Journal of Engineering, Project, and Production Management
2026, 16(5), 2026-4

Using SOM Fusion Algorithms and Development Planning Strategies for Traditional Village Classification

Kai Qin[1], Cuixia Li[2], and Qian Wang[3]
[1] Lecturer, Architecture Engineering College, Huanghuai University, Zhumadian, 463000, China
[2] Lecturer, College of Animal Husbandry Engineering, Zhumadian Agricultural Engineering Vocational College, Zhumadian, 463000, China
[3] Lecturer, Architecture and Engineering College of Huanghuai University, Zhumadian, 463000, China, E-mail: wanngqiann@outlook.com (corresponding author).

Project Management
Received January 4, 2026; revised March 4, 2026; accepted July 13, 2026
Available online July 25, 2026

Abstract: In response to the problems of inaccurate classification and homogeneous strategies in the protection and development of traditional villages, this study aims to construct a data-driven village classification and differentiated development planning strategy system. By introducing a clustering method that combines self-organizing mapping and the K-means algorithm, automated, high-precision classification of villages' multi-dimensional features can be achieved. By combining the analytic hierarchy process and the entropy weight method for weighting, a village development evaluation index system is constructed to accurately identify development shortcomings. The experimental findings demonstrate that, when testing the classification accuracy of the research method, accuracy gradually increases to 92.5% as the data volume increases to 250. When the number of clusters is 5, the central processing unit takes only 3.4 seconds. In practical application, the highest employment rate for the research method is 58.2%, from the implementation year through the 5th year of the strategy. When the neighborhood radius is 6, the highest contour coefficient is 0.72. When the data missing rate is 0%, the highest adjusted Rand index of the research method reaches 0.91. The outcomes indicate that the research method has higher classification accuracy, computational efficiency, and better robustness, providing scientific methods for precise classification and development of traditional villages, as well as a quantitative decision-making basis for differentiated management of rural revitalization.

Keywords: Village classification; SOM fusion algorithm; development planning strategy; combination weighting; development indicator system.

DOI 10.32738/JEPPM-2026-4

---

1. Introduction

With accelerating urbanization and deepening rural revitalization strategies, the protection and sustainable development of traditional villages have become important issues for national cultural security and rural governance. However, traditional villages are currently facing multiple challenges such as an imperfect classification system, a one size fits all approach to protection and development strategies, and conflicts between cultural inheritance and modernization construction. It is urgent to achieve precise identification, classification guidance, and differentiated governance through scientific methods and technological means (Wiyono et al., 2023; Li et al., 2023). High resolution remote sensing technology can be used for village morphology extraction and change detection, but it is limited by data acquisition costs and temporal consistency. The social network analysis method can analyze the governance structure of villages, but it is difficult to integrate data from material space and social dimensions. Self-Organizing Maps (SOMs) can better preserve the structural information of high-dimensional data and enable visual mapping, but their clustering results heavily rely on preset network structures, which can easily lead to boundary rigidity (Mohammadi and Mobarakeh, 2022). The Analytic Hierarchy Process (AHP) can effectively integrate expert knowledge and construct a three-level judgment matrix. Entropy Weight Method (EWM) quantifies the variability of data itself and can correct subjective weight (SW) bias by calculating information entropy (Bai et al., 2024; Wen, 2023). Therefore, to construct a traditional-village intelligent classification model driven by multi-source data and to establish a scientific and transparent development evaluation system to support differentiated and implementable development planning strategies, a clustering algorithm integrating SOM and K-means, as well as a combined weighting model integrating AHP and EWM, is proposed. The novelty of the study lies in integrating SOM and K-means for village classification, thereby addressing the challenges of clustering high-dimensional data. Propose AHP-EWM combination weighting, balance SW and objective weight (OW), and enhance the scientificity of development evaluation.

