IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 17, 2020, accepted November 9, 2020, date of publication November 16, 2020,
date of current version November 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037912

Spatial Analysis and Risk Assessment Model Research of Arthritis Based on Risk Factors: China, 2011, 2013 and 2015

ZHONGDA REN[1,2], KUN YANG[2,3], AND WEN DONG[2,3]
[1]School of Computer Science and Technology, Yunnan Normal University, Kunming 650500, China
[2]Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, Kunming 650500, China
[3]Faculty of Geography, Yunnan Normal University, Kunming 650500, China

Corresponding authors: Kun Yang (kmdcynu@163.com) and Wen Dong (dong_wen121@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 41661087.

ABSTRACT Arthritis is a public health issue that is of global concern. Arthritis is one of the chronic diseases with a high incidence of middle-aged and older adults. The patients have paid a heavy price for this and caused a substantial economic burden on society. In this study, we used spatial autocorrelation, spatial cluster analysis, multiple logistic regression, and random forest models to analyze the spatial distribution and possible risk factors for arthritis in elderly Chinese and assess arthritis risk. Global spatial autocorrelation analysis and significance test results show that Moran's I of arthritis spatial autocorrelation in 2011, 2013, and 2015 are statistically significant, so there is significant spatial autocorrelation three years. The results of local spatial autocorrelation and spatial clustering analysis show that the aggregation areas of arthritis patients are mainly in the southwest, northwest, and central China. Multivariate logistic regression analysis showed that gender, age, education level, Body Mass Index (BMI), Center for Epidemiologic Studies Depression Scale score (CES-D), altitude, region, weather temperature, hypertension, lung, liver, heart, stroke, digestive, and kidney disease were all arthritis affects factors ($P < 0.05$). Compared with the multi-factor Logistic regression model, the random forest model better assesses performance and higher fit. The fitting accuracy is 82.2% in the random forest model, which is better than the multi-factor Logistic regression model (66.6%). According to the assessment risk map generated by the random forest model, Northeast, Southwest, Northwest, South, and Central are high-risk areas for arthritis. These results provide benchmark data for the control and prevention of arthritis diseases.

INDEX TERMS Middle-aged and older adults, arthritis, spatial analysis, risk factors, logistic regression modeling, random forest modelling, risk assessment.

I. INTRODUCTION

Arthritis is a chronic autoimmune disease characterized by joint destruction and dysfunction. Arthritis has many inducing factors and complex pathogenesis, one of the common causes of dysfunction and limited activity in middle-aged and older adults. Most older people develop arthritis in their lives due to physical aging [1]. When patients have arthritis, it will seriously affect their psychological, physical, and daily life, thus reducing their quality of life [2]. Arthritis can make joints and tissues painful and stiff [3]. Besides, arthritis can affect mobility, sleep duration, and social engagement. The quality of life of patients with arthritis is worse than that of healthy people [4]. Arthritis is significantly associated with physical and emotional problems, and adults with arthritis are less likely to be physically active than adults without arthritis [5]. Also, people with arthritis often suffer from anxiety or depression [6]. Previous studies have reported that the incidence of arthritis in people is closely related to sex [7], education [8], and BMI [9]. Treating arthritis can also lead to high medical costs [10]. Because most arthritis is painful to treat due to its long course of disease and lingering, arthritis poses a severe threat to the health of middle-aged and older adults [11]. It poses a massive social and economic burden.

In a study conducted in 2011 and 2012, the overall incidence of arthritis among Chinese adults aged 45 or older

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

was 31.4% [12]. A study conducted in 2013 found that 58.48 percent of arthritis patients in southwest China had functional disabilities [13]. According to the relevant report, from 2013 to 2015, about 54.4 million (22.7%) middle-aged and older adults in the United States suffered from arthritis, 29.3% of them were aged 45-64, and 49.6% were over 65 [14]. The incidence of arthritis increases with age, so middle-aged and older people can be severely affected by arthritis [15]. With the aging process acceleration in China, the incidence rate of arthritis is relatively high, so the incidence of arthritis in middle-aged and older adults in China is worth noting [12]. According to a study, the difference between the southern and northern regions is mainly temperature and humidity. Simultaneously, in cold and humid places, the incidence of arthritis is higher [16]. With the improvement in economic conditions, the heating system in the northern regions has greatly improved and less people work outside than before. The cold weather may not affect people's joints in the northern regions; however, there are more rains in the southern regions than northern regions, and the southern region is more humid, which may be contributable to the higher incidence of arthritis [17].

At present, there are few studies on the spatial distribution characteristics and risk assessment of arthritis in middle-aged and older adults in China, and most of the studies mainly focus on the risk factors of arthritis [18]–[20]. Research the spatial distribution characteristics of arthritis and use it to determine key prevention and control areas to help relevant departments efficiently evaluate and manage arthritis. Exploring the main risk factors of arthritis can provide a scientific basis for establishing and improving regional arthritis prevention and control measures. Therefore, this paper will combine the distribution characteristics and risk factors of arthritis to further research arthritis in middle-aged and older adults in China.

Machine learning has been widely used in many fields, especially in health care [22] . In this study, a logistic regression model based on epidemiological research and a random forest model based on machine learning theory was used to analyze the related factors of arthritis in Chinese middle-aged and older adults and establish a risk assessment model. Compare the pros and cons of the two risk assessment models by sensitivity, specificity, assessment agreement rate, AUC, and Youden index, and select a model with better assessment effect to draw a map of arthritis disease risk assessment, which can be more intuitively observed Areas with a higher risk of arthritis. It is hoped that by analyzing the characteristics of arthritis spatial distribution, influencing factors, and disease risk assessment of the Chinese middle-aged and older adults, it will provide a scientific basis for developing prevention and treatment plans and measures.

II. DATA AND METHODS

A. DATA COLLECTION AND PROCESSING

This study's main subjects are 51332 cases of arthritis in the elderly in 2011, 2013, and 2015 in China. The data related to middle-aged and elderly arthritis comes from the China Health and Retirement Longitudinal Study (CHARLS) [22]. CHARLS is a major project funded by the National Natural Science Foundation of China. The CHARLS project conducted survey visits in 150 counties and 450 communities (villages) in 28 provinces (autonomous regions and municipalities) in China in 2011, 2013, and 2015. The target group is the middle-aged and older adults aged 45 and above. The CHARLS survey promotes interdisciplinary research on China's aging issues and provides a more scientific basis for formulating and improving China's relevant policies [22].

1:4 million national geographic map data comes from the National Basic Geographic Information Center (http://nfgis.nsdi.gov.cn).The temperature data comes from the National Science and Technology Basic Condition Platform-National Earth System Scientific Data Sharing Service Platform-Loess Plateau Scientific Data Center (http://loess.geodata.cn).

B. ARTHRITIS DEFINITION

Arthritis is a common chronic disease, which generally refers to inflammatory diseases that occur in human joints and surrounding tissues [2]. The clinical manifestations are joint redness, swelling, heat, pain, dysfunction and joint deformities. In addition to joint pain, arthritis is also accompanied by joint swelling and movement disorders. Even some patients have joint deformities and disability, leading to labor loss. Besides, this survey did not distinguish the types of arthritis, so we did not research by type of arthritis [12].

C. ARTHRITIS AND OTHER HEALTH CONDITIONS

The variables of interest included demographic factors (age, gender and education), self-reported diseases (hypertension, chronic lung diseases, liver diseases, heart diseases, stroke, kidney disease, and digestive diseases), biomedical measures (BMI and CES-D), health-related behavior (smoking and drinking), and other factors (regional division, annual mean temperature, average altitude). The demographic factors, self-reported diseases, biomedical measures, and health-related behavior data all come from CHARLS.

D. INDEX EVALUATION STANDARDS

BMI (body mass index) [23] was calculated as weight divided by height, expressed as kilogram per square meter. The Chinese-specific cut-offs for general adiposity were used, with underweight defined as BMI$<$18.5 kg/m^2, normal weight as BMI 18.5–23.9 kg/ m^2, overweight as BMI 24.0–27.9kg/ m^2, and general obesity as BMI$\geq$28.0kg/ m^2 [24]. The CHARLS measured the symptoms of depression using the Center for Epidemiologic Studies Depression Scale score (CES-D) [26] . The CES-D was validated previously in an older Chinese population and widely used in Chinese studies [26]. The Center for Epidemiologic Studies Depression Scale score (CES-D) items had ten questions, and each item was scored varying from 0 to 3. We summed up all items to derive a CES-D-10 score (0–30, a higher score indicating more depressive symptoms). A previous survey showed that the cutoff point of 10 had high levels of sensitivity (0.85) and specificity (0.80) in Chinese older adults [27]. Thus, we used 10 as the cutoff to generate the binary depression symptoms variable.

