Rural Landscape Design Data Analysis Based on Multi-Media, Multi-Dimensional Information Based on a Decision Tree Learning Algorithm

Ning Leng[1], Hongxin Wang[2]*
Guangdong Peizheng College, Guangdong, China[1]
University of Sanya, Hainan, China[2]*

***Abstract*—This paper analyzes and studies the design characteristics of multi-dimensional information rural scenes. For data mining and the Decision Tree (DT) calculation method, the pre-processing system and method of multi-dimensional information rural award design are put forward again. Through the analysis of the multi-dimensional value of multi-dimensional multimedia mountain villages, the form of planning and design analysis and corresponding methods are based on the analysis. Using one village as a case study, we were able to investigate the villagers, roads, services, greening, ecology, and other aspects of the village in complete detail and then implement the multi-media, multi-resource village's detailed planning and design.**

***Keywords*—Multi-media multi-dimensional information rural landscape; data mining; decision tree; data preprocessing**

I. Introduction

In recent years, with the rapid development of building automatic control systems and building management systems, which have been widely used (such as large-scale public multimedia multidimensional information rural scenery supervision and management platform and landscape measurement project successively conducted by several provinces and cities in China), extensive multimedia multidimensional information rural scenery design resources are loaded. These resources are the most direct data on the operation of buildings, systems, and equipment and are highly valued. Cities nationwide have been paying attention to the call to reform rural areas in recent years [1-3]. A loss of the original rural characteristics develops in some cities because of rural urbanization, which impacts the landscape by destroying old buildings and building high-rise structures. A decline in rural scenes has resulted from cultural pressures, resulting in a loss of rural characteristics. Destruction of ecosystems directly results from financial growth and poor resource management in rural areas [4-7]. *E.g.,* reducing land in rural fields and reducing natural spot areas. Rural ecosystems are damaged by the excessive use of chemicals such as pesticides, herbicides, and fertilizers, and the township businesses' rapid processing also increases pollution.

In order to obtain high-quality mining data, the results of the data to guide practical operation through the construction of a data pretreatment system, using the Decision Tree (DT) method to calculate the missing data of construction data, abnormal data, multi-dimensional data for data analysis, data pretreatment system and methods in the practical case, to obtain superior results.

II. Features of Big Data and Data Preprocessing System of Multi-Media Multi-Dimensional Information Rural Landscape (MMMDIRL) Design

It is already under study; the relevant research on the application of data mining in architecture focuses on regular mining using data mining algorithms. MMMDIRL analysis, landscape and load prediction, fault diagnosis, optimization of operation control, and other data pretreatment research and results are limited. This section will refer to the data preprocessing processes and methods in other fields (*e.g.,* big Internet data.) and propose the preprocessing system and methods for building data after analyzing the characteristics of multi-dimensional information rural landscape data in multimedia.

DT learning algorithm variables are represented by a triple array (r_1, r_2, r_3) $r_1 < r_2 < r_3$ consisting of zeros, whose dependent function is EQU (1).

$$\mu(x) = \begin{cases} \dfrac{x - r_1}{r_2 - r_1}, if \ r_1 \le x \le r_2 \\ \dfrac{x - r_3}{r_2 - r_3}, if \ r_2 \le x \le r_3 \\ 0, others \end{cases} \tag{1}$$

Set DT learning algorithm $\alpha = (a_1, a_2, a_3)$, $\beta = (b_1, b_2, b_3)$ according to the DT learning algorithm number of addition and extension principle of power, EQU (2).

---

[1]This research study is sponsored by these projects: project one: The Young Innovative Talents Project (Humanities and Social Sciences) of Guangdong Province, the project number is: 2019WQNCX127. Project two: Philosophy and Social Sciences "the 13th Five-Year Plan" Youth Project of Guangdong Province, the project number is: GD20YYS05. Project three: 2020-2021 academic year scientific research project of Guangdong Peizheng College; the project number is: pzxjzd02.Thank these projects for supporting this article!

