the world [61,62], which may be one of the reasons for the high prevalence of diabetes, and even of other chronic diseases, in China.

4.5. Spatial Heterogeneity of Diabetes Risk Factors

A GWR model is a simple and effective technology used to deal with spatial heterogeneity. Unlike traditional multiple linear regression, GWR lets regression parameters vary across space [63]. A GWR model was used to explore the spatial heterogeneity of diabetes risk factors. However, the results showed that there is no obvious spatial heterogeneity in the four risk factors (age, hypertension, dyslipidemia, and smoking history). This might be because this study did not incorporate socioeconomic and environmental factors into the study [64,65].

4.6. Limitations and Future Research

There are still some deficiencies in this research. For example, environmental factors, which are closely related to the prevalence of diabetes, have not been considered in this study. Besides, our approach to spatiotemporal analysis in this study was still traditional, and factors included in the model were not enough. In addition, there is still room for improvement in the accuracy of the model, and we are also trying to add other classification algorithms to our research. We will continue to advance this research, and it is believed that our research will provide accurate data support for improving the living conditions of people over 45 years old in China.

5. Conclusions

Firstly, in this paper, spatial autocorrelation and spatiotemporal clustering analysis were used to analyze the spatial distribution characteristics of diabetes. Secondly, we used the binary logistic regression model to explore the risk factors of diabetes in detail. Finally, the logistic regression model and random forest model were used to assess the risk of diabetes in people over 45 years old in China. The results showed that the clustering areas of patients with diabetes were mainly in the Beijing–Tianjin–Hebei region. The tendency to find clusters of diabetes prevalence among people over 45 years old in China is waning. Age, hypertension, dyslipidemia, and smoking history all had effects on diabetes, but the spatial heterogeneity of these factors were weak. Compared with the binary logistic model, the random forest model showed better fitness in assessing diabetes risk, and showed

that the high-risk regions are the northeast region and the Beijing–Tianjin–Hebei region. Therefore, our method can analyze the spatial distribution characteristics and influencing factors of diabetes, but there is still room for improvement in the accuracy of assessing the risk of diabetes. We will continue to follow up on this study after the data of CHARLS is updated, and we will explore more excellent methods in the following research.

Author Contributions: Conceptualization, Z.W., W.D. and K.Y.; methodology, Z.W., W.D. and K.Y.; software, Z.W. and W.D.; validation, W.D.; formal analysis, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, Z.W. and W.D.; visualization, Z.W., W.D. and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Natural Science Foundation of China (Grants Nos. 42161071, 42071381, 41661087).

Institutional Review Board Statement: We are using a secondary dataset. It has been procured from a government agency, and they have followed all the ethical protocols in collecting data.

Informed Consent Statement: Not applicable.

Data Availability Statement: “China Health and Retirement Longitudinal Study” at <http://charls.pku.edu.cn/> (accessed on 28 May 2022).

