

Article

Spatial Differentiation Characteristics of Rural Areas Based on Machine Learning and GIS Statistical Analysis—A Case Study of Yongtai County, Fuzhou City

Ziyuan Wang

School of Public Affairs, Xiamen University, Xiamen 361005, China; 1635241152@xzyz.edu.cn

Abstract: With the development of machine learning and GIS (geographic information systems) technology, it is possible to combine them to mine the knowledge rules behind massive spatial data. GIS, also known as geographic information systems, is a comprehensive discipline, which combines geography and cartography and has been widely used in different fields. It is a computer system for inputting, storing, querying, analyzing, and displaying geographic data. This paper mainly studies the spatial differentiation characteristics of rural areas based on machine learning (ML) and GIS statistical analysis. This paper studies 21 township units in Yongtai County. In this paper, ENVI remote sensing image processing software is used to carry out the geometric correction of Landsat-8 remote sensing data. ML is multidisciplinary and interdisciplinary, involving probability theory, statistics, approximation theory, convex analysis, algorithm complexity theory, and other disciplines. It is specialized in studying how computers simulate or realize human learning behavior to obtain new knowledge or skills, and reorganize existing knowledge structures to continuously improve its own performance. The purpose of using band fusion is to provide more data information for the study and improve the accuracy of land classification results. Through the extraction of evaluation elements, this paper preliminarily confirms the evaluation index object of a rural human settlement environment evaluation system from the perspective of spatial layout rationality. This paper uses a VMD-GWO-ELM-based three-stage evolutionary extreme learning machine evaluation method to simulate the model. In the same way, when the model is trained again, extra weight is given to extract the feature points to reduce the similarity. Experimental data show that GWO-SVM has good classification performance, with the cross-validation rate reaching 91.66% and the recognition rate of test samples reaching 82.41%. The results show that GIS statistics can provide a reference for environmental protection, which is conducive to land-use planning, implementation of environmental impact assessment of land-use planning, and ultimately achieving sustainable development.

Keywords: machine learning; GIS statistical analysis; spatial differentiation feature; prediction model



check for updates

Citation: Wang, Z. Spatial Differentiation Characteristics of Rural Areas Based on Machine Learning and GIS Statistical Analysis—A Case Study of Yongtai County, Fuzhou City. *Sustainability* **2023**, *15*, 4367. <https://doi.org/10.3390/su15054367>

Academic Editors: Chia-Huei Wu and Yu-Hsi Yuan

Received: 1 December 2022

Revised: 22 January 2023

Accepted: 2 February 2023

Published: 1 March 2023



Copyright: © 2023 by the author. 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/>).

1. Introduction

In GIS, geographic data not only contain spatial data related to the geometric characteristics of spatial elements but also include the attribute information of spatial elements, which enables all data related to the spatial location to be analyzed under the framework of GIS. Spatial data analysis was first developed and expanded in cartography and surveying. Because maps can contain various types of additional information, spatial data analysis intersects with many disciplines and has a positive catalytic effect on both.

Machine learning can greatly promote the research of rural spatial differentiation characteristics. Sun Yaohua comprehensively summarizes the latest progress of machine learning applications in wireless communication, which is divided into resource management of the machine learning layer, the network, mobility management of the network layer, and localization of the application layer [1]. Dev S believes that ground-based all-weather cameras have opened up new opportunities for monitoring the earth's atmosphere.

The images captured by the all-weather imager (WSI) can have high spatial and temporal resolution. It is challenging for him to extract valuable information from large amounts of image data by detecting and analyzing various entities in these images. He detailed the latest developments of these technologies and their applications in ground imaging, aiming to bridge the gap between computer vision and remote sensing with illustrative examples. His research lacks a specific practical plan [2]. Taherkhani N believes that in an urban environment, intersections are key locations in terms of road traffic accidents and the number of deaths or injuries. The vehicle ad hoc network (VANET) helps reduce traffic conflicts at intersections by sending warning messages to vehicles. He proposed a centralized and partial data congestion control strategy to use roadside units (RSU) at intersections to control data congestion. The strategy he proposed consists of three units for detecting congestion, clustering messages, and controlling data congestion. The K-means algorithm clusters messages according to the size of the message, the validity of the message, and the message type. His research is not very accurate [3]. Ruske S believes that the characterization of bioaerosols has important implications for the environment and public health sectors. For unsupervised learning, he tested hierarchical clustering with various different links. In supervised learning, he tested 11 methods, including decision trees, ensemble methods, two implementations of support vector machines, and Gaussian methods. He also found that some methods (such as clustering) fail to use other shape information provided by the instrument, while other methods (such as decision trees, ensemble methods, and neural networks) can improve performance to include such information. His research lacks the necessary data [4]. In the above studies, the analysis of machine learning in space utilization is more detailed, but there are great differences between land use and terrain in each region. Therefore, the study should be based on the actual situation of Yongzheng County in Fujian Province.

Data vector quantization is based on land-use type, the population, the urban road network data, and meteorological data to construct a regression model of land use, and provides the empirical data and theoretical basis for urban planning to better serve urban sustainable development. At the same time, in the practice of geological survey information grid, the combination of “three-dimensional map” service technology and unstructured data service technology will help to build a cloud GIS architecture to manage, access, discover, analyze, and integrate distributed geological information resources.

Through the comprehensive evaluation of the rural regional function of each township in Yongtai County, this paper reveals the spatial distribution characteristics of the rural regional function of each township in Yongtai County and finds out the factors restricting its development in the process of rural construction in Yongtai County, so as to provide a reference for government departments to adjust policies in time. At the same time, it provides methods and ideas for other regions in China to evaluate rural regional functions and determine rural development strategies. Through the comprehensive evaluation of the rural economy, tourism, and ecology, we can make clear its important position in rural development, attract the attention of relevant departments, and strengthen the development of rural resources and environmental protection.

2. Machine Learning and Rural Spatial Differentiation Characteristics

2.1. GIS Technology

With the maturity of other computer technologies, the basic functions of geographic information systems have been further expanded. In addition, society’s understanding of geographic information systems has gradually increased, and the demand has also increased significantly, which has caused the application of geographic information systems to receive widespread attention, and the research content has continued to expand and deepen. The architecture of the client–server model is shown in Figure 1 [5].

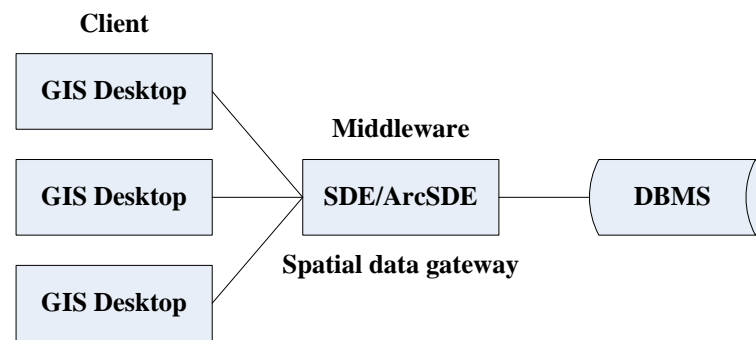


Figure 1. The architecture of the client–server model.