$$\mu_{\bar{\alpha}+\bar{\beta}}(z)=\sup\left\{\min\left\{\mu_{\bar{\alpha}}(x),\mu_{\bar{\beta}}(y)\right\}\ \ z=x+y\right\}=$$
$$\begin{cases}\dfrac{z-(a_1+b_1)}{(a_2+b_2)-(a_1+b_1)}, a_1+b_1\le z\le a_2+b_2\\ \dfrac{z-(a_3+b_3)}{(a_2+b_2)-(a_3+b_3)}, a_2+b_2\le z\le a_3+b_3\\ 0, others\end{cases} \tag{2}$$

That is, the sum of the DT learning algorithm is the DT learning algorithm, and there are EQ. (3) and EQ. (4)

$$\bar{\alpha}+\bar{\beta}=(a_1+b_1,a_2+b_2,a_3+b_3) \tag{3}$$

From $\mu_{\lambda\bar{\alpha}}(z)=\sup\left\{\mu_{\bar{\alpha}}(x)\middle|z=\lambda x\right\}$

get

$$\lambda\bar{\alpha}=\begin{cases}(\lambda a_1,\lambda a_2,\lambda a_3),\lambda\ge 0\\ (\lambda a_4,\lambda a_3,\lambda a_2),\lambda<0\end{cases} \tag{4}$$

It $\bar{\alpha}_i=(a_{i1},a_{i2},a_{i3})\ i=1,2,\cdots,m$ is the DT learning algorithm, the non-negative linear combination and DT learning algorithm programming $\bar{\alpha}_i$ are obtained, EQ. (5)

$$\sum_{i=1}^{m}\lambda_i\bar{\alpha}_i,\lambda_i\ge 0 \tag{5}$$

It is still a DT learning algorithm, and EQ. (6)

$$\sum_{i=1}^{m}\lambda_i\bar{\alpha}_i=\left(\sum_{i=1}^{m}\lambda_i a_{i1},\sum_{i=1}^{m}\lambda_i a_{i3}\right) \tag{6}$$

The DT learning algorithm is a random plan with constraints, such as random parameters, and chance is used to show precisely how probable it is that constraints will be set. A constraint programming environment enables the use of possibilities as constraints [8-10]. Random DT learning algorithm and DT learning algorithm control plan are powerful tools for solving optimization problems with random parameters and DT learning algorithm parameters. Compared with other industries, due to the early construction of landscape supervision and management platforms, design points are designed and evaluated in advance in the process of scheme customization. At the same time as data redundancy, the pre-set data attributes and accuracy in the database of the landscape management platform are avoided as far as possible. Problems such as incomplete data accuracy and inconsistent format will occur, so data redundancy, format contradiction, and accuracy are problematic [11-13].

Based on the above analysis, this paper proposes a pre-processing system of MMMDIRL design big data based on the characteristics and corresponding processing methods of MMMDIRL design big data, as shown in Fig. 1.

A. *Significance of Rural Landscape Planning (RLP) and Design*

The purpose of RLP and design is to optimize the environment of human settlements. At the same time, it is the direction of the future development of the rural environment. It is the ideal structure of the rural ecosystem. The primary purpose of RLP is to coordinate rural landscape and ecological development.

To make the relationship between man and nature more harmonious, China's new socialist rural construction is still in its infancy. RLP is the exploration and attempt of rural construction mode and the key to determining the quality of rural residents' environmental construction.

The key point is that RLP aims to optimize the rural environment and meet local ecological needs. The functions of rural landscape construction are as follows:

