International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-13

https://doi.org/10.70917/ijcisim-2026-1825
Article

# Exploration of the Ecological Wisdom in the Production Techniques of Chaozhou Dancong Tea within the Framework for Protecting Intangible Cultural Heritage of Agriculture

**Min Zhang[1,*] and Junlang Huang[1]**

[1] Institute for Advanced Studies in Humanities and Social Sciences, Beijing Normal University, Zhuhai, Guangdong, 519087, China
* Correspondence author: zhangmin@bnu.edu.cn

**Abstract:** Chaozhou single-compass tea has high production techniques, distinctive product characteristics and good industrial benefits. The study provides a basis for the ecological wisdom mining of Teozhou single-conglomerate tea by sorting out the characteristics of Teozhou single-conglomerate tea industry. The data mining method of tea production skills mainly contains association rules and clustering mining methods. The association rule adopts Apriori algorithm to mine the temporal and causal associations among tea leaves, and find out the item set that satisfies the minimum confidence level as the output of tea association rule. The tea clustering method is optimized on the basis of Density-Based Clustering (DBSCAN) algorithm from spatio-temporal data pooling, density thresholding, K-mean nearest neighbor method, and mathematical expectation method. The results showed that the confidence level of relatively low minimum temperature → large occurrence of pest class = 72.3%, and the confidence level of very high rainfall → medium occurrence of pest class = 98.9%. The largest coefficient of variation among the agronomic traits of Teochew single bush tea varieties was 100 bud weight, with a coefficient of variation of 29.34%. The highest correlation between bud length and width ratio and leaf spreading angle was found, with a correlation coefficient of 0.964. The clustering of Teochew single bush tea was divided into three categories, each of which contained five, four, and three tea tree varieties, respectively. The ecological excavation of Chaozhou single-cong tea provides a cognitive framework for the protection of agricultural non-cultural heritage.

**Keywords:** chaozhou monoculture tea; association rules; DBSCAN; Apriori algorithm; ecological wisdom mining

## 1. Introduction

Recently, the official website of the Ministry of Agriculture and Rural Development released the third batch of China's globally important agricultural cultural heritage reserve list, and the Chaozhou monoculture tea system in Guangdong Province was successfully selected [1]. It is understood that the globally important agricultural cultural heritage by the Food and Agriculture Organization of the United Nations (FAO) initiated and is responsible for identifying, is the rural areas and its environment under the long-term synergistic evolution and dynamic adaptation of the formation of the unique land use system and agricultural landscape [2]. The third batch of China's globally important agricultural heritage reserve list, a total of 30 traditional agricultural systems were selected for this list. Chaozhou monocotyledon tea culture system in Chaozhou City, Phoenix Market as the core heritage site, has a long history of more than a thousand years, is an important carrier of the inheritance of agricultural civilization in Guangdong Province, is still producing an important role in agricultural production, the development of the Phoenix monocotyledon tea leading agricultural industries, the birth of Phoenix monocotyledon tea "Guangdong" brand of agricultural products, the gold-lettered sign [3-4]. This time, the success of the Chaozhou monocotyledon tea culture system in the list shows that in the global perspective, the Chaozhou monocotyledon tea culture system has its unique inheritance value and cultural significance.

Regarding the definition of agricultural cultural heritage, the literature [5] considers that agricultural cultural heritage is a composite heritage that combines the characteristics of natural heritage, cultural heritage and intangible cultural heritage, and is a typical socio-economic-natural composite ecosystem composed of economic, biological, technological, cultural and landscape elements. Literature [6] points out that agricultural cultural heritage, as a collective public good closely related to agricultural practices, encompasses objects, places, and areas affected by agricultural activities, as well as experience-based knowledge of work, resource use and management. In Norway, agriculture is seen both as a threat to cultural heritage and as a major guardian of cultural heritage, and the policy agenda emphasizes the cultural heritage guardianship responsibilities of the agricultural sector. Literature [7] analyzes the current status of agro-cultural heritage management in China, Japan, and South Korea, and finds that Japan has a systematic inventory of cultural assets, but needs to clarify how to integrate information and data to promote the revitalization of cultural heritage; South Korea is actively exploring paths such as agricultural storytelling tourism and "Gil" tourism to respond to the gradual change of local identity; China has the longest history of cultural heritage in East Asia; and China has the longest history of cultural heritage in East Asia. China has the longest history of practicing agricultural cultural heritage in East Asia, but the process of commercialization has led to new issues such as changes in cultural identity. Literature [8] points out that the spatial distribution of agro-cultural heritage is a product of natural, economic, cultural and policy interactions, and that future conservation and management should focus on regional differentiation strategies and cross-sectoral synergistic governance.

With regard to the study of agro-nonheritage ecosystems, the literature [9] emphasizes the concept of "dynamic conservation", adapting to modern needs through sustainable innovations rather than solidifying heritage systems as static heritage, which provides a new management paradigm for the adaptation and sustainability of agro-cultural heritage in the context of global change. Literature [10] states that agro-cultural heritage aims to preserve traditional agricultural systems of global significance, and that these composite systems of human-earth harmony play a key role in guaranteeing the sustainability of family farming, the maintenance of biodiversity and cultural heritage. Literature [11] incorporates biodiversity conservation, cultural heritage maintenance, and ecosystem service enhancement into an integrated management program to transform traditional agricultural systems, from being under pressure of intensification or at risk of abandonment, into resilient agroecosystems for sustainable development. Literature [12] constructed a research paradigm for realizing the value of ecosystem products of agro-cultural heritage (EPAHS) based on the trinity of "characterization-value transformation pathway-appraisal method", and constructed a two-dimensional appraisal system including total value and structural evaluation indexes to quantify the composite value and transformation effect. Literature [13] pointed out that agricultural cultural heritage is the core support for the development of eco-agriculture, and it has unique advantages in providing ecosystem services, producing characteristic eco-agricultural products, and protecting biodiversity, and its value identification and integrated protection are the key tools for the development of characteristic eco-industries in eco-regions, and for the coordination of the relationship between eco-protection and economic growth. Literature [14] suggests that the ecological wisdom of traditional Chinese farming culture is an intrinsic driving force to promote the sustainable and high-quality development of the heritage system, and that the key strategies include the construction of eco-museums and other productive conservation pathways to realize the living heritage and adaptive use of the heritage system. Literature [15] takes the globally important agricultural cultural heritage "rice-fish symbiosis system in Qingtian, Zhejiang Province" as an object, and finds that dynamic conservation under the framework of agricultural NRM not only does not destroy the traditional agro-ecological structure, but also synergistically promotes biodiversity conservation and sustainable development of communities by enhancing the function of ecosystem services, rural infrastructure and environmental quality. Instead, it synergistically promotes biodiversity conservation and sustainable community development by enhancing ecosystem services, rural infrastructure and environmental quality.