Acknowledgments: Thanks to CHARLS for providing data support for this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Su, R.; Cai, L.; Cui, W.; He, J.; You, D.; Golden, A. Multilevel Analysis of Socioeconomic Determinants on Diabetes Prevalence, Awareness, Treatment and Self-Management in Ethnic Minorities of Yunnan Province, China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 751. [\[CrossRef\]](#)
2. Su, B.; Wang, Y.; Dong, Y.; Hu, G.; Xu, Y.; Peng, X.; Wang, Q.; Zheng, X. Trends in Diabetes Mortality in Urban and Rural China, 1987–2019: A Joinpoint Regression Analysis. *Front. Endocrinol.* **2022**, *12*, 777654. [\[CrossRef\]](#)
3. Hamat, A.; Jaludin, A.; Mohd-Dom, T.N.; Rani, H.; Jamil, N.A.; Abdul Aziz, A.F. Diabetes in the News: Readability Analysis of Malaysian Diabetes Corpus. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6802. [\[CrossRef\]](#)
4. Yuan, Q.; Wang, H.; Gao, P.; Chen, W.; Lv, M.; Bai, S.; Wu, J. Prevalence and Risk Factors of Metabolic-Associated Fatty Liver Disease among 73,566 Individuals in Beijing, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2096. [\[CrossRef\]](#)
5. Ali, A.; Alfajjam, S.; Gasana, J. Diabetes Mellitus and Its Risk Factors among Migrant Workers in Kuwait. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3943. [\[CrossRef\]](#)
6. Chung; Kim; Kwock. Dietary Patterns May Be Nonproportional Hazards for the Incidence of Type 2 Diabetes: Evidence from Korean Adult Females. *Nutrients* **2019**, *11*, 2522. [\[CrossRef\]](#)
7. El-Shareif, H. Prevalence, pattern, and attitudes of smoking among libyan diabetic males: A clinic-based study. *Ibnosina J. Med. Biomed. Sci.* **2022**, *11*, 171–175. [\[CrossRef\]](#)
8. Rabieenia, E.; Jalali, R.; Mohammadi, M. Prevalence of nephropathy in patients with type 2 diabetes in Iran: A systematic review and meta-analysis based on geographic information system (GIS). *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 1543–1550. [\[CrossRef\]](#)
9. Wang, Y.; Liang, X.; Zhou, Z.; Hou, Z.; Yang, J.; Gao, Y.; Yang, C.; Chen, T.; Li, C. Prevalence and Numbers of Diabetes Patients with Elevated BMI in China: Evidence from a Nationally Representative Cross-Sectional Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2989. [\[CrossRef\]](#)
10. International Diabetes Federation. Available online: <https://idf.org/> (accessed on 28 May 2022).
11. Murad, A.; Khashoggi, B.F. Using GIS for Disease Mapping and Clustering in Jeddah, Saudi Arabia. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 328. [\[CrossRef\]](#)
12. Masimalai, P. Remote sensing and Geographic Information Systems (GIS) as the applied public health and environmental epidemiology. *Int. J. Med. Sci. Public Health* **2014**, *3*, 1430. [\[CrossRef\]](#)
13. Ricketts, T.C. Geographic Information Systems and Public Health. *Annu. Rev. Public Health* **2003**, *24*, 1–6. [\[CrossRef\]](#)
14. Dudley, T.; Creppage, K.; Shanahan, M.; Proescholdbell, S. Using GIS to Evaluate a Fire Safety Program in North Carolina. *J. Community Health* **2013**, *38*, 951–957. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Dong, W.; Yang, K.; Xu, Q.; Liu, L.; Chen, J. Spatio-temporal pattern analysis for evaluation of the spread of human infections with avian influenza A(H7N9) virus in China, 2013–2014. *BMC Infect. Dis.* **2017**, *17*, 704. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Miranda, M.L.; Casper, M.; Tootoo, J.; Schieb, L. Putting Chronic Disease on the Map: Building GIS Capacity in State and Local Health Departments. *Prev. Chronic Dis.* **2013**, *10*, E100. [\[CrossRef\]](#)
17. Vine, M.F.; Degnan, D.; Hanchette, C. Geographic information systems: Their use in environmental epidemiologic research. *Environ. Health Perspect.* **1997**, *105*, 598–605. [\[CrossRef\]](#)