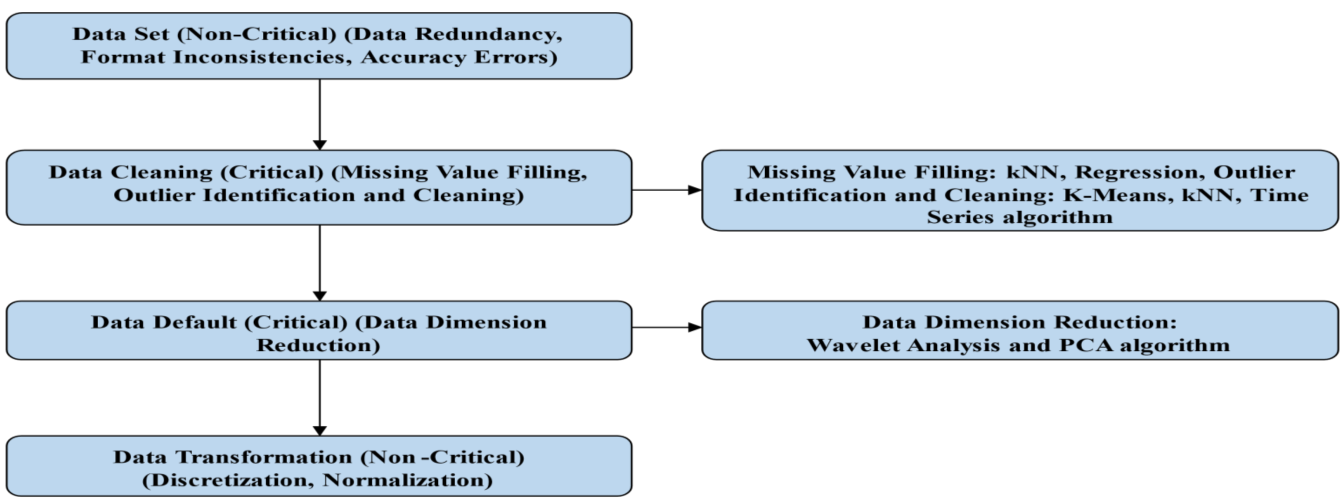

Fig. 1. Big data preprocessing technology system of MMMDIRL design.

B. Promoting Rural and Local Economic Development

RLP and design optimize and adjust the rural industrial structure, promoting the development of the local economy. The practice results in many rural areas show that while optimizing the industrial structure, the regional economy is multiplied, and the output of various crops will increase. Taking Lijiahe Township, Xuanen County, Hubei Province as an example, the adjustment of industrial structure has significantly increased the income of tobacco farmers and brought considerable tax revenue to the local finance. Another example is the "Farmhouse Fun" Ecological Sightseeing Park, which mainly displays local folk customs and promotes agricultural development by providing leisure and entertainment activities, which has extensively promoted the construction of local rural areas.

RLP focuses on the overall planning of the rural environment to realize the transformation from design to reality. RLP and design consider the balance between man and land, economic development, and environmental protection.

C. Conducive to the Inheritance and Development of National Culture

RLP is conducive to the layout of rural villages and the protection of local national culture. For example, local architectural concepts and literary works can inherit national cultural characteristics. The local architectural landscape is the carrier of national culture and fully reflects the local folk characteristics. Therefore, the rural landscape is the most essential way for people to get familiar with the traditional culture, protect the rural characteristics to the greatest extent and realize the inheritance of regional culture.

III. Building Data Preprocessing Technology Based on the DT Learning Algorithm

Because of the characteristics of multimedia multi-dimensional information rural landscape design big data, in the above mentioned, the identification of abnormal data is K-ManASKNN, such as time series and other algorithms. This paper uses only one specific method to pre-process the related data quality problems; fill in missing data, determine the DT learning algorithm, and clean outliers to reduce data volume, identify outliers, and determine the DT learning algorithm.

IV. Data Quality Analysis

Infrastructure built on multimedia landscape and electronic network maps is well-known and durable. 200 multimedia design landscapes that won professional recognition were used as a statistical sample. The formation mechanism of the landscape's ornamental nature was assumed as the classification benchmark. The sample of 100 landscape cases has 17 active multi-media landscapes; 17% of the total. While ornamental landscapes have 78 active and passive loads, passive loading in multimedia landscape architecture causes a subject of 6% of the overall. The prototype spatial interface pattern is controlled in multiple dimensions, primary visual data is gathered at eye level, and stable triangular composition is maintained to achieve 78% of the total. The absolute proportion of multimedia landscape architecture with a multi-dimensional dominant space interface is 85%. Data quality is assessed before mining in MMMDIRL designs, which frequently use plain ornamental subjects (Table I).