Big data alone marks the arrival of an era characterized not only by the pursuit of abundant material resources and convenient information services brought by the ubiquitous Internet, but also by the value discovery and value transformation that distinguishes it from material data resources, as well as the brand-new phenomena in spirituality and culture brought by big data [16-17]. Therefore, the wisdom mining of agricultural cultural heritage in the era of big data not only requires the physical mining of traditional cultural heritage, but also generates a new field. Literature [18] points out that big data technology has significant potential for application in the identification of heritage value, precision of protection strategy and adaptive use; through big data mining methods, the spatial distribution, historical value and risk mapping of industrial cultural heritage can be effectively identified, realizing the paradigm shift from “passive salvage” to “preventive protection”. Literature [19] describes a cultural heritage information system based on big data technology-SCRABS (Smart Contextual Browsing Assistant for Cultural Heritage Environment), which realizes the rich extraction, intelligent presentation and immersive interaction of cultural heritage information through big data architecture to enhance the function of cultural interpretation and participation in education in digital environment. Literature [20] adopts multi-dimensional big data mining methods to analyze the spatial and temporal distribution characteristics of cultural heritage, and analyzes the continuity and discontinuity of cultural relics in the time sequence through the continuity judgment method to reveal the “life cycle” of different cultural relics. For the analysis of public opinion and content understanding in cultural heritage scenarios, literature [21] proposes a viewpoint mining method based on NoSQL database and Apache Spark streaming processing architecture; through a large number of user comments and opinions on social media, the content of cultural heritage is efficiently characterized and accurately refined by using streaming big data technology. To briefly summarize the wisdom mining of cultural heritage, literature [22] proposes an Internet big data mining method based on IoT intelligent devices, and integrates the blockchain model to guarantee the data security and knowledge trustworthiness, which realizes the whole chain of trusted non-heritage data from “collection” to “activation”. The whole chain of non-heritage data from “collection” to “revitalization” is realized. Literature [23] takes “user-generated content” as the knowledge source of cultural IP revitalization, and realizes the innovative mode from “heritage record” to “intelligent dissemination and intellectual property derivation” through cross-modal semantic computation and emotional feedback mechanism. Through cross-modal semantic computing and emotional feedback mechanism, it realizes the innovative mode from “heritage record” to “intelligent dissemination and intellectual property derivation”, which provides a technological integration solution for the digital transformation and dynamic dissemination of traditional rural culture.

In recent years, the Chaozhou tea industry has shown a continuous expansion of the overall development scale and a gradual optimization of the industrial layout structure. In this paper, firstly, we extract the characteristics of Chaozhou tea from the industrial production area and the amount of tea production to provide a data basis for the characterization of Chaozhou tea. Secondly, the association rule mining algorithm is proposed for the linkage of tea production techniques, the main environmental factors of tea pests and diseases are set, and the changes of different characteristic environmental factors are mined. Then, the traditional density-based clustering algorithm is improved, pooling idea and parameter optimization strategy are applied to improve the clustering method, and the K-mean nearest neighbor method and mathematical expectation method are combined to realize the automated clustering threshold setting. Finally, the relevant data of Teochew single bush tea is collected for empirical research, and the tea production technology is mined.

## 2. Ecological Wisdom Mining Methods of Teochew Monocotyledon Tea

### *2.1. Characteristics of Teochew Single-Cluster Tea Industry*

The main production areas of Chaozhou monocotyledon tea are concentrated in Raoping County, Chaoan District and Xiangqiao District, among which the tea plantations in Raoping County and Chaoan District have an area of more than 4,500 $hm^2$.Since 2019, with the launching of the Provincial Modern Agricultural Industrial Park Construction Project in Guangdong Province as an opportunity, Chaozhou monocotyledon tea has continued to optimize its industrial layout, and has gradually formed two major core production areas dominated by Fenghuang monocotyledon (Chaoan District) and Lingtou monocotyledon (Raoping County).

Chaoan District is the hometown of Chinese oolong tea (famous tea), the main tea-producing towns are Phoenix Town, Dengtang Town, etc., and the species of tea planted are mainly Phoenix monocotyledon. Phoenix monocotyledon origin - Phoenix Town, the tea planting area of more than 4566 $hm^2$, with an annual output of more than 4900 tons of tea, annual output value of more than 980 million yuan. Fenghuang Town has more than one thousand ancient tea trees of Daan ancient tea tree tea plantation and Wu Adong Village 10,000 acres of contiguous ancient tea tree resource protection zone, the number of huge, concentrated distribution area is rare in the country, is unique in Guangdong, the world's rare cultivated monocotyledonous ancient tea tree resource treasury. 2014, Chaoan, Guangdong, phoenix monocotyledonous tea culture system was selected as one of the second batch of

China's important agricultural cultural heritage, filling the Guangdong Province heritage protection in the field of agriculture blank.

Raoping County is the hometown of China Lingtou Monocotyledon Tea, and the main tea-producing towns are Fubin Town, Xintang Town, Jianrao Town, etc. The varieties of tea planted are mainly Lingtou Monocotyledon Tea (White Leaf Monocotyledon Tea). Lingtou tea is famous for its unique honey rhyme style, and is a cultural tea with unique honey rhyme. 2020 Guangdong Meteorological Society awarded Lingtou tea the honorary title of "Chaozhou Lingtou tea - Lingnan Ecological and Climatic Excellence".