E. SPATIAL AUTOCORRELATION ANALYSIS AND SPATIAL DISTRIBUTION

Spatial autocorrelation is a technique tIhat can measure and analyze spatial clustering in data and calculate the degree of correlation between observations in the entire geographic space [28]. Moran's I is usually used for spatial autocorrelation analysis [29]. In this paper, Moran's I analysis aims to study the spatial distribution pattern of arthritis in China's elderly. Moran's I include global Moran's I and local Moran's I, which respectively reflect the target's spatial autocorrelation from the global and local [30]. In our study, the global Moran's I reveal the overall spatial autocorrelation of the entire country, while the local Moran's I focus on each administrative region and its surrounding areas. The global Moran's I index is between −1 and 1 [31]. When evaluating parameters, if $P \geq 0.05$, the spatial distribution of arthritis in the study area is likely to result from a random spatial process. If $P<0.05$ and the Z score's value is positive, it indicates a positive spatial correlation, then the spatial distribution of arthritis is a spatial aggregation pattern. If $P<0.05$ and the Z score's value is negative, it is expressed as a negative spatial correlation, then the spatial distribution of arthritis is a discrete spatial mode [32]. Local Moran's I is used to indicate each provincial unit's spatial autocorrelation with surrounding units [33]. Local Moran's I can detect four types of clusters, respectively reflect high-high (HH, high incidence of unit surrounded by a high incidence of unit), high-low (HL, high incidence of unit surrounded by a low incidence of unit), low-low (LL, low incidence of unit surrounded by a low incidence of unit) and low- high (LH, low incidence of unit surrounded by a high incidence of unit) clustering model [34].

Because the local spatial autocorrelation analysis only reveals each city's relative state and county unit, not the absolute incidence of arthritis. Therefore, based on the maximum and minimum incidence of arthritis in 2011, 2013, and 2015, we divided 31 municipal and county units in China into five categories on the hierarchical chart. The darker the red, the higher the incidence of arthritis.

F. SPATIAL CLUSTER ANALYSIS

Spatial cluster analysis was performed using SaTScan software to detect spatial clustering or high-risk middle-aged and elderly arthritis locations in 2011, 2013, and 2015. Use "pure spatial analysis" to test whether the incidence of arthritis is randomly distributed in space, and if not, to evaluate the statistical significance of the arthritis group. Calculate the likelihood ratio test to determine whether the cluster has the same risk within or outside a specific area. Through repeated calculations 9999 times Monte Carlo, if the P-value <0.01, then explore the clustering's statistical significance. In the data analysis, a discrete Poisson model is used because it is assumed that the number of patients with arthritis in each administrative district is Poisson distributed. The maximum spatial cluster size of data clustering is 50% and 25% of the risk population in the spatial window, and the maximum spatial cluster size of 50% of the risk population is adopted. The result of cluster relative risk (RR) of pure spatial data clustering is input and linked to ArcGIS software for visualization.

G. LOGISTIC REGRESSION MODEL

In this study, based on the Multi-factor Logistic Regression model, all risk factors were sequentially analyzed by single factor logistic regression analysis and multi-factor logistic regression analysis to estimate the model's various factors and parameters. Perform a single factor logistic regression analysis of the risk factors that may be associated with arthritis. This analysis can study the impact of each risk factor on arthritis and its correlation [35]. To further analyze the comprehensive impact of various factors on the incidence of arthritis, the statistically significant factors (P-value $<$ 0.05) selected in the single-factor logistic regression analysis were included in the multi-factor logistic regression analysis. Simultaneously, evaluate the goodness of fit of the model, analyze the significance of the model's regression coefficients, and the impact of each risk factor on the incidence. In the regression analysis, the Backward Wald method was used to exclude factors that were not statistically significant (P-value$>$0.05). The introduction and elimination criteria were both 0.05, the critical classification value was 0.5(When the critical classification value$>$0.5 has arthritis, $<$0.5 does not have arthritis.), and the maximum number of iterations was 20. After multiple iterations of the model, factor variables that are not statistically significant are deleted. The factor and multi-factor logistic regression analysis are completed by statistical analysis software SPSS 25.0 software.

H. RANDOM FOREST MODEL

The random forest can deal with nonlinear problems, has good anti-noise ability, and is not easy to fall into 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 [36], [37]. The building blocks of the decision tree-based modeling approach, the random forest model, are bootstrapped and are called bagged aggregates [38], [39]. Random forest models randomly use bagging to identify features, thereby separating each node by selecting the most critical possible to assess or predictive variables, which will improve the model's accuracy without causing over-fitting [40]. At present, the random forest model has been widely applied in predict and assess soil moisture [41], shallow water level [42], hydrology [43], and environmental management [44]. In random forest, factors with significant influence in logistic regression are included as independent variables into random forest modeling, and whether arthritis is present as the dependent variable. The total data is 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.

III. EXPERIMENTS AND RESULTS

A. CASE ANALYSIS

By analyzing the gender and age distribution of arthritis cases reported in China in 2011, 2013, and 2015 (Figure 1,2 and 3), we could obtain more intuitive information. In 2011, the overall incidence was highest in the 65-69 age group (40.7%) and women (47.8%), and highest in men (34.8%) in the 70-74 age group. In 2013, the overall incidence was highest in the 65-69 age group (42.3%) and the female age group (49.6%), while the male age group (35.6%) was highest in the 70-74 age group. In 2015, the overall incidence (54.2%), female incidence(61.1%), and male incidence(47.3%) was highest in the 70-74 age group. In addition, regardless of the age group, female's incidence is higher than that of male and the total incidence.

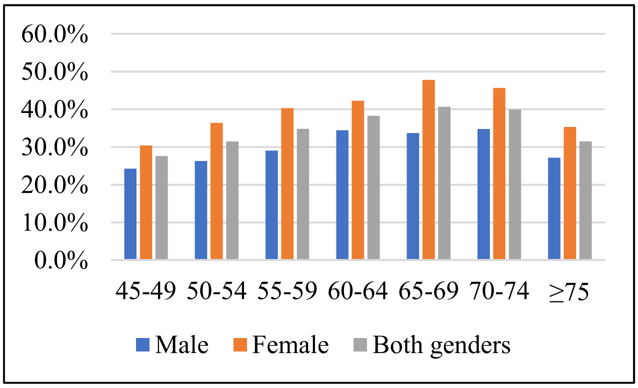

FIGURE 1. Incidence of arthritis by gender in 5-year age groups in the CHARLS 2011 national survey.

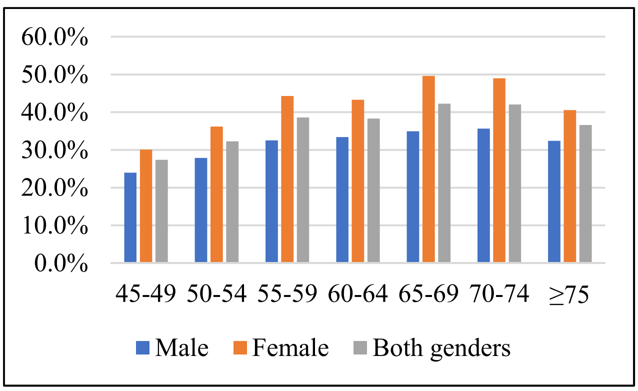

FIGURE 2. Incidence of arthritis by gender in 5-year age groups in the CHARLS 2013 national survey.

B. SPATIAL AUTOCORRELATION ANALYSIS

Spatial autocorrelation analysis is divided into global spatial autocorrelation and local spatial autocorrelation. The global spatial autocorrelation represents the whole region's geographical differences, while the local spatial autocorrelation represents the local cluster differences.