The classification and development planning of traditional villages have now received more in-depth research. Allawi and Al-Jazaeri (2023) proposed an urban-rural integration model centered on central villages to address lagging rural development. By analyzing multiple urban service indicators through geographic information systems, select central villages with radiation and driving capabilities. The findings demonstrated that it could promote the sharing of public services and the optimization of spatial structure, thereby supporting sustainable development in rural areas. Zhang et al. (2024) proposed a boundary recognition model to address the difficulty of accurately identifying and monitoring urban villages. This method generated mixed prompts using semantic segmentation models and performed fine-grained segmentation using segmented arbitrary models. The findings demonstrated that the model could efficiently delineate the boundaries of urban villages, offering a deeper understanding of their evolutionary trajectories. Meng et al. (2022) proposed an automatic classification approach grounded in deep learning to address the high cost and low efficiency of manual surveys of rural buildings. The findings demonstrated that the model could accurately and automatically identify 7 features, such as building functions and structures, providing an efficient and low-cost technical means for large-scale village research and renovation. Wang et al. (2023) proposed a tourism service facility effectiveness evaluation index system to address the imbalance between traditional village protection and tourism development. They constructed multidimensional indicators based on the concept framework of life protection. The findings demonstrated that the model could effectively assess the impact of facilities on villagers, buildings, and culture, providing a quantitative basis for coordinating protection and development efforts. Xiao et al. (2025) proposed an intelligent interest coordination and communication platform to address the conflict between modernization and heritage protection in traditional village governance. The findings demonstrated that the platform could enhance information transparency, promote multi-party negotiation, and provide an effective path for traditional villages to achieve intelligent governance and sustainable development.

The SOM algorithm has also received extensive research from scholars both domestically and internationally. Hua and Anderson-Frey (2022) introduced an analysis model based on SOM to address the complex, variable, and difficult to classify tornado and storm environments. During the process, the probability deviation is calculated using the kernel density estimation method, and SOM is used for clustering recognition. The findings demonstrated that the model could effectively reveal the spatiotemporal differentiation patterns of environmental parameters. Jhong et al. (2022) proposed an ensemble forecasting model based on an SOM and a backpropagation network to improve the accuracy of real-time flood forecasting during typhoons. During the process, SOM is used to cluster and screen weight biases. The findings demonstrated that the integrated model could provide more accurate peak flow and prediction intervals, effectively improving the reliability of flood warning. Ouzerbane et al. (2022) introduced a hydrochemical analysis method based on SOM to address the origin of salinization in coastal aquifers. During the process, SOM was used to classify the spatiotemporal evolution of 47 water sample data. The findings demonstrated that this method effectively revealed the spatial differentiation pattern of salinity parameters increasing from the center to the southwest. Sakkari et al. (2022) proposed a generalized unsupervised deep SOM model to tackle the challenges of computational complexity and insufficient generalization ability in SOM feature extraction. They built a parallel learning architecture by improving the subsampling module. The findings demonstrated that the model significantly improved classification accuracy, training efficiency, and generalization ability on the MNIST and STL-10 datasets. Liew et al. (2023) introduced an intelligent classification model based on SOM to address the problem of manual and costly wood color classification. During the process, multiple types of cameras were used to capture images of red oak and other wood, extract color and texture features, and train classification through SOM. The findings demonstrated that the classification accuracy of the model for three types of wood reached 89% -96.4%.

In summary, traditional village classification techniques focus on a single dimension, lack systematicity, and have a disconnect between strategies and classification associations. The classification results of SOM heavily rely on the preset two-dimensional grid structure and scale, which may not capture the most natural and compact clustering structure. In view of this, the research aims to achieve comprehensive intelligent classification of traditional villages driven by multi-source data through the construction of a SOM-K-means fusion model, and introduce a combined weighting model combining AHP and EWM to form differentiated and operable development planning strategies, systematically solving the problem of disconnection between classification and development strategies.

2. Methodology

2.1. Village Classification Design Based on SOM Fusion Algorithms

The protection of traditional villages faces problems such as vague classification and single development strategies, and it is urgent to introduce data-driven methods to explore their inherent type patterns from a vast amount of multidimensional features. The SOM algorithm can transform complex village features into a two-dimensional visualization map, automatically cluster similar villages through topology preservation techniques, and provide data support for developing differentiated protection plans (Liu et al., 2022). The network structure of this algorithm is shown in Fig. 1.