## *2.2. Association Rule Mining Methods for Tea Production Techniques*

Association rules are a method in data mining algorithms to find hidden association relationships between data items in data records [24]. Analysis of association rules reveals interdependencies and correlations between sets of data items. Generally formed association rules are implicit formulas shaped like $X \to Y$, where $X$ is the prior of the association rule and $Y$ is the successor, indicating that the attribute value of $X$ is associated with the attribute value of $Y$. Correlation can be further categorized into temporal correlation, causal correlation, etc. We commonly use correlation analysis mainly to measure the degree of correlation between two attributes, including gray correlation analysis also measures the degree of correlation between the comparison sequence and the reference sequence. The association rule model, on the other hand, can perform multidimensional association analysis, describing the laws and patterns that occur between attributes or between multiple attributes.

The transaction set is composed of a number of itemsets, and the probability of containing both $A$ and $B$ items in transaction set $D$ is called support. And the probability of containing $A$ items while containing $B$ items is called confidence. So support reflects the frequency of occurrence of the rule and confidence reflects the strength of the rule. Association rules that satisfy the minimum support threshold and minimum confidence threshold are meaningful. For the minimum support threshold can be determined by expert experience or after experimentation for the actual situation of the dataset. This is because if the threshold is set too low, the rules are not general, but if the threshold is set too high, the rules with more practical value will be missed. The main algorithms for association rules are: Apriori algorithm, FP-tree algorithm.Apriori algorithm needs to scan the original dataset for many times, while FP-tree algorithm is based on Apriori using tree structure without scanning it for many times to generate the frequent sets, and the algorithm is efficient. But Apriori algorithm has better scalability and is the most commonly used algorithm.

Suppose that there is a transaction set $D$, and $A$ and $B$ are subsets of the set of transaction items:

Support:

$$\sup port\left(A \Rightarrow B\right) = P\left(A \cup B\right) \tag{1}$$

Confidence:

$$confidence\left(A \Rightarrow B\right) = P\left(B \middle| A\right) \tag{2}$$

(1) Apriori algorithm

For the finite set $I = \{i_1, i_2, i_3, \cdots\cdots, i_n\}$, we call each element $i_n$ as an item, and the item set containing $n$ items is $n$ item set, the support of the item set is greater than the minimum support is called frequent item set.The basic idea of Apriori algorithm is: scan the dataset to search for candidate 1 frequent set, remove the item set that does not satisfy the minimum support, and then scan the dataset again after connecting to the frequent 1 item set, and remove the item set that is lower than the minimum support to get the frequent 2 item set. The frequent 2-item set is obtained by removing the itemsets lower than the minimum support. And so on, iteratively until no frequent $K+1$ itemsets can be found. The association rule is to find out the item set that satisfies the minimum confidence level from the frequent set as the output of the rule.

The final frequent 3-item set is generated when it can no longer be connected as $K+1$-item.

So the main process of association rule generation based on Apriori algorithm is:

1) Input dataset $D$ with minimum support $\min\sup port$ and minimum confidence $\min confidence$.

2) Scan the dataset $D$ and take all the original data as 1-item set.

3) Mining frequent $K(K \geq 1)$-item sets. Scan dataset $D$, calculate the support of the $K$-item set, and remove the item sets with support less than the minimum support to get the frequent $K$-item set.

Connect the frequent $K$ itemsets to generate $K+1$ candidate frequent sets.

4) Repeat 3) until no frequent $K+1$ itemsets can be found.

5) For each frequent itemset $K$, generate all its non-empty subsets.

6) Calculate the confidence level of each non-empty subset $c$ and generate rule $c \rightarrow (K-c)$ if the confidence level is greater than the minimum confidence threshold.

(2) Validation of association rules

The association rules are generated under certain support and confidence evaluation, so a part of uninteresting rules have been avoided from the support and confidence evaluation system. How to evaluate whether the mined association rules are useful and satisfy the objective needs of users requires an effective evaluation system from two aspects.

1) Evaluation by objective evaluation method

Support and confidence on the effectiveness of the association rules in the objective evaluation, but has not taken into account the statistical correlation between the statistics, so there will be a large number of actually independent of each other's attributes constitute the rules, and some are even negatively correlated, and these rules will seriously mislead the user's judgment and decision-making. Therefore, the degree of support and confidence based on the addition of the degree of enhancement as a three-dimensional evaluation system. In statistics, if the two transactions $A$, $B$ are independent of each other, the occurrence of $A$ transactions will not affect the $B$ transactions, that is, in the case of the two transactions are independent of the correlation rule is meaningless.

The degree of lift is the determination of the correlation between the transactions, which is how much the occurrence of transaction $A$ changes the probability of occurrence of transaction $B$. Suppose that for any two transactions $A$ and $B$, we know that if there is a $P(AB)=P(A)P(B)$, then $A$ and $B$ are independent of each other, so there is:

Degree of enhancement:

$$list(A,B)=\frac{P(A\cup B)}{P(A)P(B)} \tag{3}$$

If $list(A,B)=1$, then $A$ and $B$ are independent and the association rule is meaningless; if $list(A,B)>1$, then $A$ and $B$ are positively correlated; if $list(A,B)<1$, then $A$ and $B$ are negatively correlated.

2) Evaluation by subjective methods

On the basis of the objective evaluation of the effectiveness of the rules, combined with the user's needs and interests and then subjective evaluation. Testing with new or test data is effective, potentially useful, and novel.

## *2.3. Adaptive Cluster Mining Approach for Tea Making Techniques*

The density-based clustering (DBSCAN) algorithm is the most typical among density clustering algorithms [25]. It is able to group the higher density regions in the spatio-temporal dataset into one data set, and can ignore the influence of unclassified noise points on the clustering results, these data sets can be any kind of shapes, and a set is usually called a cluster. The idea of clustering is simply to give the maximum distance radius of the cluster and the minimum number of points needed to form a cluster, and then all the points that can be reached under the maximum density are connected and divided into the same category; figuratively speaking, it is to draw a circle, in which two parameters should be defined, one of them is the radius of the circle, and the other one is that the circle must contain at least a few points, and finally, all the points in a circle are attributed to a cluster.