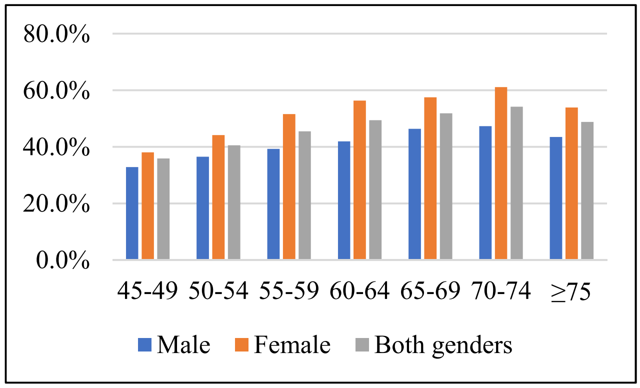

FIGURE 3. Incidence of arthritis by gender in 5-year age groups in the CHARLS 2015 national survey.

TABLE 1. Global spatial autocorrelation analysis and significance rest results.

	Moran's I	Z-score	P-value
2011	0.690929	3.807823	<0.01
2013	0.789083	4.338544	0.000014
2015	0.878335	4.807492	0.000002

1) GLOBAL SPATIAL AUTOCORRELATION

Table 1 shows the global spatial autocorrelation analysis and significance test results in 2011, 2013, and 2015. In 2011, 2013, and 2015, Moran I's spatial autocorrelation in arthritis was 0.690929, 0.789083, and 0.878335, respectively, which were statistically significant when $P \leq 0.001$. Therefore, there was significant spatial autocorrelation in these three years. In the global spatial autocorrelation analysis, Moran I in arthritis was more significant than 0, with a positive correlation, and the range in the past five years ranged from 0.690929 to 0.878335, showing an apparent upward trend. This suggests that spatial autocorrelation and geographic differences in the disease will increase over time.

2) LOCAL SPATIAL AUTOCORRELATION ANALYSIS

In 2011, 2013, and 2015, the HH clustering characteristics of arthritis were mainly found in Chongqing and Gansu, Shaanxi, Sichuan, Yunnan, Hunan, Hubei, Guizhou, and Guangxi in China. Regions with LL clustering characteristics are mainly located in Beijing, Tianjin, Heilongjiang, Jilin, Liaoning, Shaanxi, Hebei, Shanxi, Shandong, Henan, Jiangsu, Zhejiang, Jiangxi, Fujian, and Inner Mongolia Autonomous Region (Figure 2, 3, and 4).

In 2011, 2013, and 2015, arthritis's spatial characteristics did not change significantly in northern China, while HH, HL, LH, and LL clusters were more evident in southern China (Figure 4, 5, and 6). Moreover, regions with HH clustering characteristics are mostly concentrated in northwest, southwest, South, and Central China. In contrast, areas with LL

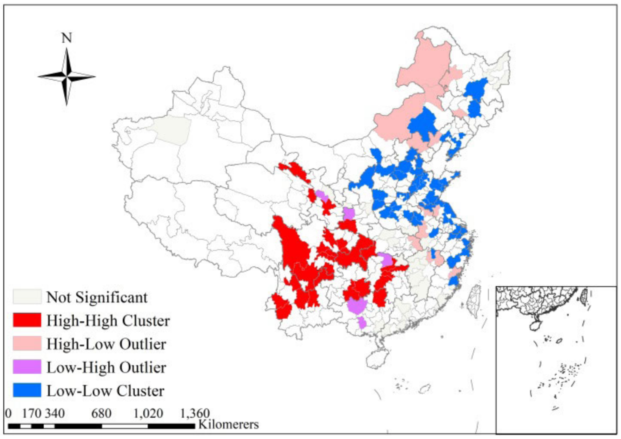

FIGURE 4. The spatial clusters of arthritis in 2011.

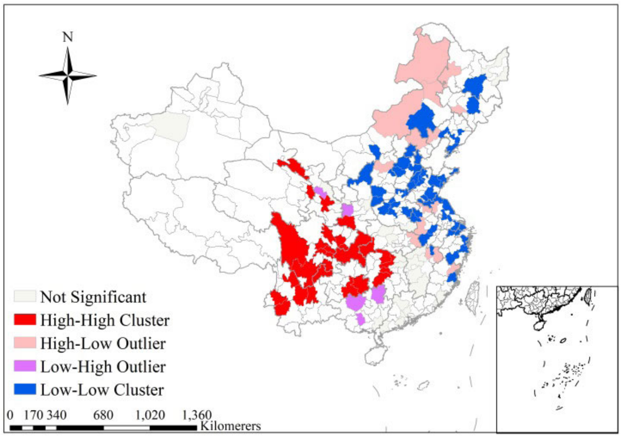

FIGURE 5. The spatial clusters of arthritis in 2013.

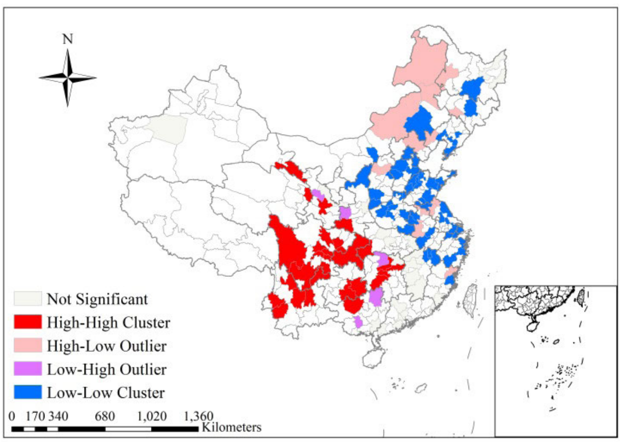

FIGURE 6. The spatial clusters of arthritis in 2015.

clustering characteristics are mostly concentrated in North, East, and northeast China. It can be seen that the incidence of arthritis in the northwest, southwest, South, and Central China is significantly higher than that in North, East, and Northeast China.

C. SPATIAL DISTRIBUTION OF ARTHRITIS IN MIDDLE-AGED AND OLDER ADULTS IN CHINA

Figures 7, 8, and 9 show the incidence of arthritis in middle-aged and older adults in different provinces in 2011, 2013, and 2015. The lowest and highest values in 2011, 2013, and 2015 were 0.000001 and 0.906250, respectively. In 2011,2013, and 2015 the high incidence of arthritis in middle-aged and older adults in China was mainly concentrated in southwest China (The darker the red, the higher the incidence of arthritis.).

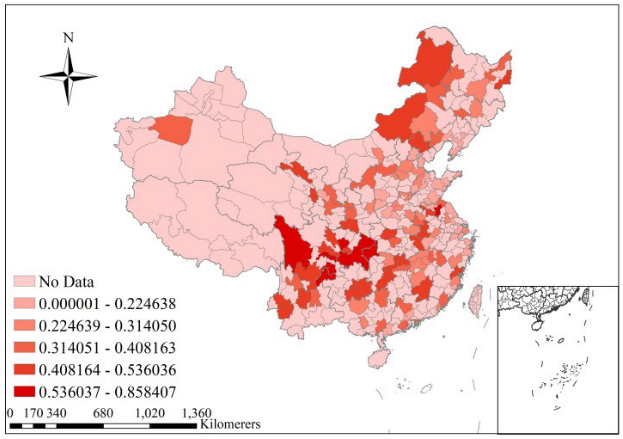

FIGURE 7. The geographical distribution of the incidence of arthritis in 2011.

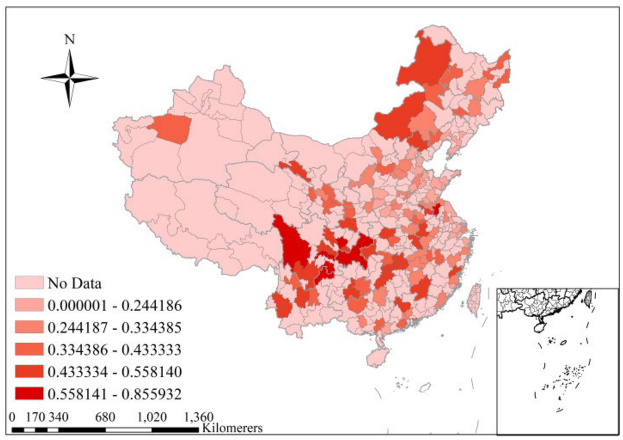

FIGURE 8. The geographical distribution of the incidence of arthritis in 2013.