Using spatial analysis technology, you can not only explore a certain spatiotemporal change law between the original spatial data and obtain some new information, but also use the analysis results as the decision-making basis for the next step of spatial behavior, and solve various practical problems related to geographic space. The development of SDM and GIS has the conditions for the integration of the two technologies. In order to be able to effectively and fully utilize the existing massive spatial data, SDM technology and GIS are organically integrated, that is, with the help of GIS for spatial data input, editing, storage, management, query, and display functions are possible, and SDM for spatial data offers powerful analysis and processing functions that enable it to analyze massive spatial data effectively and quickly and obtain useful information from massive spatial data [6].

The concept tree can be used to express the necessary background to control the process of induction. Different levels of concepts are usually organized into a class of concepts according to the order from general to special. The most general concept is the empty description, and the most special concept is the specific value in the database. Based on this, we first establish a case library for case reasoning. The case library is divided into spatial case sub-libraries and geological data case sub-libraries according to different data organization methods. According to the CBR method, the training core for the user's initial learning is determined, the training core passes through training and learning to obtain the initial classifier, then uses this initial classifier to constrain the overall training set to obtain a smaller subset, and then uses this subset for larger-scale training and learning to obtain the final classifier and classification results [7,8].

The main structure of mobile GIS is shown in Figure 2. The data of mobile GIS not only has the characteristics of large amount and complexity of GIS data but also has the characteristics of small storage space of embedded systems and simple and fast calculation and retrieval. Compared with the traditional relational database, complex spatial data has a large amount of data and uneven distribution. But in the embedded system, the storage space is very valuable, which requires a more effective structure and way to make the storage space of GIS data smaller and the memory use smaller. For a mobile GIS database, its query conditions are related to spatial location rather than attribute data, so we must adopt a more flexible mechanism than a desktop operating system to meet the requirements of embedded systems.

The system structure of GIS is shown in Figure 3. GIS is a new technology that integrates computer graphics and databases to store and process spatial information. It organically combines geographical location and related attributes and can map and analyze the existing things and events on the earth. The powerful spatial analysis ability of GIS has opened up a new realm of rural settlement research, making it an important tool for quantitative research of rural settlement at present and in the future, which is of great significance to enhance the scientificity of settlement research.

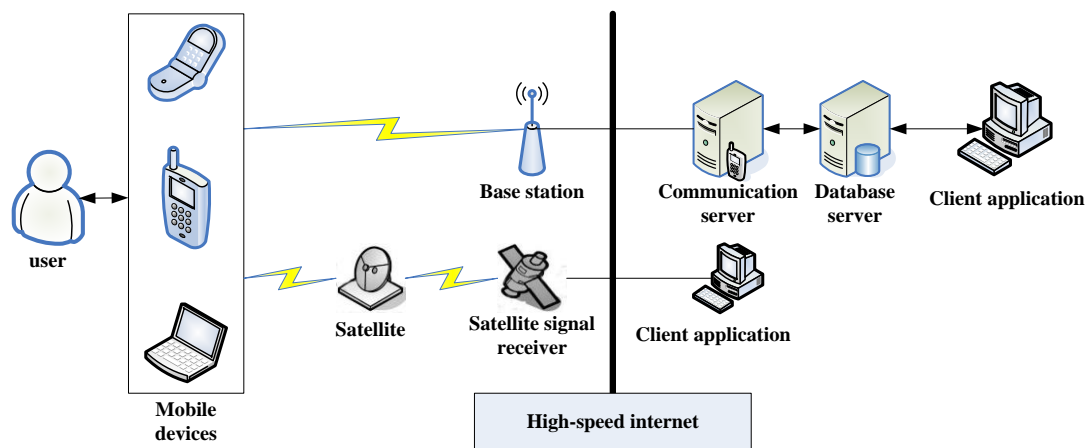


Figure 2. The main structure of mobile GIS.

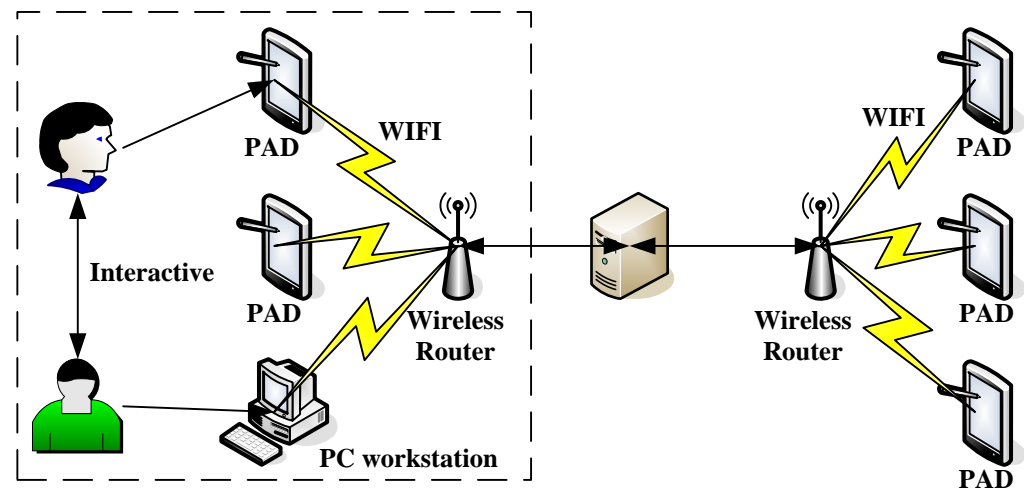


Figure 3. The system structure of GIS.

2.2. Spatial Differentiation Characteristics

The overall land-use planning is a long-term plan for the government to guide the rational use of land in order to prevent major mistakes in land use [9]. Because the process of economic and social development is long-term and gradual, it can be realized through long-term practice to adjust the land-use structure, layout, and direction according to the predetermined goals. Taking positive and effective measures to guide the spatial-temporal structure of urban land use to the ideal spatial pattern and realize the sustainable development of the city is an important problem faced by the urban land-use management departments [10].

Land resources are limited and non-renewable, but residents' demand for housing is infinite. Incomplete land planning in the early stage will naturally lead to an imbalance of residential allocation in the later stage. Residents have the right to choose their own residence. When choosing the geographical location of residence, they are bound to be affected by their own economic conditions and personal preferences. Choosing living space that can meet their various needs will lead to the phenomenon of "groups of people". People with similar economic income and similar personal preferences will gather together, and different qualities of living spaces will be isolated from each other [11]. The education park only has a university town, without relying on industries, and its location is remote. The number of buildings developed and constructed is very small, which cannot form spatial agglomeration. As a result, it does not conform to the development law of general cities [12]. Generally speaking, the rural area is a comprehensive space system, which includes residential space, industrial space, agricultural space, leisure space, and ecological

space. Therefore, this paper discusses the rural area function from three aspects: production function, life function, and ecological function. Rural regional functions not only have diversity but also have the characteristics of hierarchy and relativity. The same region has different research scales and the intensity of function is also different, so the evaluation index system is meaningful only on a certain regional scale.

2.3. Machine Learning

Find an optimal function $f(x, w_0)$ in a set of functions $\{f(x, w)\}$ to estimate the dependency relationship and minimize the expected risk.

$$R(w) = \int L(y, f(x, w)) dF(x, y) \quad (1)$$

Among them, $\{f(x, w)\}$ is called the prediction function set, and w is the generalized parameter of the function [13].