### 2.3.1. Optimization strategies

Based on the pooling idea and parameter optimization strategy, the spatio-temporal dataset is divided into multiple equidistant grid cells with equal-sized pooling windows, and feature extraction is performed for each grid cell, which combines to become a distinctive spatio-temporal dataset; then, according to the distribution characteristics of the spatio-temporal dataset data, the Eps candidate is obtained by the K-mean nearest neighbor method and mathematical expectation method, and the corresponding Eps and MinPts candidate are obtained by the mathematical expectation method. Then, according to the data distribution characteristics of the spatio-temporal data set, the Eps candidate set is obtained by K-mean nearest neighbor method and mathematical expectation method, and the corresponding MinPts candidate set is obtained by mathematical expectation method, based on which the Eps and MinPts corresponding to the spatio-temporal data set clustering number tends to be stabilized are selected as the optimal thresholds, and spatio-temporal density clustering is completed.

(1) Spatio-temporal dataset pooling

Pooling is mainly used in convolutional neural network (CNN) for image processing, with the deep development of neural network, pooling technology is also more and more used in other fields. As the saying goes, “if you can't see the mountain, you can't see the mountain”, pooling is “simplify from complexity”, pooling can filter out the key information among the miscellaneous information, which can simplify the complexity of clustering operation. Although pooling reduces the features of the input spatio-temporal data set, it still maintains the invariance of the spatio-temporal data set's features such as rotation, translation and expansion.

(2) Density threshold

Since the DBSCAN clustering algorithm is to set the data set threshold by the two parameters Eps and MinPts, the parameter Density threshold Density is introduced to realize the calculation of the threshold. The concept of density threshold Density is as follows: Eps is used as the radius to draw a circle (ball), and the number of data objects contained in this circle (ball) is MinPts.

For two-dimensional data there:

$$Density = \frac{MinPts}{\pi \cdot Eps^2} \tag{4}$$

For 3D data there is:

$$Density = \frac{MinPts}{\frac{4}{3}\pi \cdot Eps^3} \tag{5}$$

For four-dimensional data there is:

$$Density = \frac{MinPts}{\frac{1}{2}\pi^2 \cdot Eps^4} \tag{6}$$

For five-dimensional data there is:

$$Density = \frac{MinPts}{\frac{8}{15}\pi^2 \cdot Eps^5} \tag{7}$$

By analogy, the calculation of density threshold can be extended to $N$ -dimensional spatio-temporal data. In this paper, the object of study is 3-dimensional spatio-temporal data, so Equation (5) is used for calculation.

A suitable density threshold can achieve the desired clustering effect. If the density threshold is too large, the same clustering cluster will be decomposed into several different small clusters; conversely, if the density threshold is too small, the different clusters will be grouped into the same large cluster. It is found that, in general, the number of clustering clusters converges with the decrease of the density threshold, and eventually stabilizes between a certain range, so assuming that the number of clustering clusters is correct, there are the following judgments: the smaller the density threshold is, the smaller the number of noise points are, the higher the accuracy of the clustering results, and the better the clustering effect is.

(3) K-Average Nearest Neighbor Method

The K-Average Nearest Neighbor (K-ANN) method is an extension of the Nearest Neighbor (KNN) method [26]. Firstly, the distances between all data objects in dataset $D$ and the $K$ nd data object which is closest to it are calculated, and then the average is taken, and finally the K-Average Nearest Neighbor Distance of dataset $D$ is derived by calculating the K-averaged Nearest Neighbor Method.The computational procedure of the K-Average Nearest Neighbor Method is as follows:

In the first step, the distance distribution matrix of data set $D$ is calculated with the following formula:

$$D_{n\times n} = \left\{ Dist(i,j) \middle| 1 \le i \le n, 1 \le j \le n \right\} \tag{8}$$

where $D_{n\times n}$ is the real symmetric matrix of $n \times n$ ; $n$ is the number of data objects owned by dataset $D$ ; and $Dist(i,j)$ is the distance from the $i$ th data object to the $j$ th data object in dataset $D$ .

In the second step, each row of elements in the ascending sort matrix $D_{n\times n}$, where the distance vector $D_0$ composed of the 1st column elements represents the distance from the data object to itself, all of which are represented by 0. The elements of column $K$ form the K-nearest neighbor distance vector $D_K$ for all data objects.

In the third step, the average value of the elements in the vector $D_K$ is taken and the K-mean nearest neighbor distance $\overline{D_K}$ of the vector $D_K$ is calculated and taken as a candidate parameter for Eps. Finally, all the $K$ values are calculated to produce the Eps parameter list $D_{Eps}$ with the following formula:

$$D_{Eps}=\left\{\overline{D_K}\,\middle|\,1\le K\le n\right\} \tag{9}$$

(4) Mathematical expectation method

Mathematical expectation method is one of the most commonly used methods in probability theory and statistics. According to the calculated list of Eps parameters, firstly, the number of data objects within the corresponding Eps range is calculated for each Eps parameter, based on which the mathematical expectation of the number of all the data objects is found, which is taken as the neighborhood density threshold MinPts parameter of the dataset $D$, and the formula is as follows:

$$MinPts=\frac{1}{n}\sum\nolimits_{i=1}^{n}P_i \tag{10}$$

where $P_i$ is the number of data objects in the range of its Eps for the $i$ nd data object, and $n$ is the total number of data objects in data set $D$.

Through the K-mean nearest neighbor method and mathematical expectation method to calculate the Eps and MinPts parameter candidate set, based on this, according to the formula (5) will be calculated to obtain the parameter list of the density threshold Density, and then according to the Density parameter when the Density parameter tends to be stable, the corresponding value of $K$, to obtain the corresponding Eps and MinPts as the optimal threshold, you can complete the clustering threshold of the Automated setting.

### 2.3.2. DBSCAN Clustering Algorithm Implementation

The specific implementation steps of the DBSCAN adaptive clustering algorithm designed in this study are shown in Fig. 1 as follows:

(1) Input the spatio-temporal dataset $D$ and perform preprocessing operations such as normalization and pooling.