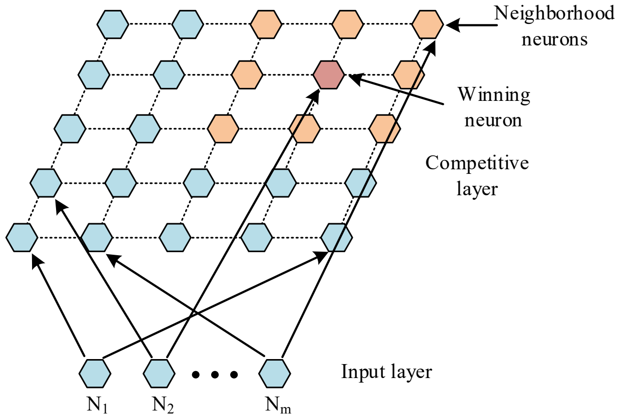


Fig. 1. The self-organizing neural network structure of SOM algorithm

As shown in Fig. 1, the SOM network maps multidimensional input data to a two-dimensional competitive layer. Through a competitive mechanism, each sample activates a winning neuron and updates the weights of that neuron and its neighboring units, so that similar samples map to adjacent positions in two-dimensional space, thereby forming a topologically ordered feature map. The core processes of this network include competition, cooperation, and weight adaptation, where the cooperation process is implemented through neighborhood functions, as shown in Eq. (1). The neighborhood function is typically defined using a Gaussian function, where $d_{c,j}$ is the distance between neurons and $\sigma(t)$ is the neighborhood radius.

$$h_{c,i}(t) = exp(-d_{c,j}^2/2\sigma(t)^2) \tag{1}$$

In Eq. (1), $h_{c,i}(t)$ is the value of the neighborhood function, $d_{c,j}$ is the geometric distance between neuron$i$and winning neuron $c$in the output layer grid, and $\sigma(t)$is the attenuation width parameter (neighborhood radius), which decays over time to ensure convergence. The learning rate $\alpha(t)$ and the neighborhood radius $\sigma(t)$ are critical for network convergence and topology preservation. Additionally, to avoid the impact of the different dimensions of the input vector's characteristics, z-score standardization is applied to the input mode vector. The specific expression is presented in Eq. (2).

$$V_{new} = (v - \phi)/\lambda \tag{2}$$

In Eq. (2), V_{new} is the standardized new value, v is the original data value of a certain feature, ϕ is the arithmetic mean of all data values of that feature, λ and is the standard deviation of all data values of that feature. To maintain consistency between the output layer topology and the high-dimensional data distribution, SOM networks dynamically adjust the weight vectors of winning neurons and their neighboring neurons to preserve the topology in the mapping space. The mathematical principle for weight update is given by Eq. (3).

$$\begin{cases} Q_i(t+1) = Q_i(t) + \alpha(t)e^{-n}[V - Q_i(t)], \; i \in L_{i^*}(t) \\ Q_i(t+1) = Q_i(t), \; i \notin L_{i^*}(t) \end{cases} \tag{3}$$

In Eq. (3), $L_{i^*}(t)$ is the neighborhood function centered around the winning neuron i^*at time$t$,$Q_i(t)$is the weight vector of the neuron i at time t, $\alpha(t)$ is the learning rate, ande^{-n} is the topological distance between neurons. To ensure the accuracy and orderly structure of village classification results, the study optimized the network by dynamically adjusting the learning rate and the range of superior neighborhoods. The update rules for these parameters are shown in Eq. (4), which includes the formulas for the learning rate $\alpha(t)$ and the neighborhood radius $\sigma(t)$.

$$\begin{cases} \alpha(t) = e^{-n}/(t+2), \; t = 1,2,\ldots,T \\ L(t) = L(0)(1 - t/T), \; t = 1,2,\ldots,T \end{cases} \tag{4}$$

In Eq. (4), $T$is the preset total number of training iterations,$L(t)$ is the neighborhood size of time$t$, and $L(0)$ is the initial neighborhood size at the initial training time. The weight vector update equation combines $h_{c,i}(t)$ and $\alpha(t)$, and its mathematical expression is shown in Eq. (5).

$$w_i(t+1) = w_i(t) + \alpha(t)h_{c,i}(t)[x(t) - w_i(t)] \tag{5}$$

In Eq. (5), $w_i(t)$ represents the weight vector before update, $w_i(t+1)$ represents the weight vector after update, and $x(t)$ represents the current input sample vector.