TABLE I. Shows the Actual Running Data of the Countryside with Multi-Media, Multi-Dimensional Information

S/N	*P*/kW	Q_{ch} /kW	T_{ci} /°C	T_{co} /°C	T_{ei} /°C	T_{eo} /°C
1	35.5	54.9	20.4	20.7	14.3	12.7
2	35.5	54.9	21.3	21.6	13.3	11.8
3	37.6	54.9	23.1	23.1	13.5	12
…	…	…	…	…	…	…
35378	97.8	504.7	28.5	33	16	10.2
35379	97.9	536.8	27.3	32.4	18.7	11.5
35380	99.5	536.8	27	31.8	21.8	14.8

A. Missing Data

The data from 35,380 groups are classified and statistically analyzed according to their attributes. Data of different attributes are missing to their extent. The degree of missing data attributes is shown in Table II. According to the analysis of data in the Table II, the missing data of the mechanical landscape attribute is the largest, with 576 missing data, accounting for 1.6%, and the loss of the remaining attribute data is small. Due to the integrity of data mining conclusions, these false data need to be filled in the pre-processing stage.

TABLE II. Distribution of Missing Data by Attribute

Missing Data Attribute	The Number of/a	Percentage of Missing Data /%
MMMDIRL (P)	574	1.67
Load (Q_{ch})	216	0.66
The outlet temperature of chilled water (T_{eo})	168	0.56
The inlet temperature of chilled water (T_{ei})	173	0.45
Cooling water outlet temperature (T_{co})	305	0.87
Cooling water inlet temperature (T_{ci})	71	0.34

B. Abnormal Data

Abnormal performance and data transmission cannot be avoided when devices such as multimedia multi-dimensional information villages and sensors are in extreme action. In data preprocessing, it is necessary to identify and clean these abnormal data.

C. Multidimensional Data

There are many design variables related to multi-dimensional information on rural landscapes. According to relevant theories and empirical models, the influencing factors of multimedia in multidimensional information rural manufacturing are attributed to five factors: load (Q_{ch}), the outlet temperature of cold water (T_{eo}), immersion temperature of cold water (T_{ei}), the outlet temperature of cold water (T_{ci}) and immersion temperature of cold water (T_{co}). There are several specific influencing factors. Whether the dimension of data continues to decline needs to be analyzed in the pre-processing stage.

At the same time, in the process of data sorting, there are fewer problems such as redundancy of points, repeatability of fields, lack of data accuracy, and format contradiction so that it is screened and processed in a simple, fast, and one-sided.

Given the above data quality analysis, this project must conduct data preprocessing, focusing on abnormal data and multidimensional data. It is the following primary research work.

V. Missing Data Filling based on the DT Learning

DT learning algorithm is a classification algorithm based on analogy learning, which learns by comparing the given check object data with similar training data [14-15]. The missing-value processing step of the DT learning algorithm compares a sample data set with labels to determine the correlation between the data and classification. The features of the sample set are compared to the newly entered unlabeled data, and the classification labels that correspond to the most similar features are extracted. This approach ensures that the data is correctly classified. Generally speaking, only the k most similar data in the first part of the data set are selected, and the most common classification among the k most similar data is selected as the new data classification to realize the filling of missing values.

A simple data analysis of Table I shows that the original running data has a continuous data missing phenomenon. As shown in Table III, see lines 34145-34149 for missing data. The missing attribute is the multi-dimensional information rural landscape (P) value, while the other attributes are complete. The kNN method populates the data to get the missing attribute values.

The classic process of taking advantage of kNN: First, have exclusive properties. (*P*, Q_{ch}, T_{ci}, T_{co}, T_{ei}, T_{eo}) For each missing data, select k=3, and finally close to ''. Distance quote: The known data values are populated as unknown attributes. Get all unknown data values. Through kNN program processing, the values of the unknown data P in the data in lines 34145-34149 are shown below.

TABLE III. Missing Data in the Actual Running Data

S/N	P/kW	Q_{ch} /kW	T_{ci} /℃	T_{co} /℃	T_{ei} /℃	T_{eo} /℃
34145	-	515.1	24.3	28	17.1	12.4
34146	-	483.0	29.8	33.7	13.4	9
34147	-	515.1	23.1	28	16.8	12.1
34148	-	496.7	24.5	33.7	13.3	8.8
34149	-	494.7	27.6	33.7	13.6	9.1

The DT learning algorithm is set for similar processing for other missing values in the original data.