(2) Calculate the distance distribution matrix of spatio-temporal dataset $D$, sort and average the elements to get the Eps parameter list, calculate the corresponding MinPts parameter and get the MinPts parameter list by dichotomous search.

(3) Cluster the spatio-temporal dataset $D$ by DBSCAN clustering algorithm, when the number of clusters tends to stabilize, the current corresponding Eps and MinPts are taken as the optimal thresholds.

(4) Output the clustering results.

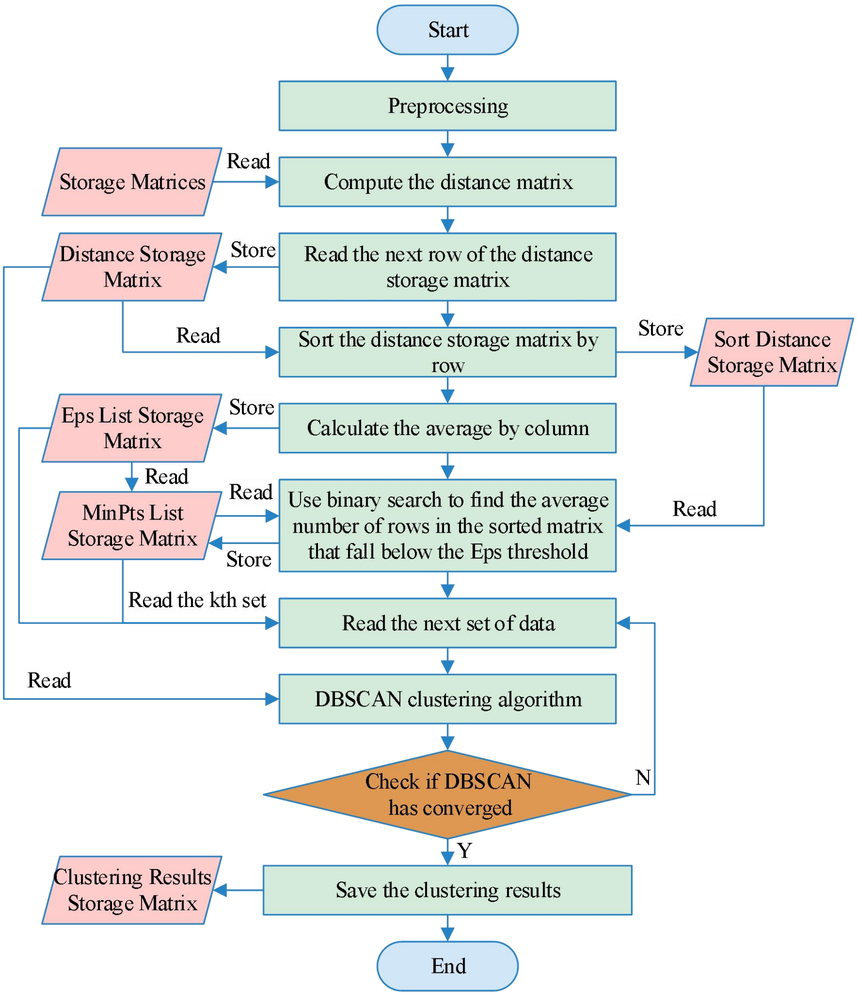


***Figure 1.*** The implementation process of the DBSCAN adaptive clustering algorithm

## 3. Ecological Wisdom Mining Practices of Teochew Monocotyledon Tea

The features of tea pests and diseases of Teochew monocotyledon tea were extracted by association rules, and then the shapes in tea production were clustered and analyzed using the improved DBSCAN algorithm.

### *3.1. Association rule extraction results*

The data of the five factors of average temperature, maximum temperature, minimum temperature, rainfall and sunshine hours are non-discrete numerical attributes, so it is necessary to discretize these data attributes in terms of 0, 1, 2, and 3 grades, and they denote light occurrence, moderate occurrence, heavy occurrence, and heavy occurrence, respectively. The discretization levels for each factor are shown in Table 1. In this way, the recorded data were converted according to the generalized conversion criteria as per the generalized conversion criteria and the conversion resulted in the processed data. The temperature factor contains three factors of average, maximum and minimum temperature and the rest of the factors are rainfall, hours of sunshine, and number of combined pests and diseases.

***Table 1.*** Factor discretization level

| Level | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| A:Average temperature (℃) | <18.5 | 18.5~20.5 | 20.5~22.5 | >22.5 |
| B:Maximum temperature (℃) | <26.5 | 26.5~28.5 | 28.5~30.5 | >30.5 |
| C:Minimum temperature (℃) | <14.5 | 14.5~16.5 | 16.5~18.5 | >18.5 |
| D:Rainfall (mm) | <98.5 | 98.5~190.5 | 190.5~280.5 | >280.5 |
| E:Sunshine duration (h) | <82.5 | 82.5~92.5 | 92.5~105.5 | >105.5 |
| F:Total insect count | <84 | 84~98 | 98~112 | >112 |

We used the Apriori algorithm to conduct association analysis on the processed data sample records. Specifically, we associated each of the five climatic conditions, namely "average temperature", "maximum temperature", "minimum temperature", "precipitation", and "sunshine duration", with the final tea pest occurrence grade. Based on the obtained sample data of feature factors and pest grades, we processed the data using the Apriori association rule mining algorithm. When the minimum support frequency was set to 0.22 and the minimum confidence level was set to 52%, the association rules mined are shown in Table 2.

(1) There is a high correlation between the average temperature and the pest level. When the average temperature is suitable, the confidence level of the corresponding pest level being severe is as high as 76.6%, while when the average temperature is relatively high, the confidence level of the corresponding pest level being severe is 66.4%. This indicates that suitable temperature has a significant impact on the occurrence of pests. Therefore, when the temperature is generally high or during droughts, it is not conducive to the growth and reproduction of the green leafhopper.

(2) The correlation between the maximum temperature and the level of infestation was not very large, which indicated that occasional high temperatures would not affect the occurrence of leafhopper infestation in tea plantations.