The prediction function can be defined as:

$$L(y, f(x, w)) = \begin{cases} 0, & \text{if } y = f(x, w) \\ 1, & \text{if } y \neq f(x, w) \end{cases} \quad (2)$$

The loss function can be defined as:

$$L(y, f(x, w)) = (y - f(x, w))^2 \quad (3)$$

Assuming that the probability distribution is uniform, that is, using sample data to define empirical risk:

$$R_{emp}(w) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i, w)) \quad (4)$$

The expression of the Lagrange function is as follows:

$$L(w, b, a) = \frac{1}{2}(w \cdot w) - \sum_{i=1}^n a_i(y_i[(w \cdot x_i) + b] - 1) \quad (5)$$

The maximum value of the function is:

$$Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j (x_i \cdot x_j) \quad (6)$$

If a_i^* is the optimal solution, then:

$$w^* = \sum_{i=1}^n a_i^* y_i x_i \quad (7)$$

The expression of the optimal classification function is:

$$f(x) = \text{sgn}\{(w^* \cdot x) + b^*\} = \text{sgn}\left\{\sum_{i=1}^n a_i^* y_i (x_i \cdot x) + b^*\right\} \quad (8)$$

The expression of the radial basis function is as follows:

$$k(x, y) = \exp\left(-\frac{|x - y|^2}{256\sigma^2}\right) \quad (9)$$

Among them, σ represents variance.

The calculation method of the influence score of the central town is carried out by the exponential decay method. The specific calculation is as follows:

$$f_{im} = M_m^{1-r_{im}} (m = 1, 2, 3) \quad (10)$$

In the formula, f_{im} is the influence degree score of the m -th central town of the i -th grading unit [14].

The standard deviation ellipse can be expressed as:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (11)$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (12)$$

Hypothesis: \bar{x}_i and \bar{y}_i are the difference between the xy coordinates and the average center. The accuracy of spatial differentiation is calculated as follows:

$$A = \left(\sum_{i=1}^n \bar{x}_i^2 - \sum_{i=1}^n \bar{y}_i^2 \right) \quad (13)$$

$$B = \sqrt{\left(\sum_{i=1}^n \bar{x}_i^2 - \sum_{i=1}^n \bar{y}_i^2 \right)^2 + 4 \left(\sum_{i=1}^n \bar{x}_i \bar{y}_i \right)} \quad (14)$$

3. Experiments on Rural Spatial Differentiation Characteristics

3.1. Data Sources

In the experiment, we mainly observe the degree of agreement between the NAES's similarity index results for categorical variables and the common sense judgment results [15,16]. This paper studies 21 township units in Yongtai County. The original data mainly come from the Yongtai County statistical yearbook and China County (city) social and economic statistical yearbook provided by the Yongtai County Bureau of Statistics; the land data and other vector data are mainly provided by the land and Resources Bureau of Yongtai County, including land-use change database data, the third land survey data (2006–2020), DEM data, and administrative boundary vector data; tourism statistics are provided by Yongtai County Tourism Bureau; part of the index data is composed of the indirect data obtained from the calculation and analysis of the original data and other data.

3.2. Data Preprocessing

This paper uses ENVI remote sensing image processing software to perform geometric correction on Landsat-8 remote sensing data. Band fusion is used to allow images to provide more data for research and to improve the accuracy of land classification results. With the help of ArcGIS 10.2, we perform unsupervised classification for interpretation, combined with the corresponding land-use vector data layer to view the size of the error and to correct the pattern of rural settlements [17]. We use the datasets and add a method to bind data from a variety of different data sources to MapX so that external data sources and the map can be connected. This solves the problem of integrating spatial data and attribute data in spatial data mining. In the development example, it is necessary to associate the attribute data of layers such as customers, cities, and states with the corresponding spatial objects, and perform unified analysis and processing [18].

3.3. Determination of Evaluation Indicators

The rural spatial index system of Yongtai County is shown in Table 1. This paper constructs a three-level framework of the eco-environmental assessment index structure. The first level index consists of structure, function, and coordination. The second level index is composed of several factors according to the selection principle of the evaluation index. The third level index is to select several factors under the second level index to form the whole evaluation index system. The most advanced comprehensive index (level 0) is the ecological environment comprehensive index (ECI), which is used to evaluate the ecological degree of rural areas. There is an antagonistic relationship between the three functions, that is, there must be a short board in the process of their rising development, such as the protection of the ecological environment, which limits the development of the production function of rural areas to a certain extent. If we vigorously develop production, the level of economic development will be improved, the ecological environment will be damaged, and the living function of rural areas will be reduced [19,20].

Table 1. Rural spatial index system of Yongtai County.

Target Layer	Domain Layer	Criterion Layer	Index Layer
Rural dungeon function (A)	Production function (B1)	Agricultural production function (C1)	Cultivated land area per capita (D1)
			Per capita food production (D2)
			Per capita non-food crops (D3)
			Grain yield (D4)
			Reclamation Index (D5)
		Economic development function (C2)	Industrial output value of township enterprises (D6)
			Income of township and village enterprises per capita (D7)
			Per capita agricultural output value (D8)
			Non-agricultural employment ratio of rural population (D9)
			Per capita net income of rural residents (D10)
	Life function (B2)	Life support function (C3)	Basic education attraction index (D11)
			Traffic accessibility (D12)
		Leisure tourism function (C4)	Land resource carrying capacity (D13)
			Richness of tourism resources (D14)
	Ecological function (B3)	Ecological conservation function (C5)	Forest coverage (D15)
			Total regional ecological service value (D16)
			Ecological service value per land (D17)
			Wetland area ratio (D18)
			Biological abundance index (D19)
			Terrain slope (D20)

3.4. Spatial Analysis Modeling

In this paper, a three-stage evolutionary limit learning machine evaluation method based on VMD-GWO-ELM is used to simulate the model. The method is divided into three stages: in the first stage, the variational mode decomposition is used to decompose the non-stationary time series information to obtain more convenient input variables for model learning, so as to solve the non-stationary problem of time series; in the second stage, the gray wolf algorithm is used to optimize the parameter values of the limit learning machine

model, obtain the optimal weight and bias of the input layer and the hidden layer, and solve the adaptive problem of model parameters; in the third stage, the improved limit learning machine is used to classify and predict the target problems [21,22].

3.5. Feature Pattern Classification

When the model is first trained, the role of the training data is mainly the first point, to extract the feature points of the model, but insufficient training data may cause errors in the feature point extraction process. Such errors will inevitably have a greater impact on the results [23,24]. Then the model comparison results are relabeled, and the labeling information is used for the retraining of the model. At this time, the second role of the training data is reflected. For the similar legend symbols, the similarity calculated by the model is higher. The low legend is marked as a positive data pair. When the model is retrained, extra weight is given when extracting feature points to improve its similarity [25,26]. In the same way, legends that are not similar pairs of legends but are incorrectly matched are marked as reverse data pairs. When the model is retrained, extra weight is given when extracting feature points to reduce its similarity [27,28].

3.6. Rural Regional Function Evaluation

The spatial single function and comprehensive function reflect the strength of the single function and comprehensive function, respectively, in the regional rural regional function, which can reflect the difference between the single function of a town compared with other towns in space [29,30]. It is a comprehensive index of comprehensive functions of rural areas. This paper uses the weighted sum model to obtain the single function index and the comprehensive function index of rural areas. In order to ensure the comparability of the two periods of data in time, this paper combines the data from 2013 and 2015 to determine the maximum and minimum values of each index, uses the extreme normalization method to standardize the two periods of data, and then uses the weighted sum method to obtain the function index of each rural regional function.