18. Xu, S.; Ming, J.; Xing, Y.; Gao, B.; Yang, C.; Ji, Q.; Chen, G. Regional differences in diabetes prevalence and awareness between coastal and interior provinces in China: A population-based cross-sectional study. *BMC Public Health* **2013**, *13*, 299. [\[CrossRef\]](#)
19. Li, J.; Wang, S.; Han, X.; Zhang, G.; Zhao, M.; Ma, L. Spatiotemporal trends and influence factors of global diabetes prevalence in recent years. *Soc. Sci. Med.* **2020**, *256*, 113062. [\[CrossRef\]](#)
20. Cao, G.; Cui, Z.; Ma, Q.; Wang, C.; Xu, Y.; Sun, H.; Ma, Y. Changes in health inequalities for patients with diabetes among middle-aged and elderly in China from 2011 to 2015. *BMC Health Serv. Res.* **2020**, *20*, 719. [\[CrossRef\]](#)
21. Zhang, X.; Chen, X.; Gong, W. Type 2 diabetes mellitus and neighborhood deprivation index: A spatial analysis in Zhejiang, China. *J. Diabetes Investig.* **2019**, *10*, 272–282. [\[CrossRef\]](#)
22. Alcalá-Rmz, V.; Galván-Tejada, C.E.; García-Hernández, A.; Valladares-Salgado, A.; Cruz, M.; Galván-Tejada, J.I.; Celaya-Padilla, J.M.; Luna-Garcia, H.; Gamboa-Rosales, H. Identification of People with Diabetes Treatment through Lipids Profile Using Machine Learning Algorithms. *Healthcare* **2021**, *9*, 422. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Samet, S.; Laouar, M.R.; Bendib, I.; Eom, S. Analysis and Prediction of Diabetes Disease Using Machine Learning Methods. *Int. J. Decis. Support Syst. Technol.* **2022**, *14*, 1–19. [\[CrossRef\]](#)
24. Zhou, M.; Astell-Burt, T.; Bi, Y.; Feng, X.; Jiang, Y.; Li, Y.; Page, A.; Wang, L.; Xu, Y.; Wang, L.; et al. Geographical variation in diabetes prevalence and detection in china: Multilevel spatial analysis of 98,058 adults. *Diabetes Care* **2015**, *38*, 72–81. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Li, L.; Ding, H.; Li, Z. Does Internet Use Impact the Health Status of Middle-Aged and Older Populations? Evidence from China Health and Retirement Longitudinal Study (CHARLS). *Int. J. Environ. Res. Public Health* **2022**, *19*, 3619. [\[CrossRef\]](#)
26. Yu, J.; Yi, Q.; Chen, G.; Hou, L.; Liu, Q.; Xu, Y.; Qiu, Y.; Song, P. The visceral adiposity index and risk of type 2 diabetes mellitus in China: A national cohort analysis. *Diabetes Metab. Res. Rev.* **2022**, *38*, e3507. [\[CrossRef\]](#)
27. He, B.; Li, Z.; Xu, L.; Liu, L.; Wang, S.; Zhan, S.; Song, Y. Upper arm length and knee height are associated with diabetes in the middle-aged and elderly: Evidence from the China Health and Retirement Longitudinal Study. *Public Health Nutr.* **2022**, 1–9. [\[CrossRef\]](#)
28. Liu, X.; Fang, W.; Li, H.; Han, X.; Xiao, H. Is Urbanization Good for the Health of Middle-Aged and Elderly People in China?—Based on CHARLS Data. *Sustainability* **2021**, *13*, 4996. [\[CrossRef\]](#)
29. China Health and Retirement Longitudinal Study. Available online: <http://charls.pku.edu.cn/en/> (accessed on 26 July 2022).
30. Freitas, W.W.L.; de Souza, R.M.C.R.; Amaral, G.J.A.; De Bastiani, F. Exploratory spatial analysis for interval data: A new autocorrelation index with COVID-19 and rent price applications. *Expert Syst. Appl.* **2022**, *195*, 116561. [\[CrossRef\]](#)
31. Cheruiyot, K. Detecting spatial economic clusters using kernel density and global and local Moran's I analysis in Ekurhuleni metropolitan municipality, South Africa. *Reg. Sci. Policy Pract.* **2022**, *14*, 307–327. [\[CrossRef\]](#)