D. SPATIAL CLUSTERS OF ARTHRITIS IN CHINA

Table 2 and Figure 10 show the statistically significant space clusters for 2011, 2013, and 2015. Space scan tests identified three arthritis space clusters between January 1, 2011, and December 31, 2015 ($P < 0.01$). From January 1, 2011, to December 31, 2015, the most likely clusters were observed, including 22 locations with a radius of 749.69 km, mainly distributed in Yunnan, Sichuan, Chongqing, Guangxi, Hubei,

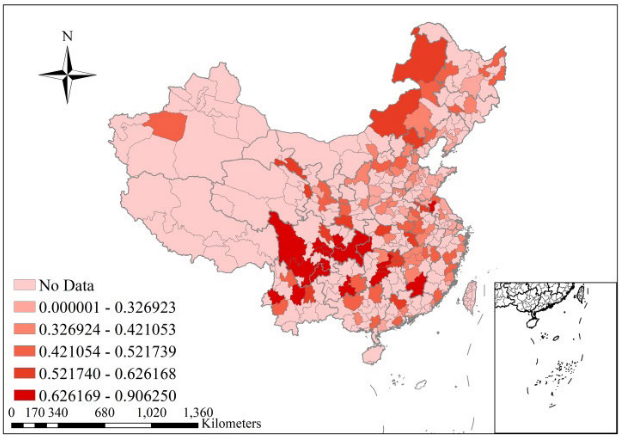

FIGURE 9. The geographical distribution of the incidence of arthritis in 2015.

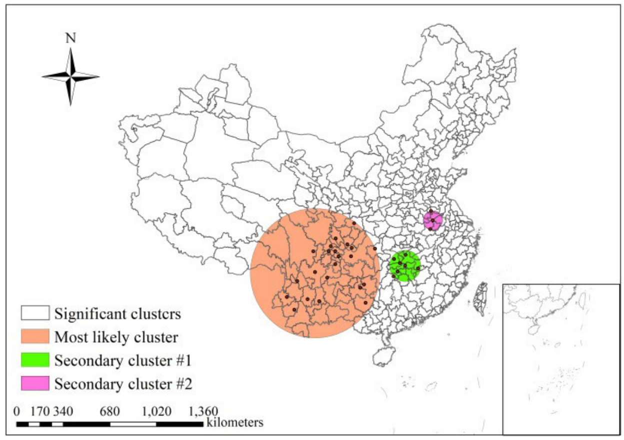

FIGURE 10. Space clusters of arthritis in the elderly in China in 2011, 2013 and 2015.

TABLE 2. Results of spatial clustering test for the incidence of arthritis in Chinese middle-aged and elderly people in 2011, 2013 and 2015.

	Most likely cluster	Secondary cluster1	Secondary cluster2
Region	Baoshan, Chuxiong, Zhaotong, Kunming, Lincang, Lijiang, Guang'an, Chengdu, Liangshan, Nanchong, Yibin, Ziyang, Neijiang, Ganzi, Mianyang, Meishan, Chongqing, Hechi, Enshi, Hanzhong, Qiannan, Qiandong	Yichun, Shaoyang, Yueyang, Changsha, Yiyang, Loudi, Changde	Huainan, Suzhou, Lu'an
Provinces	Yunnan, Sichuan, Chongqing, Guangxi, Hubei, Shaanxi, Guizhou	Jiangxi, Hunan	Anhui
Coordinates	27.887752 N, 102.273503 E	28.234889 N, 112.945473 E	32.631847 N, 117.006389 E
Radius (km)	749.69	181.73	113.42
LLR	513.483511	45.186210	8.991361
P	<0.01	<0.01	<0.01
Number of cases observed	5763	1336	479
Assessment number of cases	3876.80	1026.26	392.94
RR	1.68	1.32	1.22

Shaanxi, and Guizhou. Secondary cluster1 contains 7 locations with a radius of 181.73km, spread primarily on Jiangxi and Hunan. Secondary cluster2 includes three areas, mainly concentrated in Anhui Province.

E. SUBJECTS INCIDENCE OF ARTHRITIS

Among 33,934 subjects, the incidence of arthritis was 38.9%. Among the 13399 arthritis patients, men accounted for 42.81%, and women accounted for 57.2%. The highest proportion of patients in the 60-64 age group is 19.5%. The proportion of arthritis patients with gastrointestinal diseases (38.0%) is higher than those without arthritis (20.0%). The proportion of arthritis patients with hypertension (34.0%) is higher than those without arthritis (27.0%). The proportion of people with arthritis who drink alcohol (58.0%) is higher than those without arthritis (56.0%). The proportion of smokers (39.3%) among arthritis patients is lower than those without arthritis (43.0%). The education level of arthritis patients is mainly middle and low education level, and their illiteracy accounts for 30.1%, followed by elementary school education level (24.3%). Patients with a BMI between 18.5 and 23.9 accounted for the highest proportion (48.9%), followed by patients with BMI between 24.0 and 27.9 (30.9%).36.7% of patients showed depression. Geographically, the proportion of arthritis patients living in the southwest region is 26.2%, which is higher than that in other areas. When the annual average temperature of the respondent's residence is 15.1∼20.0°C, the incidence of arthritis is the highest, being 52.1%. According to altitude, patients living in areas below 200 meters above sea level account for the highest proportion (46.6%).

F. LOGISTIC REGRESSION ANALYSIS OF FACTORS ASSOCIATED WITH ARTHRITIS

The results showed that gender, age, educational level, BMI, CES-D, altitude, geographical location, and average air temperature all had effects on the incidence of arthritis. The risk was higher in women than in men (OR=1.23, P<0.001), and in the 65-69 age group than in other age groups (OR=1.45, P<0.001). Patients with the digestive disease had a higher risk of arthritis than those with other chronic conditions (OR=2.05, P<0.01). The risk of arthritis in middle-aged and older adults decreases with increasing educational attainment. The highest risk was associated with a BMI of 28.0 OR greater (OR=1.96, P<0.01). Those with depressive symptoms had a higher risk of arthritis than those without (OR=1.48, P<0.01). The risk of arthritis in the middle and older adults in southwest China was higher than that in other regions (OR=2.99, P<0.01), followed by central China

TABLE 3. Characteristics of the participants by arthritis status in the 2011, 2013 and 2015 national survey.

	Arthritis			
	No (n=20535)	Yes (n=13399)	x^2	P
Gender			141.20	<0.01
Male	10143（0.49）	5736（0.43）		
Female	10392（0.51）	7663（0.57）		
Age			151.20	<0.01
45~49	3761（0.18）	1906（0.14）		
50~54	3325（0.16）	2032（0.15）		
55~59	3882（0.19）	2518（0.19）		
60~64	3591（0.17）	2609（0.19）		
65~69	2432（0.12）	1948（0.15）		
70~	3544（0.17）	2386（0.18）		
Hypertension			178.03	<0.01
No	14937（0.73）	8837（0.66）		
Yes	5598（0.27）	4562（0.34）		
Chronic lung disease			382.33	<0.01
No	18460（0.90）	11067（0.83）		
Yes	2075（0.10）	2332（0.17）		
Heart disease			432.56	<0.01
No	18091（0.88）	10694（0.80）		
Yes	2444（0.12）	2705（0.20）		
Stroke			43.05	<0.01
No	19979（0.97）	12864（0.96）		
Yes	556（0.03）	535（0.04）		
Live disease			157.60	<0.01
No	19755（0.96）	12483（0.93）		
Yes	780（0.04）	916（0.07）		
Kidney disease			465.79	<0.01
No	19426（0.95）	11806（0.88）		
Yes	1109（0.05）	1593（0.12）		
Digestive disease			1263.2	<0.01
No	16348（0.80）	8310（0.62）		
Yes	4187（0.20）	5089（0.38）		
Smoking			50.54	<0.01
Yes	8876（0.43）	5270（0.39）		
No	11659（0.57）	8129（0.61）		
Drinker			11.66	<0.01
Yes	11598（0.56）	7819（0.58）		
No	8937（0.44）	5580（0.42）		
Education			221.83	<0.01
Illiteracy	5472（0.27）	4083（0.30）		
Less than elementary school	3571（0.17）	2679（0.20）		
Elementary school	4869（0.24）	3259（0.24）		
Middle school	4290（0.21）	2305（0.17）		
High School	1547（0.08）	731（0.05）		
Above vocational school	786（0.04）	342（0.03）		
BMI			33.12	<0.01
<18.5	1268（0.06）	859（0.06）		
18.5~23.9	10470（0.51）	6553（0.49）		
24.0~27.9	6378（0.31）	4143（0.31）		
≥28.0	2419（0.12）	1844（0.14）		
CES-D			473.54	<0.01
Normal	15243（0.74）	8488（0.63）		
Depression	5292（0.26）	4911（0.37）		
Regional division			1243.7	<0.01
Northeast	1545（0.08）	770（0.06）		
North	2724（0.13）	1596（0.12）		
East	6953（0.34）	3132（0.23）		
South	1978（0.10）	1030（0.08）		
Central	3288（0.16）	2275（0.17）		
Northwest	1483（0.07）	1093（0.08）		
Southwest	2564（0.12）	3503（0.26）		
Annual mean temperature (℃)			153.94	<0.01
0~5	530（0.03）	420（0.03）		
5.1~10.0	2620（0.13）	1621（0.12）		
10.1~15.0	4787（0.23）	2573（0.19）		
15.1~20.0	9461（0.46）	6981（0.52）		
>20.0	3137（0.15）	1804（0.13）		
Average altitude(m)			754.77	<0.01
<200	12570（0.61）	6245（0.47）		
201~500	2615（0.13）	2528（0.19）		
501~1000	2970（0.14）	2272（0.17）		
1001~2000	2040（0.10）	1991（0.15）		
2001~	340（0.02）	363（0.03）		