3.7. Rural Regional Function Orientation

According to the classification results of the production function, life function, and ecological function, this paper constructs the conceptual model of rural regional functional zoning. According to the evaluation results of the rural regional function, the production function (x), life function (y), and ecological function (z) are divided into four levels: strong, strong, weak, and weak, respectively, corresponding to grade I, II, III, and IV functional areas in the classification standard, and represented by three-dimensional spatial coordinates, which are weak comprehensive functional area, single functional dominant area, dual functional area, and strong comprehensive functional area.

4. Analysis of the Characteristics of Regional Rural Spatial Differentiation

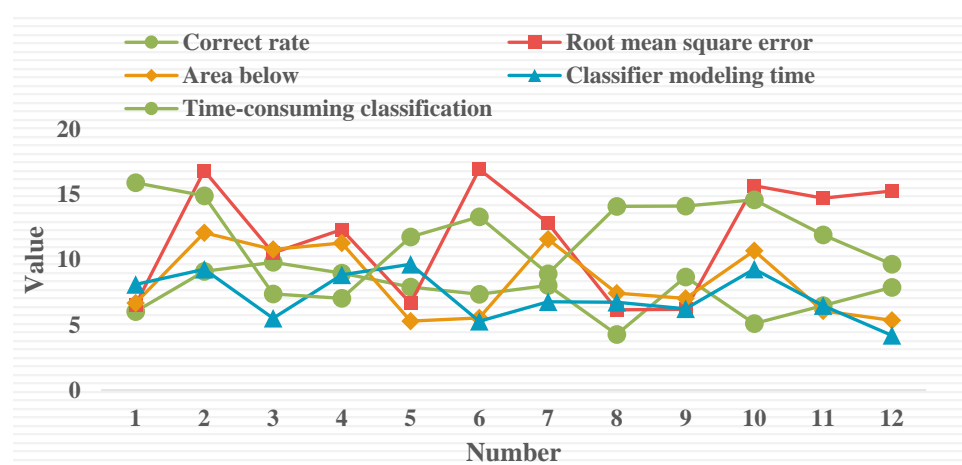
4.1. GIS Statistical Analysis

In the experiment, each algorithm is run five times with 10-fold cross-validation on the same dataset, which is equivalent to executing the algorithm 50 times, taking the average of the correct rate as the final result and finding the AUC by using the average of the first category for comparison. The experimental results are shown in Table 2. AUC (area under the curve) is a model evaluation index, ACC (accuracy) refers to accuracy, DBN (deep belief network) refers to deep belief network, and NBC (naive Bayes classifier) refers to the naive Bayes classification algorithm. The experimental results show that for most datasets, the joint operation not only improves the accuracy rate, but also increases the AUC value and comprehensively improves the performance of the classifier. The joint operation actually plays a role similar to the kernel function in SVM here, that is, combining two attributes into a new attribute that can provide more information that is beneficial to classification.

Table 2. Experimental results.

AUC		ACC	
DNB	NBC	DNB	NBC
0.7397	0.7352	85.45	83.97
0.9392	0.7676	65.00	54.50
0.9658	0.9719	95.61	99.25
0.8664	0.8569	97.82	98.14
0.5239	0.5067	56.81	56.03
0.6485	0.0887	68.89	53.70

The running results of the NBC on the selected datasets are shown in Figure 4. Five of the datasets did not find any deletable attributes, so in this round of experiments, only twenty-five datasets were used. The experiment tested a total of five indicators: accuracy, root mean square error, the area under the receiver operating characteristic curve, classifier modeling time, and classification time. Among them, the higher the correct rate, the better; the root mean square error represents the degree of dispersion of the classification results, and the smaller the error, the more stable the classification results; the ROC curve has been used in previous experiments, and the evaluation criteria for the classification results are more accurate than just focusing on accuracy. The rate is more comprehensive, and the larger the value, the better. For the final modeling time and classification time, the smaller the value, the better.

**Figure 4.** The results of the NBC running on the selected dataset.

The descriptive statistical results of PAHs in the subsidence area of Yongtai County are shown in Table 3. It can be seen from the table that the content of PAHs in the soil of the subsidence area ranges from 6.81 ng/gdw to 408.79 ng/gdw. The coefficient of variation of PAH content in the two subsidence areas was more than 1.0, which indicates strong variation. Although the conventional statistical analysis of PAH content in subsidence area can summarize the whole picture and overall characteristics of soil PAH content, it cannot reflect its local variation characteristics, that is, it can only reflect the whole sample to a certain extent, but cannot quantitatively describe the randomness, structure, independence, and correlation of soil PAH content. From the perspective of spatial distribution characteristics, the overall distribution characteristics are strong in the north and weak in the south. Specifically, grade I functional areas of agricultural production are mainly distributed in northeast villages and towns, including Hongxing Township, Baiyun Township, and Danyun Township. There are four townships in the grade II agricultural production functional area, which are mainly located in the northwest and central towns. The grade

III agricultural production functional area includes eight townships, mainly located in the central and eastern townships of Yongtai County. The fourth-grade functional areas of agricultural production are mainly distributed in the towns of central and southern China, including Wutong Town, Chek Tin Township, Camphor Town, and Chengfeng Town. Grade I agricultural production functional areas are mainly distributed in Hongxing Township, Baiyun Township, and Danyun Township. The second-grade agricultural production functional area includes eight townships, mainly located in the north of Yongtai County. The third-grade agricultural production functional area includes six townships, mainly located in the central and western regions of Yongtai County. The fourth-grade functional areas of agricultural production are mainly distributed in the towns of central and southern China, including Wutong Town, Chek Tin Township, camphor town, and Chengfeng town. Therefore, the northern township of Yongtai County has a stronger agricultural production function, while the southern township is relatively weak in agricultural production development. From the perspective of temporal variation characteristics, the overall spatial differentiation has little change, but the differentiation pattern has changed. The average agricultural production function index in Yongtai County is 0.010714 and the median is 0.0099720; the average in 2015 is 0.010715 and the median is 0.010721, which indicates that the agricultural production function index in Yongtai County is rising over time. The coefficient of variation of the agricultural production function index in Yongtai County is 0.273, and the coefficient of variation in 2015 is 0.243, which indicates that spatial differentiation is more obvious.

Table 3. Descriptive statistics of PAHs in subsidence soil of Yongtai County.