32. Eccles, K.M.; Bertazzon, S. Applications of geographic information systems in public health: A geospatial approach to analyzing MMR immunization uptake in Alberta. *Can. J. Public Health* **2015**, *106*, e355–e361. [\[CrossRef\]](#)
33. Xue, D.; Yue, L.; Ahmad, F.; Draz, M.U.; Chandio, A.A.; Ahmad, M.; Amin, W. Empirical investigation of urban land use efficiency and influencing factors of the Yellow River basin Chinese cities. *Land Use Policy* **2022**, *117*, 106117. [\[CrossRef\]](#)
34. Ghosh, K.; Dhillon, P.; Agrawal, G. Prevalence and detecting spatial clustering of diabetes at the district level in India. *J. Public Health* **2019**, *28*, 535–545. [\[CrossRef\]](#)
35. Zhang, Y.; Shen, Z.; Ma, C.; Jiang, C.; Feng, C.; Shankar, N.; Yang, P.; Sun, W.; Wang, Q. Cluster of Human Infections with Avian Influenza A (H7N9) Cases: A Temporal and Spatial Analysis. *Int. J. Environ. Res. Public Health* **2015**, *12*, 816–828. [\[CrossRef\]](#)
36. Li, C.; Liu, M.; An, Y.; Tian, Y.; Guan, D.; Wu, H.; Pei, Z. Risk assessment of type 2 diabetes in northern China based on the logistic regression model. *Health Care* **2021**, *29*, 351–358. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Khodakarami, L.; Pourmanafi, S.; Soffianian, A.R.; Lotfi, A. Modeling Spatial Distribution of Carbon Sequestration, CO₂ Absorption, and O₂ Production in an Urban Area: Integrating Ground-Based Data, Remote Sensing Technique, and GWR Model. *Earth Space Sci.* **2022**, *9*, e2022EA002261. [\[CrossRef\]](#)
38. Yang, L.; Yu, K.; Ai, J.; Liu, Y.; Yang, W.; Liu, J. Dominant Factors and Spatial Heterogeneity of Land Surface Temperatures in Urban Areas: A Case Study in Fuzhou, China. *Remote Sens.* **2022**, *14*, 1266. [\[CrossRef\]](#)
39. Boateng, E.Y.; Otoo, J.; Abaye, D.A. Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *J. Data Anal. Inf. Processing* **2020**, *08*, 341–357. [\[CrossRef\]](#)
40. Ren, Z.; Yang, K.; Dong, W. Spatial Analysis and Risk Assessment Model Research of Arthritis Based on Risk Factors: China, 2011, 2013 and 2015. *IEEE Access* **2020**, *8*, 206406–206417. [\[CrossRef\]](#)
41. Daghistani, T.; Alshammari, R. Comparison of Statistical Logistic Regression and RandomForest Machine Learning Techniques in Predicting Diabetes. *J. Adv. Inf. Technol.* **2020**, *11*, 78–83. [\[CrossRef\]](#)
42. DeCesare, N.J.; Hebblewhite, M.; Schmiegelow, F.; Hervieux, D.; McDermid, G.J.; Neufeld, L.; Bradley, M.; Whittington, J.; Smith, K.G.; Morgantini, L.E.; et al. Transcending scale dependence in identifying habitat with resource selection functions. *Ecol. Appl.* **2012**, *22*, 1068–1083. [\[CrossRef\]](#)
43. Sandie, A.B.; Tchatchueng Mbougua, J.B.; Nlend, A.E.N.; Thiam, S.; Nono, B.F.; Fall, N.A.; Senghor, D.B.; Sylla, E.H.M.; Faye, C.M. Hot-spots of HIV infection in Cameroon: A spatial analysis based on Demographic and Health Surveys data. *BMC Infect. Dis.* **2022**, *22*, 334. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Jesri, N.; Saghaipour, A.; Koohpaei, A.; Farzinnia, B.; Jooshin, M.K.; Abolkheirian, S.; Sarvi, M. Mapping and Spatial Pattern Analysis of COVID-19 in Central Iran Using the Local Indicators of Spatial Association (LISA). *BMC Public Health* **2021**, *21*, 2227. [\[CrossRef\]](#) [\[PubMed\]](#)