TABLE 3. *(Continued.)* Characteristics of the participants by arthritis status in the 2011, 2013 and 2015 national survey.

(OR=1.76, P<0.01). Besides, annual mean air temperature and mean altitude were risk factors for arthritis.

G. RISK ANALYSIS OF ARTHRITIS

In this study, sensitivity, specificity, assessment agreement rate, AUC (area under the ROC curve), and Youden index were used to evaluate the assessment model. The multi-factor Logistic regression analysis was used to establish the arthritis risk assessment model on the training set to assess the test set. The sensitivity was 0.372, the specificity was 0.849, the assessment consensus rate was 0.666, the AUC was 0.687, and the Youden index was 0.221. The random forest algorithm was used to establish an arthritis risk assessment model on the training set to assess the test set. The sensitivity was 0.822, the specificity was 0.746, the assessment agreement rate was 0.822, the AUC was 0.872, and the Youden index was 0.569(Table 5 and Figure 11). Therefore, it can be seen that the assessment consensus rate of the random forest model for a disease is higher than the diabetes risk assessment model established by multi-factor Logistic regression analysis. In addition, the random forest model cannot explain the function direction of independent variables and the relative risk degree of influencing factors, but multi-factor Logistic regression analysis can define the model and variables well.

The data set is randomly divided into a training set and a test set according to 7:3. The training set is used to establish a random forest model and a multi-factor Logistic regression model, and the test set is used to test the assessment effects of the multi-factor Logistic regression model and the random forest model. The assessment value of the arthritis incidence probability is obtained through the model calculation, and the risk assessment map of the test set is obtained by using the inverted distance weight space interpolation method based

TABLE 4. Multivariate factors affecting the occurrence of arthritis in middle-aged and older people Logistics linear regression analysis.

	P	OR	95%CI Lower	95%CI Upper
Gender	<0.01	1.23	1.17	1.29
Age	<0.01			
50~54	<0.01	1.229	1.131	1.335
55~59	<0.01	1.295	1.195	1.403
60~64	<0.01	1.344	1.239	1.458
65~69	<0.01	1.450	1.328	1.584
70~	<0.01	1.300	1.193	1.417
Hypertension	<0.01	1.24	1.18	1.31
Chronic lung disease	<0.01	1.45	1.35	1.55
Heart disease	<0.01	1.51	1.41	1.62
Stroke	<0.01	123	1.08	1.41
Liver disease	<0.01	1.36	1.23	1.52
Kidney disease	<0.01	1.71	1.57	1.86
Digestive disease	<0.01	2.05	1.95	2.16
Education	<0.01			
Less than elementary school	0.136	1.06	0.98	1.13
Elementary school	0.751	1.01	0.94	1.08
Middle school	0.020	0.91	0.85	0.99
High School	0.028	0.88	0.79	0.99
Above vocational school	<0.01	0.77	0.67	0.89
BMI	<0.01			
18.5~23.9	0.132	1.08	0.98	1.19
24.0~27.9	<0.01	1.20	1.08	1.33
≥28.0	<0.01	1.39	1.23	1.56
CES-D	<0.01			
Depression	<0.01	1.42	1.31	1.54
Regional division	<0.01			
North region	0.041	1.15	1.01	1.31
East region	<0.01	1.28	1.12	1.47
South region	<0.01	1.59	1.37	1.85
Central region	<0.01	1.76	1.53	2.03
Northwest region	<0.01	1.60	1.36	1.88
Southwest region	<0.01	2.99	2.53	3.52
Annual mean temperature (℃)	<0.01			
5.1~10.0	<0.01	0.74	0.63	0.87
10.1~15.0	<0.01	0.56	0.47	0.67
15.1~20.0	<0.01	0.71	0.60	0.84
>20.0	<0.01	0.63	0.53	0.75
Average altitude	<0.01			
201~500（m）	<0.01	1.29	1.19	1.41
501~1000（m）	<0.01	1.29	1.18	1.41
1001~2000（m）	0.003	1.19	1.06	1.33
2001~（m）	0.017	1.26	1.04	1.52
Constant	<0.01	0.20		

TABLE 5. Multi-factor logistic regression model and random forest model fitting results.

	Multi-factor logistic regression model fitting results	Random forest model fitting results
Sensitivity	0.372	0.822
Specificity	0.849	0.746
Assessment consistency	0.666	0.822
AUC	0.687	0.872
Youden Index	0.221	0.569

on the assessment probability value. To verify whether the model's expected risk result is consistent with the actual incidence of arthritis, ArcGIS 10.6 is used to overlay the actual incidence of the test set with the arthritis risk assessment map. In the assessment map of arthritis risk generated by the random forest algorithm model (Figure 12), the high-risk assessment areas are mainly located in Sichuan, Chongqing,

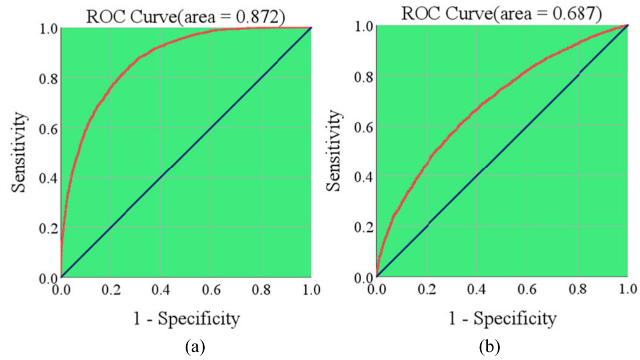

FIGURE 11. ROC curve of random forest model (a) and multi-factor logistic regression model (b).

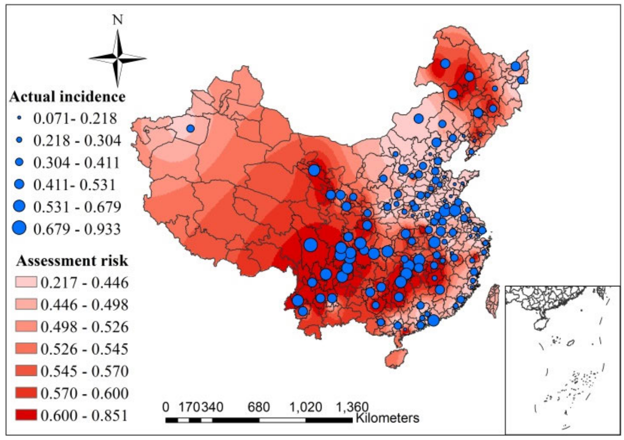

FIGURE 12. Disease risk assessment (Random forest model).

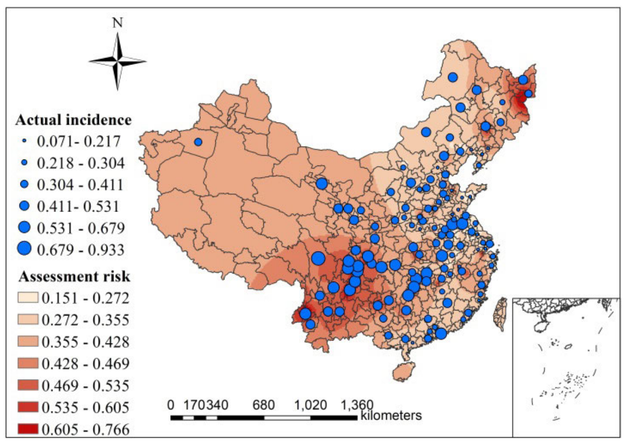

FIGURE 13. Disease risk assessment (Multi-factor logistic regression model).