Number of Samples	Standard Deviation	Max	Minimum	Median	Average Value	Coefficient of Variation (%)
48	112.6358	408.79	6.81	31.2715	74.8134	1.506
44	1725.006	5911.57	106.33	363.35	1025.476	1.682

The cluster analysis is based on the proximity of the geomorphic window in the spatial coordinate system for classification. This paper uses the hierarchical clustering method to classify the geomorphic types of the study area. First, the geomorphology of the study area is divided into five categories. The statistical values of various geomorphic factors are shown in Table 4 and Figure 5. ESTDmean refers to the average of the standard deviation of elevation, SMEANmean refers to the average of the slope, SSTDmean refers to the average of the standard deviation of slope, and ARmean refers to the average of the area rate. On the basis of the above classification, karst landforms can be further divided into subcategories to highlight the spatial differentiation characteristics of karst landforms under the influence of multiple factors. Among them, landform types I, II, and III are the most widely distributed in the study area. This type of landform is the product of the late development of early karst cycles. It is well preserved due to the influence of water flow traceability and erosion. It is in the form of remnant mound troughs, gentle mound troughs, karst basins, and other landforms, and the standard deviation of elevation is 18.27 m. The average slope is 11.79° , the slope standard deviation is 9.47, and the area ratio is 1.05. These parameters all indicate that this type of karst landform has the characteristics of wide and gentle wavy karst planing surfaces, its groundwater level is shallow, and the Quaternary loess residual. The thickness of the material is large, the development of the landform is dominated by the lateral dissolution of the current, and it is continuously leveled. However, due to the decline of the erosion benchmark, there are sinkholes and funnels on the edge, and the landform type transitions to the Type II landform. From the secondary cluster analysis of the geomorphic factors of this type of karst landform unit, it can be further divided into Type I-1, remnant mound troughs and basins; Type I-2, gentle mound troughs and karst depressions; and Type I-3. In terms of space, the differentiation of Type I landforms is not only affected by the deep cut of the lateral valley as the base

level of regional excretion, but also by the lateral valley as the base level of local excretion. Therefore, from when the upstream watershed of the karst system transitions to the regional drainage base level and the local drainage base level, the Type I landform transitions from the Type I-1 landform to the Type I-3 landform. The degree of adaptation of the karst development in the descending area and strengthening of the local erosion base level shows the greater the depth of the karst development.

Table 4. Topographic factors of terrain types.

Category	ESTDmean	SMEANmean	SSTDmean	ARmean	Area Ratio (%)
I	18.27	11.79	9.47	1.05	30.34
II	33.16	20.263	9.59	1.10	52.39
III	67.39	23.24	10.39	1.14	15.43
IV	75.46	25.27	8.26	1.46	1.55
V	19.60	20.32	18.96	1.24	0.29

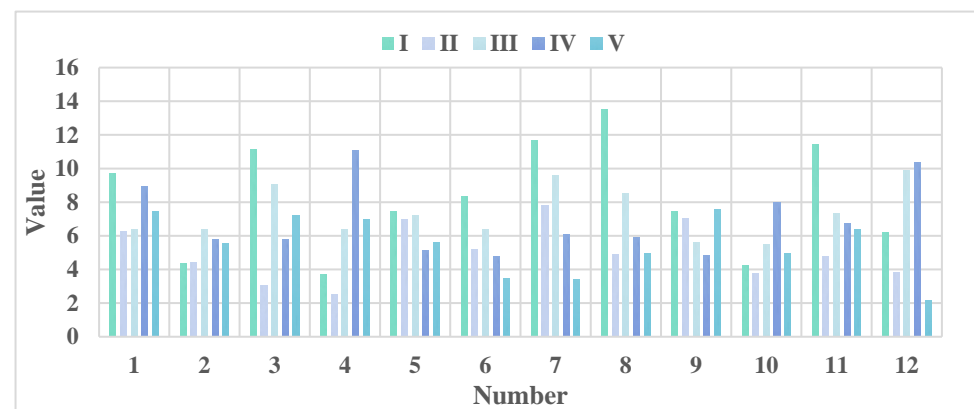


Figure 5. Topographic factors of terrain types.

4.2. Spatial Differentiation Feature Analysis

Because factor ecological analysis is an analysis method in the category of induction, it requires multivariate statistical analysis, and the selection of variables must be universal and relatively comprehensive. In the research of urban social space, most scholars choose urban demographic data as social factors. The research on rural social space emphasizes the network space of daily practice behavior. Therefore, the selected influencing factors involve population, residence, employment, schooling, etc., while taking into account the spatial attributes of rural social space, the spatial distance and topography are considered. The addition of factors can better reflect that rural social space is a spatial manifestation of rural residents' daily life practice. The correlation between the soil erodibility K value and organic matter content is shown in Figure 6. The formation and change of rural settlements are often affected by multiple factors. The impact of various indicators on the settlement distribution is also different, so it is necessary to grade and assign weights to the indicators and explore the significant degree of correlation between the impact factors and settlements, as well as between different impact factors. From the perspective of temporal variation characteristics, the overall spatial differentiation has little change, but the differentiation pattern has changed. The K value of the agricultural production function index of each township is 10.714, with an average of 20.715, indicating that the agricultural production function index of each township is rising over time. From the perspective of spatial distribution, it is strong in the middle and low on both sides. Specifically, the coefficient of variation of the rural life guarantee function in Yongtai County is greater than 0.5, which indicates that its spatial differentiation is obvious. The first-level functional areas of the life guarantee function include Zhangcheng town and Chengfeng town. The

second-level functional areas of the life support function include Dayang Town, Tongan town, and Wutong Town, all located in the central part of Yongtai County. The level III functional area of the life security function includes five townships, which are mainly located in the periphery of level I and II functional areas. The level IV functional area of the life security function includes 11 townships, with the proportion as high as Yongtai County, and the number of townships is generally above. From the perspective of time change, the ecological pattern of Yongtai County is relatively stable, and the change in the functional index is not obvious.

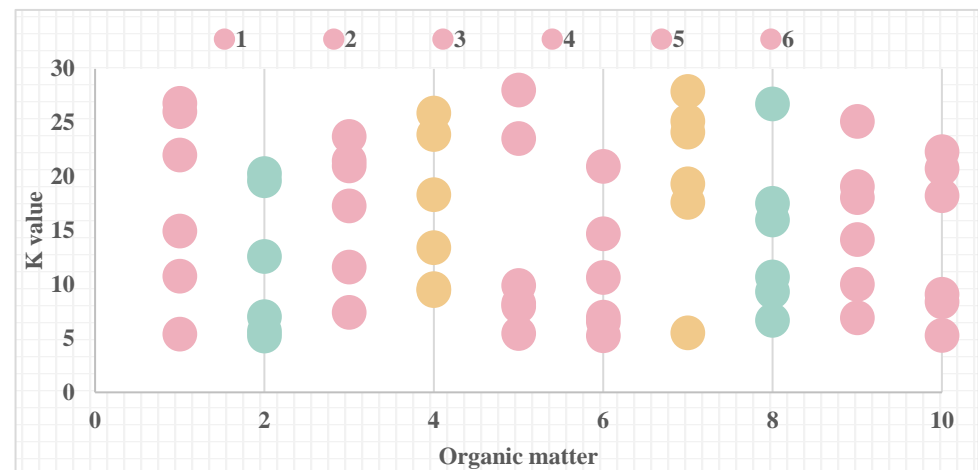


Figure 6. Correlation between the soil erodibility K value and organic matter content.

The coefficient of variation of the content of soil texture components of the main soil types is shown in Table 5. Through the scattered point trend prediction regression analysis, the relationship between the K value and the sand content is analyzed to obtain an exponential relationship distribution; the relationship between the K value and the silt content is analyzed to obtain a power function relationship between the K value and the silt grain; the relationship between the K value and N1 content is analyzed; and the relationship between K value and N1 tends to be a power function.

Table 5. Coefficient of variation of soil texture component content.