VI. Abnormal Data Identification and Cleaning Based on the DT Learning Algorithm

DT learning algorithm analysis, the essence of the principle is to collect similar things, not similar things, into distinct kinds of processes. Data set 'D' contains 'n' objects, generated clusters 'K', and the clustering algorithm divides them into clusters EQ. (7) and EQ. (8).

$$k\left(k \le n\right) \; C_1,\cdots,C_k \; 1 \le i, j \le k \; C_i \subset D \tag{7}$$

$$C_i \cap C_j = \varnothing \tag{8}$$

The DT learning algorithm identifies data outliers by assuming that objects belong to large, dense, small, sparse, or none of these clusters. An overview of the above ideas as a general method of outgroup point (outlier) identification is as follows.

As illustrated in Fig. 2, a data object is regarded as an outlier if it does not fit into any of the clusters that have been created.

Fig. 2. a is abnormal data.

Outliers are depicted in Fig. 3 and occur when a data object is located far from the center of the cluster to which it belongs.

Fig. 3. a and b are abnormal data.

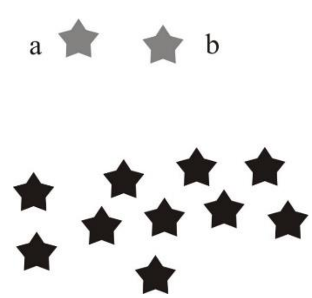

Fig. 4. a and b are abnormal data.

As seen in Fig. 4, if a data object is included in a minor or sparse cluster, then every object included in that cluster is considered an outlier.

In general, if the abnormal data does not belong to any cluster, as shown in Fig. 2, or belongs to a smaller cluster than the others, as shown in Fig. 4, the cluster data is deleted and ignored (even if the deletion does not affect the data sample capacity). As shown in Fig. 3, abnormal data must be cleaned correctly if it is far from the cluster center. In this paper, based on the DT learning algorithm, the cluster subset is clustered, and the abnormal data is replaced with the central value of the subset group to which the abnormal data belongs.

In order to study the relationship between multi-dimensional information about rural landscapes and a single variable, it is necessary to classify the data in detail. When the rural load rate of multimedia multidimensional information is fixed (66%) and the immersion temperature of cooling water (28.3℃), the rural landscape of multimedia multidimensional information constitutes a changing 2D array. The total number of data groups is 176, as shown in Table IV.

The DT learning algorithm was applied to perform cluster analysis on the above data, and the number of categories k=4 of the DT learning algorithm was selected to achieve the clustering results, as shown in Fig. 5.

Based on the above abnormal data identification principle, the cluster results in Fig. 5 are analyzed, and it is found that there are abnormal data in Fig. 5, which requires sub-cluster data cleaning and processing.

Take the cluster in the upper left corner of Fig. 5. The data set comprises 56 groups of data objects and data objects. (7.6,58.4), (7.5,58.4) are abnormal data (separated from the center of each cluster), and the tree learning algorithm of the cluster object is re-determined to take the number of cluster categories k=4. If cluster results, data object (7.5,58.3), (7.6,58.5) and other owning cluster center is (8.2,58.3), and cluster center (8.1,59.1) replaces (7.5,58.5), (7.6,58.4) and other abnormal data, abnormal data is cleaned.

Similarly, the data in the lower left, upper right, and right corners of Fig. 5 is cleaned using a similar determination tree learning algorithm. Fig. 6 shows the results of re-clustering the cleaned data. Comparing Fig. 6 vs. Fig. 5, it is found that the data distribution is centralized, which is a good clustering result.

TABLE IV. SHOWS THE ACTUAL RUNNING DATA OF THE COUNTRYSIDE WITH MULTI-DIMENSIONAL MULTIMEDIA INFORMATION

S/N	P/kW	Q_{ch} /kW	T_{ci} /°C	T_{co} /°C
1	59.5	354.6	29.5	8.7
2	60.6	354.7	29.6	8.8
5	59.7	354.7	29.6	8.9
175	60.8	354.7	29.6	11.2
176	61.1	354.7	29.6	11.2

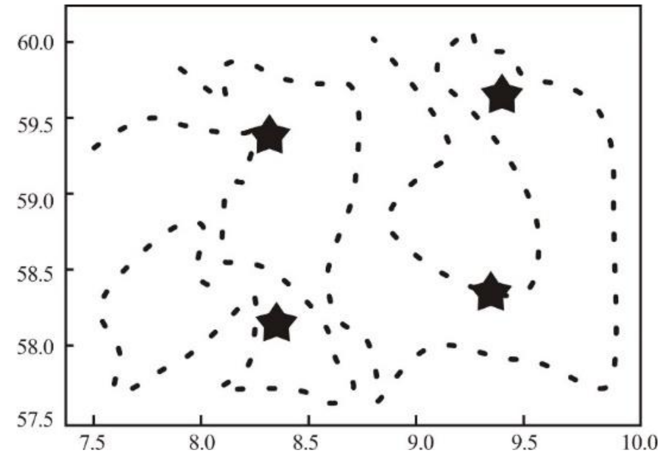

Fig. 5. Results of DT learning algorithm for raw data.