Guizhou, Guangxi, Hubei, Hunan, Gansu, Shaanxi, Liaoning, Jilin, Heilongjiang, and eastern Inner Mongolia Autonomous Region. The assessment results are consistent with the actual incidence. The multi-factor Logistic regression model generated arthritis risk assessment map (Figure 13), the assessment results are far from the real incidence rate.

IV. DISCUSSION

In China, arthritis poses a severe threat to the people and society, especially the elderly. Before the age of 69, the incidence of arthritis is gradually increasing with age. Although the incidence of arthritis after 69 years of age has declined compared with the previous, the incidence is still very high, so the age growth is an essential factor in arthritis [16]. Also, in this study, it can be seen that the high incidence of arthritis in the middle-aged and elderly in China is mainly in the age group of 65-69 years, indicating that this age group has more arthritis patients.

Gender is also one of the essential factors affecting the incidence of arthritis. It can be seen from this study that the incidence of arthritis in Chinese middle-aged and older women is higher than that of men, which is consistent with the results of a previous study [12]. The higher incidence of female arthritis may also be related to their physiological reasons. Although all patients with arthritis have disorders of the internal environment of sex hormones, the difference between men and women is more prominent. Therefore, changes in women's hormone levels may be one of the reasons for the higher incidence of arthritis in women. Consequently, it can be seen that arthritis is still a public health issue worthy of attention in our country [4]. If this situation is not improved, more and more middle-aged and older adults will have arthritis. It will inevitably place a heavy burden on society.

In 2011, 2013, and 2015, the Moran's I coefficient of arthritis incidence in my country was between 0.690929-0.878335 and showed an increasing trend. The research results show that the incidence of arthritis in my country is non-random in spatial distribution, and there is a positive spatial autocorrelation. The degree of local spatial autocorrelation is different from the incidence of arthritis, but there is a correlation in some aspects. The local spatial autocorrelation analysis is based on the correlation and comprehensive consideration of regional geography, population, and other factors. It can describe the relationship between the incidence of arthritis in a specific spatial area and the surrounding spatial area [45]. Local spatial autocorrelation analysis found that the High-High distribution pattern of arthritis is mainly concentrated in Gansu, Sichuan, Chongqing, Yunnan, and Hunan. These provinces are also areas with high arthritis incidence, and these "hot spots" areas should continue to be strengthened for monitoring and prevention. Low-Low distribution models are mainly concentrated in Beijing, Tianjin, Hebei, Liaoning, Shanxi, Shandong, Henan, Jiangsu, Zhejiang, and Anhui. For areas with "cold spots" where arthritis is affected, especially areas with Low-High distribution, one cannot relax vigilance to avoid an increase in arthritis incidence.

As a supplement to local spatial autocorrelation analysis, spatial scanning statistics detect the specific conditions of arthritis incidence space and aggregation range simultaneously and give a risk assessment. The retrospective discrete Poisson distribution model scanned three spatiotemporal clustering regions. There are some discrepancies with spatial autocorrelation analysis results due to the different principles of analysis methods. The spatial scan detected that Most likely clusters were mainly distributed in Yunnan, Sichuan, Chongqing, Guangxi, Hubei, Shaanxi, and Guizhou. Secondary cluster1 is primarily distributed in Jiangxi and Hunan. Secondary cluster2 mainly gathers in Anhui, which is consistent with the research results of local spatial autocorrelation.

With the increase of age, the rise in body mass, the prolongation of the disease course, the patient's body's immune function is abnormal, the function of many systems is weakened, and multiple diseases are prone to occur. In previous studies, it was also found that arthritis can induce other chronic diseases. Nakamura *et al.* [46] reported that chronic inflammation in arthritis patients combined with drug exposure and its toxic effects increases kidney disease risk. This study found that hypertension, chronic pneumonia, liver disease, heart disease, kidney disease, and gastrointestinal disease are all risk factors for arthritis. Li *et al.* [12] also found this relationship but did not make an explanation. However, the scholar found that the relationship between arthritis and liver disease is still uncertain, and it exists in both directions with other chronic diseases.

In the study, it was found that the lower the academic background, the higher the incidence of arthritis in the elderly. It may be because the increase in education level often leads to higher income and better living conditions, which affects the elderly's health. Besides, the level of education reflects the patient's understanding and acceptance of chronic diseases to a certain extent. People with high education level may pay more attention to their daily maintenance, thereby reducing the risk of illness.

People who are overweight and obese also have a relatively high risk of arthritis, especially for osteoarthritis. From a weight-bearing point of view, the weight load caused by obesity will overload the weight-bearing joints (knee joint, hip joint, etc.). This may cause uneven stress on the joint surface and disorder of joint function, resulting in the loss of cartilage and the formation of osteophytes, which in turn leads to osteoarthritis [47].

In addition, domestic studies have confirmed that long-term pain in patients with arthritis can cause varying degrees of mental health problems such as anxiety and depression. This study found that people with depressive symptoms have a high risk of arthritis, which may be due to the chronic exposure of patients with depression to negative emotions. Its hypothalamic-pituitary-adrenal dysfunction and peripheral release of glucocorticoids are accompanied by activation of the sympathetic nervous system and decreased activation of the parasympathetic nervous system, leading to autoimmune diseases and increasing the risk of arthritis [48].

The study also found that the average altitude, geographical location, and average temperature are strictly related to the incidence of arthritis in the elderly. Respondents living below 200 meters above sea level have a much lower risk of arthritis than those surveyed living above 200 meters. The risk of disease in Southwest China is much higher than in other areas, which is consistent with the results of local spatial autocorrelation and spatial scanning. The result that the incidence was higher in southern regions was not consistent with Tie's findings [20]. It was supposed that the difference between southern regions and northern regions was mostly temperature and humidity, and a higher incidence of arthritis can occur where it is cold and damp [20]. With the improvement in economic conditions, the heating system in the northern regions has greatly improved and fewer people work outside than before. The cold weather may not affect people's joints in the northern regions; however, there are more rains in the southern regions than northern regions, and the southern region is more humid, which may be contributable to the higher incidence of arthritis [16]. There is spatial heterogeneity in the influencing factors of arthritis incidence. The critical prevention and control areas for arthritis are mainly concentrated in South, Central, Northwest, and Southwest China. Different provinces (municipalities and autonomous regions) should carry out health education and health promotion in response to the risk factors of arthritis. Besides, this survey did not distinguish between arthritis types, so we cannot estimate the incidence by type of arthritis.

The traditional data analysis method is difficult to avoid the interaction between the respective variables. As an emerging machine learning algorithm, the random forest algorithm runs steadily, does not require data sets, and does not have overfitting and collinearity problems. Therefore, it is widely used in disease risk assessment. Casanova *et al.* [49] used data from the Jackson Heart Research Cohort to predict arthritis risk through multi-factor Logistic regression analysis and random forest algorithm. They found that the random forest algorithm's prediction accuracy was higher than the multi-factor Logistic regression analysis.

In the random forest algorithm model's arthritis risk assessment map, the assessment result is consistent with the actual incidence rate. The multi-factor Logistic regression model generated an arthritis risk assessment map. The assessment results and the actual incidence have a large gap. So the random forest algorithm model can achieve a more accurate arthritis risk assessment effect. The use of data mining algorithms to establish a concise and accurate arthritis risk assessment model is an innovative research point on assessing arthritis disease risk in middle-aged and older adults in China. The random forest algorithm has a good effect in assessing arthritis risk, but multivariate logistic regression analysis has an intuitive explanation for arthritis risk factors. Therefore, the advantages of the two models should be combined in practical applications to make them play the most significant value in disease risk assessment. Because CHARLS did not investigate information such as geography and environment, the impact of these two factors on arthritis was supplemented in this article, making the research results more convincing. Finally, the discovery of risk factors should be interpreted with caution, because different forms of arthritis have various risk factors, so our findings may differ from the research on specific kinds of arthritis.