Soil	Coefficient of Variation (%)			
	<0.002 mm	0.05–0.002 mm	0.1–0.05 mm	2–0.05 mm
Black soil	68.19	26.85	46.15	26.94
Chernozem	74.79	28.40	46.24	16.89
Albic soil	16.59	13.23	34.75	13.26
Dark brown earth	77.82	38.30	36.12	25.69

The test results of the GWO-SVM model are compared with the results of the PSO-SVM, GA-SVM, and ABC-SVM models. The classification results are shown in Table 6 and Figure 7. SVM (support vector machine) refers to support vector machine, GA-SVM refers to genetic algorithm optimization support vector machine, PSO-SVM refers to particle swarm optimization support vector machine, ABC-SVM refers to intelligent bee colony optimization support vector machine, and GWO-SVM refers to gray wolf optimization support vector machine. The results show that GWO-SVM has good classification performance. Its cross-validation rate reaches 91.66%, and the recognition rate of test samples reaches 82.41%. Both the cross-validation rate and the recognition rate are the best among several comparative models. It shows that the use of GWO-SVM has better evaluation accuracy

for the evaluation of small sample datasets. When the sample distance is greater than the range, the samples become completely independent.

Table 6. Average classification results.

	SVM	GA-SVM	PSO-SVM	ABC-SVM	GWO-SVM
C	9117	5242.26	915.12	315.41	941.38
σ	0.261	0.316	57.06	0.26	0.265
Cross-validation rate	64.66%	86.2%	87.93%	89.65%	91.66%
Recognition rate	40.68	80.34	80.91	81.52	82.41
C	554.99	981.0869	5863.0957	422.0964	1011.7
σ	57.2	359.2969	26.0763	45.9438	31.5
Cross-validation rate	65.23	74.84	75.43	75.82	81.35
Recognition rate	64.9805	70.8171	73.8132	74.9805	76.97

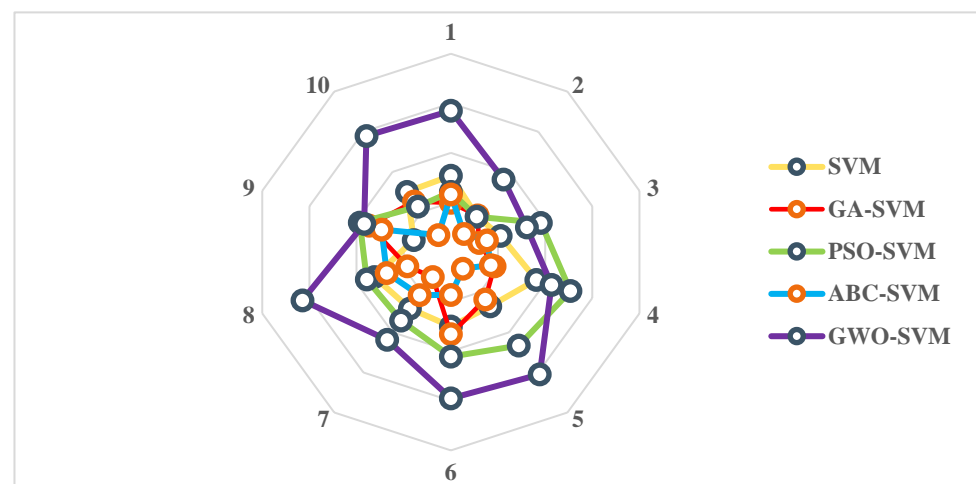


Figure 7. Average classification results.

The change in the landscape index of rural settlement patterns is shown in Figure 8. From the perspective of time change, the growth trend of rural life security function in Yongtai County is obvious, and the spatial differentiation characteristics have obvious changes. The average and median life security function in 2013 were less than that in 2015, and the number of villages and towns in grade I and II functional areas in 2015 was significantly more than that in 2013, indicating the overall improvement of life security function in Yongtai County in 2015. From the perspective of spatial distribution pattern change, although the number of each functional grade area has changed at the two evaluation time points, the relative strength relationship between the township functional index has not changed significantly. The analysis of per capita income is helpful to understand the degree of wealth and poverty in a region. From the source of income, a higher degree of industrial agglomeration and a more developed industrial and agricultural system will bring higher income, which can reflect the level of urbanization in a region; the level of urbanization is closely related to the distribution characteristics of rural settlements. Logically, per capita income has a positive impact on the density of rural settlements. Nearly 75% of the rural settlements are distributed in gentle areas with slopes of less than 2° , and less than 2% of the rural settlements are distributed in areas with slopes greater than 10° . The impact of the slope on the spatial distribution of rural settlements is more obvious than that of altitude. Low-slope areas are conducive to the construction of residential areas and infrastructure, and agricultural production can be better carried out, so it is easier to form settlements.

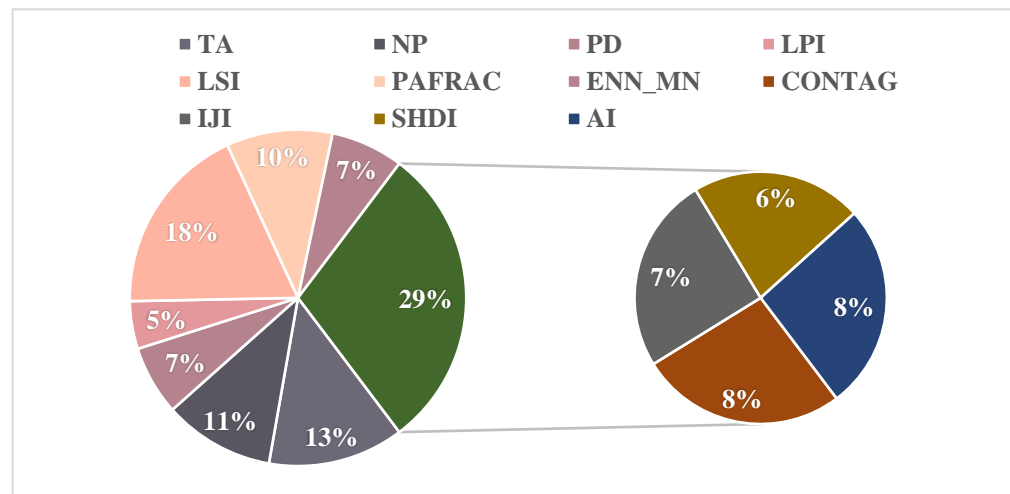


Figure 8. Changes in the landscape index of rural settlement patterns.

Non-normally distributed data will cause the semivariogram to produce a proportional effect, enlarge the impact of errors, raise the sill value and nugget value, and change the correlation of the spatial structure. Therefore, in order to eliminate the proportional effect, the Hg data of the soil surface layer after removing the specific value meets the normal distribution after logarithmic transformation, and the Hg data of the deep soil layer is also processed in the same way. In the case of data not belonging to a normal distribution, it is generally necessary to transform them. If the data are transformed and the transformed data still do not meet the conditions of normal distribution, Krüger interpolation should not be used. The impact of spatial relationship similarity and name similarity on data links is shown in Figure 9. It can be seen from the figure that when measuring the result of the data link, the utilization rate of spatial topological relationship similarity is obviously higher than other similarities. However, name similarity is more different in data linking, and it is more convenient to analyze whether the link is successful. Therefore, spatial topological relationship similarity and name similarity are more important in data linking and matching processing than category similarity.