45. Lee, S.; Moon, J.; Jung, I. Optimizing the maximum reported cluster size in the spatial scan statistic for survival data. *Int. J. Health Geogr.* **2021**, *20*, 33. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Zhou, T.; Liu, X.; Liu, Y.; Li, X. Meta-analytic evaluation for the spatio-temporal patterns of the associations between common risk factors and type 2 diabetes in mainland China. *Medicine* **2019**, *98*, e15581. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A.A.; Ogurtsova, K.; et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **2019**, *157*, 107843. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Oza, A.; Bokhare, A. Diabetes Prediction Using Logistic Regression and K-Nearest Neighbor. In Proceedings of the Congress on Intelligent Systems; Springer: Singapore, 2022; pp. 407–418. [\[CrossRef\]](#)
49. Sergeev, A.V.; Weckman, G.R. Cardiovascular Disease Treatment Outcomes in Patients with Diabetes: Prediction Models Using Artificial Neural Networks and Logistic Regression. *Ann. Epidemiol.* **2015**, *25*, 705. [\[CrossRef\]](#)
50. Nayak, B.S.; Sobrian, A.; Latiff, K.; Pope, D.; Rampersad, A.; Lourenço, K.; Samuel, N. The association of age, gender, ethnicity, family history, obesity and hypertension with type 2 diabetes mellitus in Trinidad. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2014**, *8*, 91–95. [\[CrossRef\]](#)
51. Aswin, M.; Mohan, V. Diabetes and Hypertension: What Is the Connection? In *Hypertension and Cardiovascular Disease in Asia*; Ram, C.V.S., Teo, B.W.J., Wander, G.S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 159–169. [\[CrossRef\]](#)
52. Asante, D.O.; Walker, A.N.; Seidu, T.A.; Kpogo, S.A.; Zou, J. Hypertension and Diabetes in Akatsi South District, Ghana: Modeling and Forecasting. *BioMed Res. Int.* **2022**, 2022, 9690964. [\[CrossRef\]](#)
53. Yuan, Y.; Zhou, X.; Lu, J.; Guo, X.; Ji, L. Lipid control in adult Chinese patients with type 2 diabetes: A retrospective analysis of time trends and geographic regional differences. *Chin. Med. J.* **2022**, *135*, 356–358. [\[CrossRef\]](#)
54. Cheng, F.; Li, Y.; Zheng, H.; Tian, L.; Jia, H. Mediating Effect of Body Mass Index and Dyslipidemia on the Relation of Uric Acid and Type 2 Diabetes: Results from China Health and Retirement Longitudinal Study. *Front. Public Health* **2022**, *9*, 823739. [\[CrossRef\]](#)
55. Tomic, D.; Shaw, J.E.; Magliano, D.J. The burden and risks of emerging complications of diabetes mellitus. *Nat. Rev. Endocrinol.* **2022**, *18*, 525–539. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Wu, Y.; Sun, L.; Zhuang, Z.; Hu, X.; Dong, D. Mitochondrial-Derived Peptides in Diabetes and Its Complications. *Front. Endocrinol.* **2022**, *12*, 808120. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Liang, W.; Chikritzhs, T. Alcohol Consumption during Adolescence and Risk of Diabetes in Young Adulthood. *BioMed Res. Int.* **2014**, 2014, 795741. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Medina-Chávez, J.H.; Vázquez-Parrodi, M.; Santoyo-Gómez, D.L.; Azuela-Antuna, J.; Garnica-Cuellar, J.C.; Herrera-Landero, A.; Baladrán-Duarte, D.A. Integrated Care Protocol: Chronic complications of diabetes mellitus 2. *Rev. Med. Del Inst. Mex. Del Seguro Soc.* **2022**, *60*, S19–S33.
59. Tonstad, S. Cigarette smoking, smoking cessation, and diabetes. *Diabetes Res. Clin. Pract.* **2009**, *85*, 4–13. [\[CrossRef\]](#)
60. Yatsuya, H. Avoid clinical inertia: Importance of asking and advising patients with diabetes who smoke about quitting. *J. Diabetes Investig.* **2021**, *12*, 317–319. [\[CrossRef\]](#)
61. Zhou, Y.H.; Mak, Y.W.; Ho, G.W.K. Effectiveness of Interventions to Reduce Exposure to Parental Secondhand Smoke at Home among Children in China: A Systematic Review. *Int. J. Environ. Res. Public Health* **2019**, *16*, 107. [\[CrossRef\]](#)
62. Xiao, L.; Jiang, Y.; Zhang, J.; Parascandola, M. Secondhand Smoke Exposure among Nonsmokers in China. *Asian Pac. J. Cancer Prev.* **2020**, *21*, 17–22. [\[CrossRef\]](#)
63. Zhao, Y.; Li, Z.; Hu, X.; Yang, G.; Wang, B.; Duan, D.; Fu, Y.; Liang, J.; Zhao, C. Spatial heterogeneity of county-level grain protein content in winter wheat in the Huang-Huai-Hai region of China. *Eur. J. Agron.* **2022**, *134*, 126466. [\[CrossRef\]](#)
64. Murad, A.; Faruque, F.; Naji, A.; Tiwari, A.; Helmi, M.; Dahlan, A. Modelling geographical heterogeneity of diabetes prevalence and socio-economic and built environment determinants in Saudi City—Jeddah. *Geospat Health* **2022**, *17*. [\[CrossRef\]](#)
65. Isfandiari, M.A.; Wahyuni, C.U.; Pranoto, A. Tuberculosis Predictive Index for Type 2 Diabetes Mellitus Patients Based on Biological, Social, Housing Environment, and Psychological Well-Being Factors. *Healthcare* **2022**, *10*, 872. [\[CrossRef\]](#) [\[PubMed\]](#)