V. CONCLUSION

Firstly, in this paper, spatial autocorrelation and spatial clustering analysis were used to analyze arthritis's spatial distribution characteristics. Secondly, we used the logistic regression model to explore the risk factors of arthritis in detail. Finally, the logistic regression model and the random forest model were used to assess arthritis risk in middle-aged and older adults in China. The results showed that the aggregation areas of patients with arthritis were mainly southwest, northwest, and Central China. Education level, BMI, CES-D, altitude, region, temperature, sex, age, hypertension, liver, heart, kidney, digestive, stroke, and lung disease all had effects on arthritis. The random forest algorithm's assessment accuracy is higher than that of the logistic regression model, and the assessment results show that the high-risk regions are located in the northeast, southwest, northwest, and central China. Therefore, our method can not only analyze the spatial distribution characteristics and influencing factors of arthritis but also accurately assess the risk of arthritis. In addition, CHARLS only investigated the data of two provinces in China in 2012. Our research scope is the spatial distribution characteristics and risk assessment of Chinese middle-aged and older adults, so we did not use the 2012 data. In the 2014 survey, CHARLS did not have chronic diseases (including arthritis), so there is no 2014 data in our study. 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.

REFERENCES

[1] R. C. Lawrence, D. T. Felson, C. G. Helmick, L. M. Arnold, H. Choi, R. A. Deyo, S. Gabriel, R. Hirsch, M. C. Hochberg, G. G. Hunder, J. M. Jordan, J. N. Katz, H. M. Kremers, F. Wolfe, and N. Arthritis Data Workgroup, "Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II," *Arthritis Rheumatism*, vol. 58, no. 1, pp. 26–35, Jan. 2008.

[2] Y. Rao, X. Xu, D. Liu, C. Reis, I. Newman, L. Qin, M. Sharma, J. Shen, and Y. Zhao, "Health-related quality of life in patients with arthritis: A cross-sectional survey among middle-aged adults in Chongqing, China," *Int. J. Environ. Res. Public Health*, vol. 15, no. 4, p. 768, Apr. 2018.

[3] L. Fu, J. Zhang, L. Jin, Y. Zhang, S. Cui, and M. Chen, "A case–control study of rheumatoid arthritis revealed abdominal obesity and environmental risk factor interactions in Northern China," *Mod. Rheumatol.*, vol. 28, no. 2, pp. 249–257, 2018.

[4] G. Gong and J. Mao, "Health-related quality of life among Chinese patients with rheumatoid arthritis: The predictive roles of fatigue, functional disability, self-efficacy, and social support," *Nursing Res.*, vol. 65, no. 1, pp. 55–67, 2016.

[5] S. Shinan-Altman and S. Afuta-Goldstein, "Contribution of the self-regulation model to understanding the health related quality of life of rheumatoid arthritis patients," *Qual. Life Res.*, vol. 29, no. 2, pp. 403–412, Feb. 2020.

[6] C. A. Hitchon, L. Zhang, C. A. Peschken, L. M. Lix, L. A. Graff, J. D. Fisk, S. B. Patten, J. Bolton, J. Sareen, R. El-Gabalawy, J. Marriott, C. N. Bernstein, and R. A. Marrie, "Validity and reliability of screening measures for depression and anxiety disorders in rheumatoid arthritis," *Arthritis Care Res.*, vol. 72, no. 8, pp. 1130–1139, Aug. 2020.

[7] G. Guo, T. Fu, R. Yin, L. Zhang, Q. Zhang, Y. Xia, L. Li, and Z. Gu, "Sleep quality in Chinese patients with rheumatoid arthritis: Contributing factors and effects on health-related quality of life," *Health Qual. Life Outcomes*, vol. 14, no. 1, p. 151, Dec. 2016.

[8] A. K. M. Kamruzzaman, M. R. Chowdhury, M. N. Islam, I. Sultan, S. Ahmed, A. Shahin, M. M. Alam, M. A. Azad, M. M. Hassan, M. Z. Amin, S. Sinha, H. I. Ahmad, M. N. Shazzad, S. N. Ahmad, S. A. Haq, and J. J. Rasker, ''The knowledge level of rheumatoid arthritis patients about their disease in a developing country. A study in 168 Bangladeshi RA patients,'' *Clin. Rheumatol.*, vol. 39, no. 4, pp. 1315–1323, Apr. 2020.

[9] X. Jiang, M. E. C. Sandberg, S. Saevarsdottir, L. Klareskog, L. Alfredsson, and C. Bengtsson, ''Higher education is associated with a better rheumatoid arthritis outcome concerning for pain and function but not disease activity: Results from the EIRA cohort and Swedish rheumatology register,'' *Arthritis Res. Therapy*, vol. 17, no. 1, p. 317, Dec. 2015.

[10] E. Yelin *et al.*, ''National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions–United States, 2003,'' *Mmwr Morbidity Mortality Weekly Rep.*, vol. 56, no. 1, pp. 4–7, 2007.

[11] B. Zhou, G. Wang, Y. Hong, S. Xu, J. Wang, H. Yu, Y. Liu, and L. Yu, ''Mindfulness interventions for rheumatoid arthritis: A systematic review and meta-analysis,'' *Complementary Therapies Clin. Pract.*, vol. 39, May 2020, Art. no. 101088.

[12] C. Li, T. Liu, W. Sun, L. Wu, and Z.-Y. Zou, ''Prevalence and risk factors of arthritis in a middle-aged and older Chinese population: The China health and retirement longitudinal study,'' *Rheumatology*, vol. 54, no. 4, pp. 697–706, Apr. 2015.

[13] S. Zhao, Y. Chen, and H. Chen, ''Sociodemographic factors associated with functional disability in outpatients with rheumatoid arthritis in southwest China,'' *Clin. Rheumatol.*, vol. 34, no. 5, pp. 845–851, May 2015.

[14] K. E. Barbour, C. G. Helmick, M. Boring, and T. J. Brady, ''Vital signs: Prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation—United States, 2013–2015,'' *Morbidity Mortality Weekly Rep.*, vol. 66, no. 9, pp. 246–253, 2017.

[15] K. E. Barbour, C. G. Helmick, M. Boring, X. Zhang, H. Lu, and J. B. Holt, ''Prevalence of doctor-diagnosed arthritis at state and county levels—United States, 2014,'' *Morbidity Mortality Weekly Rep.*, vol. 65, no. 19, pp. 489–494, 2016.

[16] X. Sun, X. Zhen, X. Hu, Y. Li, S. Gu, Y. Gu, and H. Dong, ''Osteoarthritis in the middle-aged and elderly in China: Prevalence and influencing factors,'' *Int. J. Environ. Res. Public Health*, vol. 16, no. 23, p. 4701, Nov. 2019.

[17] J. Goossens, B. Coustet, E. Palazzo, P. Dieude, and S. Ottaviani, ''Overweight and obesity affect clinical assessment of synovitis in rheumatoid arthritis: Comparison of ultrasonography and clinical exam,'' *Clin. Exp. Rheumatol.*, vol. 37, no. 1, pp. 49–54, 2019.

[18] L. Olaru, L. Soong, S. Dhillon, and E. Yacyshyn, ''Coexistent rheumatoid arthritis and gout: A case series and review of the literature,'' *Clin. Rheumatol.*, vol. 36, no. 12, pp. 2835–2838, Dec. 2017.

[19] D. F. McWilliams, M. Marshall, K. Jayakumar, S. Doherty, M. Doherty, W. Zhang, P. D. W. Kiely, A. Young, and D. A. Walsh, ''Erosive and osteoarthritic structural progression in early rheumatoid arthritis,'' *Rheumatology*, vol. 55, no. 8, pp. 1477–1488, Aug. 2016.

[20] T. Xiao-Jia, Z. Ru-Geng, Z. Meng, H. Ya-Jun, G. Hong-Liang, W. Zhi-Zhou, and M. Guo-Ju, ''Prevalence of knee osteoarthritis in the middle-aged and elderly in China: A meta-analysis,'' *Chin. J. Tissue Eng. Res.*, vol. 22, no. 4, p. 650, 2018.

[21] H. Ahmed, E. M. G. Younis, A. Hendawi, and A. A. Ali, ''Heart disease identification from patients' social posts, machine learning solution on spark,'' *Future Gener. Comput. Syst.*, vol. 111, pp. 714–722, Oct. 2020.

[22] H. Luo, X. Ren, J. Li, K. Wu, Y. Wang, Q. Chen, and N. Li, ''Association between obesity status and successful aging among older people in China: Evidence from CHARLS,'' *BMC Public Health*, vol. 20, no. 1, pp. 1–10, Dec. 2020.