(3) There is a certain correlation between the minimum temperature and the infestation level, the confidence level of relatively low minimum temperature for the corresponding infestation level of large occurrence is as high as 72.3%, while the minimum temperature is a little lower for its corresponding infestation level of serious confidence level is also 64.7%, which indicates that the minimum temperature has a certain effect on the occurrence of the infestation, so that the minimum temperature is generally low, which reduces the average air temperature, and becomes the suitable environment for the growth of the little green leafhopper.

(4) There is a certain correlation between rainfall and pest grade, and the confidence level of very heavy rainfall for the corresponding medium occurrence of pest grade is as high as 98.9%, which indicates that very heavy rainfall is unfavorable to the growth of the little green leafhopper, while the appropriate rainfall provides favorable conditions for the environment for the growth of the little green leafhopper, therefore, the little green leafhopper will reproduce more in the environment with a certain degree of humidity, and at that time, it should be prepared in advance to prevent pests.

(5) There is a high correlation between the number of hours of sunshine and the grade of pest occurrence, the confidence level of the corresponding grade of pest occurrence is as high as 77.4% when the number of hours of sunshine is small, while the confidence level of the corresponding grade of pest occurrence is medium when the number of hours of sunshine is low is 99.2%, which indicates that the number of hours of sunshine is small and has a greater impact on the occurrence of pests, so from the viewpoint of geographic location of the tea plantation, pests occurring in tea gardens shaded by large trees are more likely to occur in tea gardens. The chances of insect infestation in tea gardens shaded by large trees are considerably greater than in other tea gardens.

***Table 2.*** The discovered association rules

| Previous item | → | Next item | Confidence |
|---|---|---|---|
| A3 | → | F3 | 66.4% |
| A1 | → | F3 | 76.6% |
| C2 | → | F3 | 72.3% |
| C3 | → | F2 | 64.7% |
| D3 | → | F1 | 98.9% |
| D1 | → | F3 | 78.3% |
| E0 | → | F3 | 77.4% |
| E2 | → | F1 | 99.2% |
| A2&E0 | → | F3 | 81.4% |

## *3.2. DBSCAN clustering mining results*

The mean, standard deviation and coefficient of variation were calculated for 8 agronomic traits of 14 Teochew tea varieties, and the results of data description are shown in Table 3.The mean coefficient of variation was 14.70%, among which the coefficient of variation for 100 buds weight was the largest, 29.34%, and the 100 buds weight was closely related to the appearance and quality of tea leaves. The coefficients of variation for shoot aspect ratio and leaf spreading angle were the next highest, 20.63% and 19.22%, respectively, and the coefficient of variation for tea color was the smallest, 7.38%.The eight traits can be divided into three categories, which are phenotypic traits of the new shoots (including shoot aspect ratio (T1), 100 buds weight (T2), and leaf spreading angle (T3)), biochemical traits (amino acids (T4), tea polyphenols (T5), water leachate (T5), and water extracts (T5)), and biochemical traits (amino acid (T4), tea polyphenols (T5), and water extracts (T4)). , water leachate (T6)) and sensory quality traits (shape (T7), soup color (T8)), among which the coefficient of variation for sensory quality traits was low.

***Table 3.*** Data description results

| Serial | Length width ratio | Hundred buds | Angle of leaf | Amino acid % | Tea % | Water % | Shape | Tea infusion color |
|---|---|---|---|---|---|---|---|---|
| 1 | 6.11 | 11.87 | 15.17 | 4.98 | 19.9 | 45.55 | 21.60 | 9.16 |
| 2 | 5.78 | 11.92 | 14.85 | 4.65 | 22.35 | 43.33 | 21.78 | 10.52 |
| 3 | 7.88 | 11.09 | 13.99 | 6.14 | 22.8 | 44.41 | 21.32 | 6.89 |
| 4 | 8.38 | 11.52 | 16.45 | 3.89 | 23.25 | 45.44 | 20.75 | 9.62 |
| 5 | 8.62 | 13.43 | 14.07 | 3.52 | 22.48 | 45.34 | 24.49 | 8.86 |
| 6 | 7.10 | 13.55 | 17.2 | 3.49 | 21.78 | 44.38 | 21.81 | 9.60 |
| 7 | 8.05 | 12.49 | 16.08 | 5.42 | 23.1 | 43.85 | 19.71 | 9.39 |
| 8 | 6.73 | 13.47 | 15.71 | 4.69 | 22.2 | 43.98 | 23.76 | 10.64 |
| 9 | 9.88 | 10.86 | 14.76 | 4.79 | 22.49 | 44.11 | 20.91 | 9.13 |
| 10 | 6.37 | 14.71 | 14.66 | 3.76 | 21.8 | 44.87 | 22.27 | 10.61 |
| 11 | 7.20 | 11.33 | 15.91 | 5.64 | 21.52 | 43.63 | 21.47 | 8.05 |
| 12 | 7.14 | 11.41 | 14.06 | 4.04 | 21.12 | 44.38 | 22.64 | 9.51 |
| 13 | 7.29 | 10.66 | 14.52 | 4.38 | 20.41 | 45.84 | 19.79 | 9.40 |
| 14 | 6.88 | 11.89 | 14.09 | 5.15 | 22.47 | 44.67 | 21.53 | 9.16 |
| Avg | 7.39 | 12.16 | 15.11 | 4.61 | 21.98 | 44.56 | 21.7 | 9.32 |
| Sd | 1.09 | 1.20 | 1.01 | 0.81 | 0.97 | 0.77 | 1.32 | 1.00 |
| CV | 20.63% | 29.34% | 19.22% | 14.55% | 7.38% | 8.36% | 9.34% | 8.82% |

The similarity coefficients among the tea plant traits are shown in Fig. 2. The results showed that the correlation between shoot aspect ratio and spreading angle was high, with a correlation coefficient of 0.964. Once again, amino acids were significantly negatively correlated with tea polyphenols, with a correlation coefficient of -0.902. The correlation coefficients of tea polyphenols were positively correlated with the water extract and the color of the soup, with the correlation coefficients of 0.763 and 0.665, respectively.