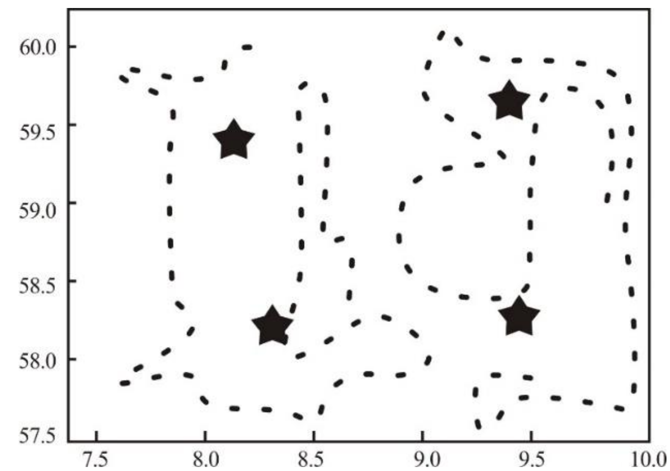

Fig. 6. Re-clustering results after cleaning the abnormal data in Fig. 5.

VII. DATA DIMENSION REDUCTION BASED ON THE DT LEARNING ALGORITHM

Principal Component Analysis (PCA) is a method of multiple statistical analysis that uses a linear transformation to reduce the number of different indexes to a small number of comprehensive indexes uncorrelated. By analyzing the characteristics of the matrix, the original data is projected in linear space to reduce the dimension of the data. The central principle is based on each component's cumulative dispersion contributions to determine the principal component.

The core process of principal component analysis is as follows: (a) The covariance matrix of the eigen-centralization matrix. (b) Computing the eigenvector associated with the covariance matrix's eigenvalue is necessary. (c) Calculate the dispersion contribution rate of each component. (d) Calculate the cumulative dispersion contribution rate.

In the core process, primary element analysis accumulates dimensional scattered data contribution rate to a manually set limit. This data influence represents the entire data set and creates high-order relegation data.

The most original problem of regression research is to study the relationship between multi-dimensional information rural landscape (P) and load (Q_{ch}), the outlet temperature of cold water (T_{eo}), immersion temperature of cold water (T_{ei}), the outlet temperature of cold water (T_{co}), immersion temperature of cold water (T_{ci}) and other factors. From the perspective of variables, it is related to six-dimensional variables. If we dig down all six variables, there may be some insignificant factors. It will do useless work and will waste many computing costs. This section used the PCA method to analyze six-dimensional variables to obtain the main influencing factors. The results of the DT learning algorithm are shown in Table IV.

The results of the DT learning algorithm in Table V are analyzed, showing that the cumulative dispersion contribution rate of the upper four variables reaches 99.93%. Since the dispersion contribution rates of the first four components and the last two components show a significant difference in the number of digits, the first four components are the main components, and the six-dimensional data in this data set is maintained at four-dimensional data. Through the follow-up data, freezing waters found that there is a direct relationship between the multi-dimensional information rural landscape (P) and load (Q_{ch}), the outlet temperature of cold water (T_{eo}), and the immersion temperature of cold water (T_{ci}), and the corresponding mathematical relationship is obtained. The rationality and applicability of data degradation of the DT learning algorithm have been proved here.