[23] Q. Cai, F. Chen, T. Wang, F. Luo, X. Liu, Q. Wu, Q. He, Z. Wang, Y. Liu, L. Liu, J. Chen, and L. Xu, ''Obesity and COVID-19 severity in a designated hospital in Shenzhen, China,'' *Diabetes Care*, vol. 43, no. 7, pp. 1392–1398, Jul. 2020.

[24] X.-X. Li, L.-X. Wang, H. Zhang, S.-W. Jiang, Q. Fang, J.-X. Chen, and X.-N. Zhou, ''Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010,'' *BMC Public Health*, vol. 14, no. 1, p. 257, Dec. 2014.

[25] Y. Li, L. Zhao, D. Yu, and G. Ding, ''Exposure to the Chinese famine in early life and depression in adulthood,'' *Psychol., Health Med.*, vol. 23, no. 8, pp. 952–957, Sep. 2018.

[26] M. He, J. Ma, Z. Ren, G. Zhou, P. Gong, M. Liu, X. Yang, W. Xiong, Q. Wang, H. Liu, and X. Zhang, ''Association between activities of daily living disability and depression symptoms of middle-aged and older Chinese adults and their spouses: A community based study,'' *J. Affect. Disorders*, vol. 242, pp. 135–142, Jan. 2019.

[27] Y. Li, L. Zhao, D. Yu, and G. Ding, ''Associations between serum uric acid and depression among middle-aged and elderly participants in China,'' *Psychol., Health Med.*, vol. 24, no. 10, pp. 1277–1286, Nov. 2019.

[28] L. Wang, J. Xing, F. Chen, R. Yan, L. Ge, Q. Qin, L. Wang, Z. Ding, W. Guo, and N. Wang, ''Spatial analysis on hepatitis c virus infection in mainland China: From 2005 to 2011,'' *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e110861.

[29] Y. Cai, Y. Wei, X. Han, Z. Han, S. Liu, Y. Zhang, Y. Xu, S. Qi, and Q. Li, ''Spatiotemporal patterns of hemorrhagic fever with renal syndrome in Hebei Province, China, 2001-2016: CAI *et al.*,'' *J. Med. Virol.*, vol. 91, no. 3, pp. 337–346, Mar. 2019.

[30] F. R. Martins-Melo, M. C. C. Pinheiro, A. N. Ramos, C. H. Alencar, F. S. D. M. Bezerra, and J. Heukelbach, ''Spatiotemporal patterns of schistosomiasis-related deaths, Brazil, 2000–2011,'' *Emerg. Infectious Diseases*, vol. 21, no. 10, pp. 1820–1823, Oct. 2015.

[31] B. Zhu, J. Liu, Y. Fu, B. Zhang, and Y. Mao, ''Spatio-temporal epidemiology of viral hepatitis in China (2003–2015): Implications for prevention and control policies,'' *Int. J. Environ. Res. Public Health*, vol. 15, no. 4, p. 661, Apr. 2018.

[32] X. Wu, S. Hu, A. B. Kwaku, Q. Li, K. Luo, Y. Zhou, and H. Tan, ''Spatio-temporal clustering analysis and its determinants of hand, foot and mouth disease in Hunan, China, 2009–2015,'' *BMC Infectious Diseases*, vol. 17, no. 1, pp. 1–9, Dec. 2017.

[33] L. Anselin, ''Local indicators of spatial association-LISA,'' *Geograph. Anal.*, vol. 27, no. 2, pp. 93–115, Sep. 2010.

[34] L. Anselin, I. Syabri, and Y. Kho, ''GeoDa: An introduction to spatial data analysis,'' *Geograph. Anal.*, vol. 38, no. 1, pp. 5–22, Jan. 2006.

[35] K. Zou, F.-K. Xiao, H.-Y. Li, Q. Zhou, L. Ban, M. Yang, C.-F. Kuo, and W. Zhang, ''Risk of cardiovascular disease in Chinese patients with rheumatoid arthritis: A cross-sectional study based on hospital medical records in 10 years,'' *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180376.

[36] F. Kuang-Nana *et al.*, ''A review of technologies on random forests,'' *Stats Inf. Forum*, 2011.

[37] A. Lahouar and J. B. H. Slama, ''Hour-ahead wind power forecast based on random forests,'' *Renew. Energy*, vol. 109, pp. 529–541, Aug. 2017.

[38] L. Breiman, ''Bagging predictors,'' *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[39] B. Xiang, C. Zeng, X. Dong, and J. Wang, ''The application of a decision tree and stochastic forest model in summer precipitation prediction in Chongqing,'' *Atmosphere*, vol. 11, no. 5, p. 508, May 2020.

[40] H. Yao, X. Li, H. Pang, L. Sheng, and W. Wang, ''Application of random forest algorithm in hail forecasting over Shandong Peninsula,'' *Atmos. Res.*, vol. 244, Nov. 2020, Art. no. 105093.

[41] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, ''Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition,'' *Geoderma*, vol. 330, pp. 136–161, Nov. 2018.

[42] J. Koch, H. Berger, H. J. Henriksen, and T. O. Sonnenborg, ''Modelling of the shallow water table at high spatial resolution using random forests,'' *Hydrol. Earth Syst. Sci.*, vol. 23, no. 11, pp. 4603–4619, Nov. 2019.

[43] I. D. Moore, R. B. Grayson, and A. R. Ladson, ''Digital terrain modelling: A review of hydrological, geomorphological, and biological applications,'' *Hydrol. Processes*, vol. 5, no. 1, pp. 3–30, Jan. 1991.

[44] J. C. Ascough, H. R. Maier, J. K. Ravalico, and M. W. Strudley, ''Future research challenges for incorporation of uncertainty in environmental and ecological decision-making,'' *Ecol. Model.*, vol. 219, nos. 3–4, pp. 383–399, Dec. 2008.

[45] H.-X. Rao, X. Zhang, L. Zhao, J. Yu, W. Ren, X.-L. Zhang, Y.-C. Ma, Y. Shi, B.-Z. Ma, X. Wang, Z. Wei, H.-F. Wang, and L.-X. Qiu, ''Spatial transmission and meteorological determinants of tuberculosis incidence in Qinghai Province, China: A spatial clustering panel analysis,'' *Infectious Diseases Poverty*, vol. 5, no. 1, p. 45, Dec. 2016.

[46] T. Nakamura, S.-I. Higashi, K. Tomoda, M. Tsukano, and M. Shono, ''Etanercept can induce resolution of renal deterioration in patients with amyloid a amyloidosis secondary to rheumatoid arthritis,'' *Clin. Rheumatol.*, vol. 29, no. 12, pp. 1395–1401, Dec. 2010.

[47] L. Jiang, X. Xie, Y. Wang, Y. Wang, Y. Lu, T. Tian, M. Chu, and Y. Shen, ''Body mass index and hand osteoarthritis susceptibility: An updated meta-analysis,'' *Int. J. Rheumatic Diseases*, vol. 19, no. 12, pp. 1244–1254, Dec. 2016.

[48] M.-C. Lu, H.-R. Guo, M.-C. Lin, H. Livneh, N.-S. Lai, and T.-Y. Tsai, ''Bidirectional associations between rheumatoid arthritis and depression: A nationwide longitudinal study,'' *Sci. Rep.*, vol. 6, no. 1, p. 20647, Feb. 2016.

[49] R. Casanova, S. Saldana, S. L. Simpson, M. E. Lacy, A. R. Subauste, C. Blackshear, L. Wagenknecht, and A. G. Bertoni, ''Prediction of incident diabetes in the Jackson heart study using high-dimensional machine learning,'' *PLoS ONE*, vol. 11, no. 10, Oct. 2016, Art. no. e0163942.

ZHONGDA REN received the B.S. degree from the Jilin University of Finance and Economics, China, in 2017. He is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. His current research interests include GIS-based epidemiological spatiotemporal analysis and modeling assessment.

KUN YANG received the master's degree from the University of New South Wales, Australia, in 1998. He is currently a Professor with the School of Faculty of Geography, Yunnan Normal University. His research interests include GIS and remote sensing image processing.

WEN DONG is currently an Associate Professor and a master's Tutor with the School of Faculty of Geography, Yunnan Normal University, China. Her research interests include GIS-based epidemiological spatiotemporal analysis and modeling assessment.