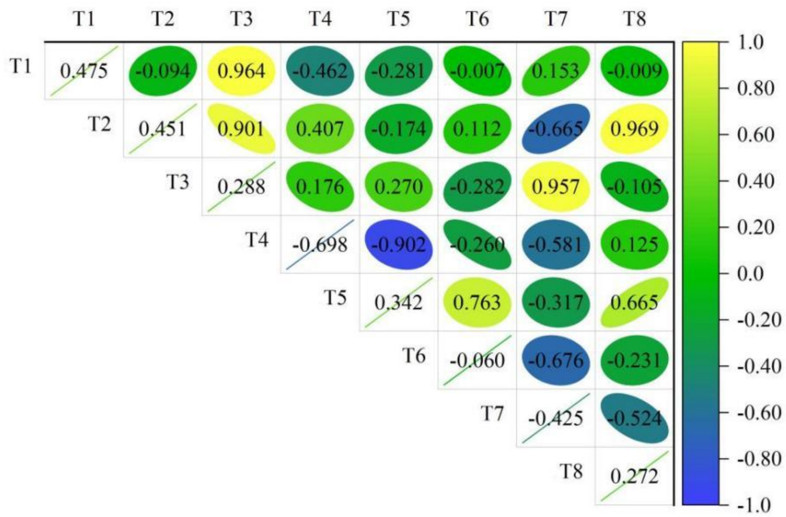


***Figure 2.*** Similarity coefficient among the characteristics of tea plants

The five principal components of the results of the principal component analysis of agronomic traits of tea tree varieties can reflect most of the information of the eight trait indexes, therefore, the first five principal components were selected as the comprehensive indexes, and the eigenvector matrix of the tea tree new-topping traits is shown in Figure 3. According to the eigenvector matrix of tea tree new tip traits can be obtained as the functional equation of the 5 principal components, taking tea tree new tip trait T1 as an example, T1=-0.063*PC1-0.160*PC2-0.062*PC3-0.180*PC4+0.063*PC5.

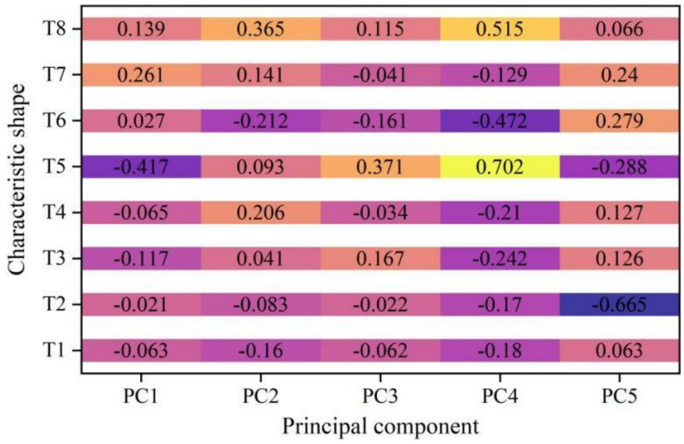


***Figure 3.*** The characteristic vector of the new shoot traits of tea plants

The main traits of new growth, biochemistry and tea-making quality of 14 Chaozhou single bush tea tree varieties were clustered according to the similarity coefficient method. The clustering results of different tea tree varieties are shown in Figure 4. A total of three clusters were divided into three clusters, of which cluster 1 contained five varieties, in the order of [3,8,9,12,13], and cluster 2 and cluster 3 contained four and three varieties, respectively. The shape characteristics of different tea trees can be fully explored through the cluster analysis of Teochew monocotyledon tea production techniques.

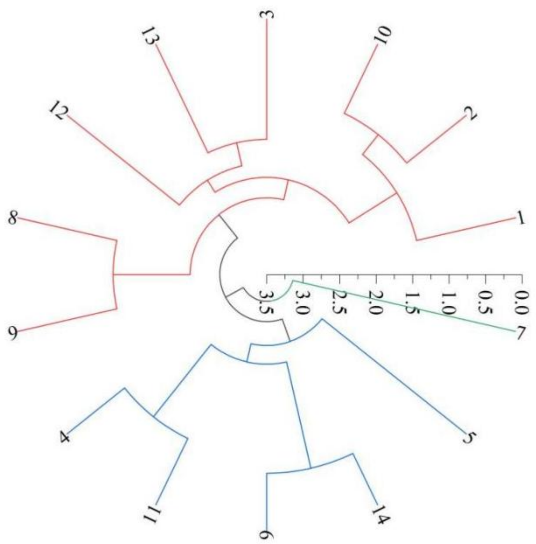


***Figure 4.*** The clustering results of different tea tree varieties

## 4. Conclusion

In this paper, we applied association rules and improved DBSCAN clustering algorithm to analyze the ecological mining of tea production technology characteristics of TeoChew single-column tea from two aspects, namely, tea pests and diseases and agronomic shapes of tea trees. The research conclusions are as follows:

(1) The confidence level that the average temperature is suitable to point to the pest grade as a major occurrence is 76.6%, and the confidence level that the average temperature is higher to guide the pest grade as a major occurrence is 66.4%, and the growth and reproduction of the little green leafhopper will be affected when the temperature is generally higher or drought.

(2) Eight key agronomic traits were extracted for Chaozhou single bush tea, namely, bud length and breadth ratio, 100 buds weight, leaf spreading angle, amino acid, tea polyphenol, water leachate, shape and soup color, which belonged to the categories of phenotypic traits of new shoots, biochemical quality traits and sensory quality traits, among which the correlation coefficients of the bud length and breadth ratio and leaf spreading angle reached 0.964.

(3) According to the agronomic traits of tea tree varieties divided into five principal components,

clustering to get three clusters, the improved DBSCAN clustering method enhances the mining of key features of tea tree shape.

**Funding**
This work was supported by the "Chaozhou Culture Research" Special Project of the Guangdong Provincial Philosophy and Social Sciences Planning (Project Title: Researching and Planning the Inheritance and Development of Chaozhou DanCong Tea Making Techniques from a Cultural Ecology Perspective) (Grant No. GD22CZZ14).