TABLE V. RATE OF VARIANCE CONTRIBUTION AND CUMULATIVE RATE OF VARIANCE CONTRIBUTION

S. No. of Composition	Rate of Variance Contribution (%)	Rate of Cumulative Variance Contribution (%)
1	83.951	82.851
2	17.758	99.309
3	2.014	100.023
4	1.47	100.193
5	1.304	100.197
6	1.303	100

VIII. CONCLUSION

Given the characteristics of multi-dimensional information data in rural landscape design, the general steps and methods of data preprocessing applied in some other fields cannot be directly applied to the big data preprocessing of buildings. The data preprocessing system of multi-dimensional information rural landscape design suitable for multimedia and the corresponding data quality problem processing methods are presented.

DATA AVAILABILITY

On request, the corresponding author will provide access to the data used to support the findings of this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING STATEMENT

This research study is sponsored by these projects: Project one: The Young Innovative Talents Project (Humanities and Social Sciences) of Guangdong Province, the project number is 2019WQNCX127. Project two: Philosophy and Social Sciences "the 13th Five-Year Plan" Youth Project of Guangdong Province, the project number is GD20YYS05. Project three: 2020-2021 academic year scientific research project of Guangdong Peizheng College; the project number is pzxjzd02.Thank these projects for supporting this article!

REFERENCES

[1] K. Rajala, M. G. Sorice & V. A. Thomas, “The meaning(s) of place: identifying the structure of sense of place across asocial–ecological landscape”, People and Nature, vol. 32, no. 9, pp. 310-316, 2020.

[2] A. A. Chibilev & V. P. Petrishchev et al., “The soil-ecological index as an integral indicator for the optimization of the land-use structure”, Geography and Natural Resources, vol. 37, no. 4, pp. 348–354, 2016.

[3] Q. Zhang, S. Wang, & K. Jian, “Design of campus wetland landscapes: a case study of the bailuxi wetland of University of Sanya in Sanya, Hainan, China”, Journal of Landscape Research, vol. 55, no. 11, pp. 393-399.

[4] Z. Wang, “Study on the multimedia application in college aerobics teaching: a learning interactive perspective”, Revista De La Facultad De Ingenieria, vol. 32, no. 2, pp. 759-767, 2017.

[5] J. Kim, “Subdivision design and landscape structure: case study of the woodlands, texas, us”, Urban Forestry & Urban Greening, vol. 38, pp. 232-241.

[6] Y. Chen, “Coastal mountain landscape and urban plant planning based on remote sensing imaging”, Arabian Journal of Geosciences, vol. 14, no. 8, pp. 1-15, 2021.

[7] Costanza & Robert, “Whole watershed health and restoration: applying the Patuxent and GWYNN Falls landscape models to designing a sustainable balance between humans and the rest of nature”, Journal of Landscape Research, vol. 8, no. 2, pp. 333-344, 2015.

[8] Y. U. Ziping, “Landscape design of the Ming Dynasty outer city wall in Nanjing from the perspective of experience tourism”, Journal of Landscape Research, vol. 10, no. 6, pp. 137-139, 2018.

[9] Corlett & T. Richard, “The role of rewilding in landscape design for conservation”, Current Landscape Ecology Reports, vol. 1, no. 2, pp. 127-133, 2016.

[10] R. Oppermann, E. Aguirre, R. Bleil, J. D. Calabuig & A. Schraml, “A rapid method for monitoring landscape structure and ecological value in European farmlands: the Lisa approach”, Landscape Online, vol. 90, pp. 1-24, 2021.

[11] M. Lin, T. Lin, L. Jones, X. Liu & X. Lu, “Quantitatively assessing ecological stress of urbanization on natural ecosystems by using a landscape-adjacency index”, Remote Sensing, vol. 13, no. 7, pp. 727-733, 2021.

[12] L. Z. Bian, “Analysis principles of landscape design small city”, Applied Mechanics and Materials, pp. 744-746, 2015.

[13] M. B. Lott, “Collective memory as an aesthetic landscape: the costume and scenic design of William Shakespeare’s as you like it”, Dissertations & Theses – Gradworks, vol. 8, no. 1, pp. 11-15, 2015.

[14] L. Valetti, A. Pellegrino & C. Aghemo, “Cultural landscape: Towards the design of a nocturnal lightscape”, Journal of Cultural Heritage, vol. 12, no. 21, pp. 200-203, 2019.

[15] Woudstra & Jan, “Designing the garden of GEDDES: the master gardener and the profession of landscape architecture”, Landscape & Urban Planning, vol. 2, no. 1, pp. 1-7, 2018.