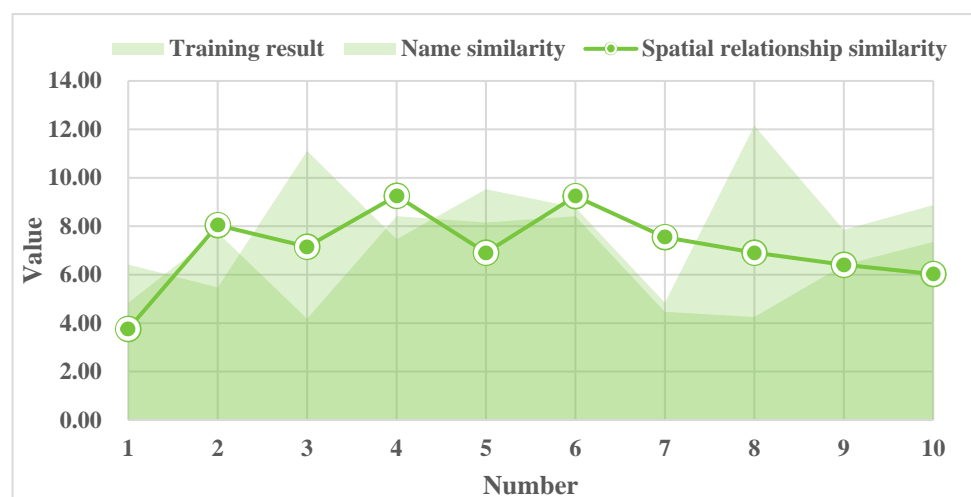


Figure 9. The impact of spatial relationship similarity and name similarity on data links.

The prediction accuracy evaluation of the model is shown in Table 7 and Figure 10. It can be inferred from the results that the prediction accuracy of the three prediction models is more than 88%, and the radial basis function neural network prediction model is better than the least squares support vector machine prediction model and the random forest

prediction model. From the prediction accuracy of different soil layers, the prediction effect of the least-squares support vector machine prediction model and radial basis function neural network prediction model in the 0~20 cm soil layer is better than that in the 20~40 cm soil layer, which may be caused by the more obvious effect of meteorological factors on surface soil and the heavy influence of meteorological factors on all prediction factors; however, the prediction effect of the random forest prediction model in the 20~40 cm soil layer is slightly better than that in the 0~20 cm soil layers. Through the accuracy analysis and comparison, it can be found that the radial basis function neural network prediction model has the best prediction effect in this study area.

Table 7. Model prediction accuracy evaluation.

Predictive Model	Soil Depth	R ²	RMSE	Forecast Accuracy (%)
RF	0~20 cm	0.408	10.830	88.030
	20~40 cm	0.443	10.842	88.451
LS-SVM	0~20 cm	0.811	6.202	92.981
	20~40 cm	0.787	6.787	92.703
RBFNN	0~20 cm	0.909	4.255	95.121
	20~40 cm	0.888	4.793	94.834

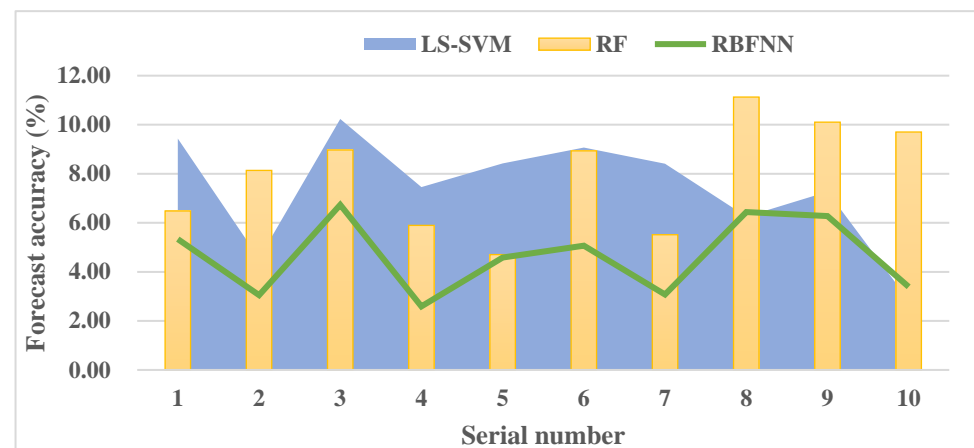


Figure 10. Model prediction accuracy evaluation.

The statistical values of the ecological functions of each township in Yongtai County are shown in Table 8. From the perspective of spatial characteristics, the rural areas are more functional in the central and northeast areas than in other areas. The coefficient of variation is less than 0.2 and that in 2015 is less than in 2013, which indicates that the spatial differentiation of comprehensive function in rural areas is less significant and has a decreasing trend. Specifically, in 2013, the first level functional zone includes Tangqian Township and Chengfeng town. The second level functional zone includes six towns. The third level functional zone includes five towns. The level IV functional zone includes eight towns. In 2015, the first-grade functional areas include seven townships, namely Tangtang Township, Chengfeng Town, Song Kou Town, Tongan Town, Wutong Town, Dayang Town, and Qing Liang Town. The second-level functional zone includes four towns. The third-level functional zone includes seven towns. The level IV functional zone consists of two towns. From the perspective of time change, the comprehensive function of rural areas has been enhanced as a whole, and the spatial differentiation pattern has changed. In 2015, the number of townships in the grade I functional zone was significantly higher than that in 2013, and that of the class IV functional zone was significantly lower than that in 2013. In

general, the multi-function of rural areas increased over time, but the growth of different towns was different.

Table 8. Statistical values of ecological functions of towns and towns in Yongtai County.

Statistical Indicators	Average	Median	Range	Standard Deviation	Coefficient of Variation
2013	0.022871	0.023443	0.015193	0.003761	0.164424
2015	0.024861	0.024295	0.011625	0.003149	0.126645

5. Conclusions

Yongtai County is generally high in rural areas, which indicates that the economic development and urbanization level of the area needs to be further improved. The government should pay more attention to the work and give policy support to the areas with strong rural characteristics, so as to continuously improve the urbanization level of these areas and realize the coordinated development among the towns in Yongtai County.

At the same time, the evaluation results of the rural regional function are basically consistent with the actual development of Yongtai County. In this paper, the production function, production function, and life function are taken as the three functional fields, and the agricultural production function, economic development function, life security function, and ecological conservation function are taken as the criteria. The agricultural production function index, economic development function index, production function index, life security function index, leisure tourism function index, life security function index, ecological function index, and rural regional comprehensive function index of each township in Yongtai County were calculated, and the grading evaluation was carried out from the two aspects of time (2013 and 2015) and space. Rural regional function zoning and rural regional leading function positioning are based on the advantageous functions of each township. However, in the process of rural development, each function promotes the other, but there are also contradictions and impulses, so there must be a “short board” hindering its development. Scientific rural regional function orientation is not only to determine the leading function and play its important role in rural development but also to identify obstacle functions, overcome the “stumbling block” in rural development, and accelerate the realization of rural revitalization.