**About the Author**
Min Zhang was born in Maoming, Guangdong, China in 1980. She received her Ph.D. degree from Sun Yat-sen University. She is currently a lecturer at the Beijing Normal University, with research interests including intangible cultural heritage, anthropology, and folklore studies.

**References**

1. Chunlin, H. U. A., Jiuhong, Z. H. A. N. G., & Canqiang, Z. H. A. N. G. (2022). Agricultural heritage protection and development policy from the perspective of government documents. Journal of Resources and Ecology, 14(1), 102-113.
2. Food, F. A. O. (2008). Food and Agriculture organization of the United Nations. A framework for land evaluation Soils Bulletin, 32.
3. Guo, X., & Min, Q. (2023). Analysis of landscape patterns changes and driving factors of the guangdong chaoan fenghuangdancong tea cultural system in China. Sustainability, 15(6), 5560.
4. Qiu, J., & Chen, S. (2024). The innovative application of Internet of Things technology in the cultural inheritance and industrial development of Phoenix Dancong tea. International Journal of New Developments in Engineering and Society, 8(5), 502.
5. Qingwen, M. I. N. (2021). Agri-cultural heritage: An interdisciplinary field with development prospects. Journal of Resources and Ecology, 12(4), 437-443.
6. Daugstad, K., Rønningen, K., & Skar, B. (2006). Agriculture as an upholder of cultural heritage? Conceptualizations and value judgements—A Norwegian perspective in international context. Journal of Rural Studies, 22(1), 67-81.
7. Kajihara, H., Zhang, S., You, W., & Min, Q. (2018). Concerns and opportunities around cultural heritage in east Asian globally important agricultural heritage systems (GIAHS). Sustainability, 10(4), 1235.
8. Xu, L., Liao, Y., & Liu, J. (2025). Study on the evolution of the spatial patterns and driving factors of national agricultural cultural heritage in China. Plos one, 20(1), e0313926.
9. Agnoletti, M., & Santoro, A. (2022). Agricultural heritage systems and agrobiodiversity. Biodiversity and Conservation, 31(10), 2231-2241.
10. Reyes, S. R. C., Miyazaki, A., Yiu, E., & Saito, O. (2020). Enhancing sustainability in traditional agriculture: Indicators for monitoring the conservation of globally important agricultural heritage systems (GIAHS) in Japan. Sustainability, 12(14), 5656.
11. Assandri, G., Bogliani, G., Pedrini, P., & Brambilla, M. (2018). Beautiful agricultural landscapes promote cultural ecosystem services and biodiversity conservation. Agriculture, Ecosystems & Environment, 256, 200-210.
12. Ye, R., Liu, M., & Li, Z. (2025). Designing a theoretical framework for ecosystem products of agricultural heritage systems. Agriculture, 15(7), 755.
13. Peng, Q., & Zhang, B. (2023). Protection of agricultural cultural heritage: A new opportunity for the development of ecological agriculture. Academic Journal of Agricultural Sciences, 4(1).
14. Li, M., Zhang, Y., Xu, M., He, L., Liu, L., & Tang, Q. (2019). China eco-wisdom: A review of sustainability of agricultural heritage systems on aquatic-ecological conservation. Sustainability, 12(1), 60.

15. Bin, W. A. N. G., Yehong, S. U. N., & Wenjun, J. I. A. O. (2021). Ecological benefit evaluation of agricultural heritage system conservation—a case study of the Qingtian rice-fish culture system. Journal of Resources and Ecology, 12(4), 489-497.
16. Yan, J. (2025). Innovative Cultural Transmission Paths Analysis for Local Opera Based on Data Mining. International Journal of High Speed Electronics and Systems, 34(02), 2440027.
17. Pernia, D. L., Demeyer, S., Schalm, O., & Anaf, W. (2018, June). A data mining approach for indoor air assessment, an alternative tool for cultural heritage conservation. In IOP Conference Series: Materials Science and Engineering (Vol. 364, No. 1, p. 012045). IOP Publishing.
18. Li, Y., Ma, X., & Song, J. (2024). Global industrial cultural heritage protection and mining strategies based on big data. Journal of Computational Methods in Science and Engineering, 24(4-5), 2909-2926.
19. Amato, F., Moscato, V., Picariello, A., Colace, F., Santo, M. D., Schreiber, F. A., & Tanca, L. (2017). Big data meets digital cultural heritage: Design and implementation of scrabs, a smart context-aware browsing assistant for cultural environments. Journal on Computing and Cultural Heritage (JOCCH), 10(1), 1-23.
20. Li, P., Shi, Z., Ding, Y., Zhao, L., Ma, Z., Xiao, H., & Li, H. (2021). Analysis of the temporal and spatial characteristics of material cultural heritage driven by big data—Take museum relics as an example. Information, 12(4), 153.
21. Vonitsanos, G., Kanavos, A., Mohasseb, A., & Tsolis, D. (2019, July). A nosql approach for aspect mining of cultural heritage streaming data. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-4). IEEE.
22. Xie, Y. (2021, November). Internet Big Data Mining and Research on Intangible Cultural Heritage Activation with Smart Devices. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 724-727). IEEE.
23. Yu, Z. (2025, June). Intelligent Generation of Intellectual Property Based on User-Generated Content and Big Data Mining in Culture-Tourism Integration in Rural Areas. In 2025 IEEE International Conference on Computation, Big-Data and Engineering (ICCBE) (pp. 669-673). IEEE.
24. Pitka, T., Bucko, J., Krajči, S., Krídlo, O., Guniš, J., Šnajder, Ľ., ... & Eliaš, P. (2025). Time analysis of online consumer behavior by decision trees, GUHA association rules, and formal concept analysis. Journal of Marketing Analytics, 13(1), 29-52.
25. Kulkarni, O., & Burhanpurwala, A. (2024, February). A survey of advancements in DBSCAN clustering algorithms for big data. In 2024 3rd International conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC) (pp. 106-111). IEEE.
26. Chen, Y., Tan, P., Li, M., Yin, H., & Tang, R. (2024). K-means clustering method based on nearest-neighbor density matrix for customer electricity behavior analysis. International Journal of Electrical Power & Energy Systems, 161, 110165.