According to the evaluation results of the rural regional function, this paper constructs the research framework of rural regional function positioning and completes the accurate positioning of the rural regional function. This research provides a theoretical basis for targeted rural development strategy research and promoting rural development and construction. The maximum value of the rural regional function in space cannot be simply determined as the dominant function. The maximum value of the rural regional function in space is an important basis to determine the leading function of rural regions, but simply taking the maximum value of a single function space as the leading function makes the result of function positioning one-sided. In short, in future research, we should further understand the deep characteristics of the culture of Yongzheng County from the perspective of spatial diversity, and use new methods to quantitatively identify the spatial distribution characteristics of place names. Protecting the local ecology is the basis of development, reflecting the folk style characteristics of the whole region and the deep historical culture charm.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Sun, Y.; Peng, M.; Zhou, Y.; Huang, Y.; Mao, S. Application of machine learning in wireless networks: Key techniques and open issues. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3072–3108. [[CrossRef](#)]
2. Dev, S.; Wen, B.; Lee, Y.H.; Winkler, S. Ground-Based Image Analysis: A Tutorial on Machine-Learning Techniques and Applications. *IEEE Geoenviron. Remote Sens. Mag.* **2016**, *4*, 79–93. [[CrossRef](#)]
3. Taherkhani, N.; Pierre, S. Centralized and Localized Data Congestion Control Strategy for Vehicular Ad Hoc Networks Using a Machine Learning Clustering Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3275–3285. [[CrossRef](#)]
4. Ruske, S.; Topping, D.O.; Foot, V.E.; Kaye, P.H.; Stanley, W.R.; Crawford, I.; Morse, A.P.; Gallagher, M.W. Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer. *Atmos. Meas. Tech.* **2017**, *10*, 695–708. [[CrossRef](#)]
5. Buczak, A.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2017**, *18*, 1153–1176. [[CrossRef](#)]
6. Helma, C.; Cramer, T.; Kramer, S.; De Raedt, L. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. *J. Chem. Inf. Comput.* **2018**, *35*, 1402–1411.
7. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [[CrossRef](#)]
8. Goswami, R.; Dufort, P.; Tartaglia, M.C.; Green, R.E.; Crawley, A.; Tator, C.H.; Wennberg, R.; Mikulis, D.J.; Keightley, M.; Davis, K.D. Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. *Brain Struct. Funct.* **2016**, *221*, 1911–1925. [[CrossRef](#)]
9. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365. [[CrossRef](#)]
10. Assouline, D.; Mohajeri, N.; Scartezzini, J.L. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Sol. Energy* **2017**, *141*, 278–296. [[CrossRef](#)]
11. Sacha, D.; Sedlmair, M.; Zhang, L.; Lee, J.A.; Peltonen, J.; Weiskopf, D.; North, S.C.; Keim, D.A. What You See Is What You Can Change: Human-Centered Machine Learning By Interactive Visualization. *Neurocomputing* **2017**, *268*, 164–175. [[CrossRef](#)]
12. Mullainathan, S.; Obermeyer, Z. Does Machine Learning Automate Moral Hazard and Error? *Am. Econ. Rev.* **2017**, *107*, 476–480. [[CrossRef](#)] [[PubMed](#)]
13. Patel, M.J.; Khalaf, A.; Aizenstein, H.J. Studying depression using imaging and machine learning methods. *Neuroimage Clin.* **2016**, *10*, 115–123. [[CrossRef](#)] [[PubMed](#)]
14. Hu, S.; O'Hagan, A.; Sweeney, J.; Ghahramani, M. A spatial machine learning model for analysing customers' lapse behaviour in life insurance. *Ann. Actuar. Sci.* **2021**, *15*, 367–393. [[CrossRef](#)]
15. Luo, X.; Liu, J.; Zhang, D.; Chang, X. A large-scale web QoS prediction scheme for the Industrial Internet of Things based on a kernel machine learning algorithm. *Comput. Netw.* **2016**, *101*, 81–89. [[CrossRef](#)]
16. Lemley, J.; Bazrafkan, S.; Corcoran, P. Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision. *IEEE Consum. Electron. Mag.* **2017**, *6*, 48–56. [[CrossRef](#)]
17. Van Ginneken, B. Fifty years of computer analysis in chest imaging: Rule-based, machine learning, deep learning. *Radiol. Phys. Technol.* **2017**, *10*, 23–32. [[CrossRef](#)]
18. Luo, G. Automatically explaining machine learning prediction results: A demonstration on type 2 diabetes risk prediction. *Health Inf. Syst. Syst.* **2016**, *4*, 1–9. [[CrossRef](#)]
19. Jamali, A.A.; Ferdousi, R.; Razzaghi, S.; Li, J.; Safdari, R.; Ebrahimie, E. DrugMiner: Comparative analysis of machine learning algorithms for prediction of potential druggable proteins. *Drug Discov. Today* **2016**, *21*, 718–724. [[CrossRef](#)]
20. Sui, H.; Li, L.; Zhu, X.; Chen, D.; Wu, G. Modeling the adsorption of PAH mixture in silica nanopores by molecular dynamic simulation combined with machine learning. *Am. J. Hematol.* **2016**, *85*, 1950–1959. [[CrossRef](#)]
21. Chou, J.S.; Ngo, N.T. Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Appl. Energy* **2016**, *177*, 751–770. [[CrossRef](#)]
22. Taylor, R.A.; Pare, J.R.; Venkatesh, A.K.; Mowafi, H.; Melnick, E.R.; Fleischman, W.; Hall, M.K. Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Acad. Emerg. Med.* **2016**, *23*, 269–278. [[CrossRef](#)]
23. Dickson, M.E.; Perry, G.L.W. Identifying the controls on coastal cliff landslides using machine-learning approaches. *Environ. Model. Softw.* **2016**, *76*, 117–127. [[CrossRef](#)]
24. Valdes, G.; Solberg, T.D.; Heskell, M.; Ungar, L.; Simone, C.B., 2nd. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. *Phys. Med. Biol.* **2016**, *61*, 6105–6120. [[CrossRef](#)]
25. Plawiak, P.; Sosnicki, T.; Niedzwiecki, M.; Tabor, Z.; Rzecki, K. Hand Body Language Gesture Recognition Based on Signals from Specialized Glove and Machine Learning Algorithms. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1104–1113. [[CrossRef](#)]
26. Azam, M.S.; Raihan, M.A.; Rana, H.K. An Experimental Study of Various Machine Learning Approaches in Heart Disease Prediction. *Int. J. Comput. Appl.* **2020**, *175*, 16–21.

27. Mosquera, R.; Castrillón, O.D.; Parra Osorio, L. Prediction of psychosocial risks in colombian teachers of public schools using machine learning techniques. *Inf. Technol.* **2018**, *29*, 267–281. [[CrossRef](#)]
28. Li, B.; Huang, J.; Feng, Y.; Wang, F.; Sang, J. A machine learning-based approach for improved orbit predictions of LEO space debris with sparse tracking data from a single station. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56.6*, 4253–4268. [[CrossRef](#)]
29. Mateo-Garcia, G.; Veitch-Michaelis, J.; Smith, L.; Oprea, S.V.; Schumann, G.; Gal, Y.; Baydin, A.G.; Backes, D. Towards global flood mapping onboard low cost satellites with machine learning. *Sci. Rep.* **2021**, *11*, 7249. [[CrossRef](#)]
30. Wang, C.; Platnick, S.; Meyer, K.; Zhang, Z.; Zhou, Y. A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations. *Atmos. Meas. Tech.* **2020**, *13*, 2257–2277. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.