The classification results of SOM networks are limited by their fixed topological grid structure, which can lead to inflexible category boundaries. However, the K-means algorithm can dynamically determine clustering centers based on data distribution, making category division more accurate and compact. The clustering structure of this algorithm is shown in Fig. 2.

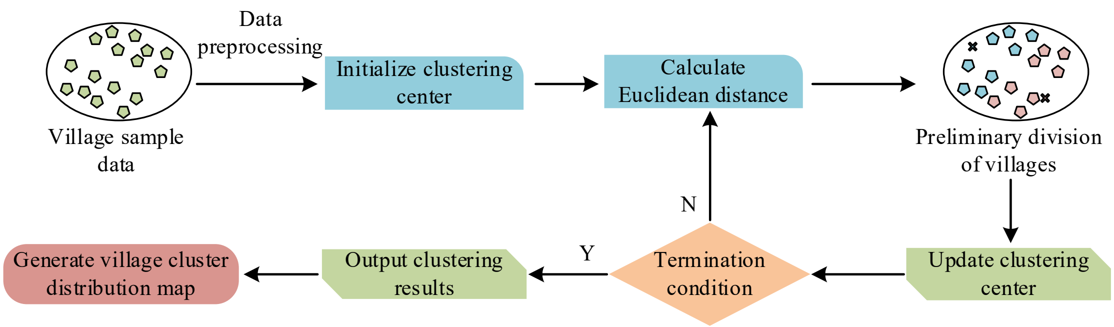


Fig. 2. The clustering structure of K-means algorithm

As shown in Fig. 2, the algorithm first preprocesses the multidimensional village sample data and then initializes K cluster centers (Tissot, 2025; Hebbi and Mamatha, 2023). The algorithm iteratively calculates the distance from each village to each center, assigns it to the nearest cluster, and updates the cluster center meaning until the center point changes by less than a threshold or reaches the maximum number of iterations. Finally, it outputs clustering results and evaluates their effectiveness.

To objectively assess clustering effectiveness and determine the optimal number of village types, the study optimizes the objective function by minimizing the sum of squared errors, as given by Eq. (6).

$$E = \sum_{m}^{i=1} \sum_{u \in C_i} |u - \eta_i|_2^2 \tag{6}$$

In Eq. (6), E is the sum of squares of total errors, C_i is the ith cluster, and η_i is the cluster center of the ith cluster. Considering that high-dimensional village datasets contain a large number of redundant or unrelated features, the study adopts a correlation based feature selection algorithm to screen the most representative features in the dataset. The definition of the estimated value of its individual feature is shown in Eq. (7).

$$Z_{\text{merit}} = nb_{cf} / \sqrt{n + n(n-1)b_{ft}} \tag{7}$$

In Eq. (7), Z_{merit} is the "value" score of the feature subset to be evaluated, b_{cf} is the average correlation between all features in the feature subset and the target category, and b_{ft} is the average correlation between all feature pairs within the feature subset. Finally, the SOM algorithm is integrated with the K-means algorithm to construct a village classification process based on SOM-K-means, as shown in Fig. 3.

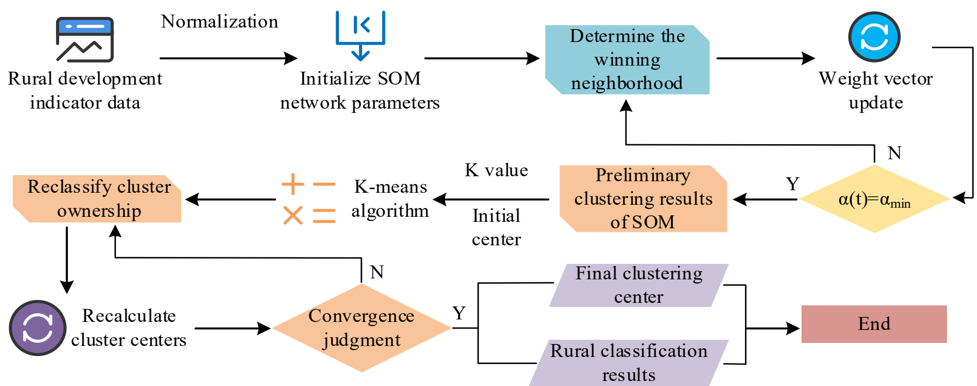


Fig. 3. Village classification process based on SOM-K-means (https://iconpark.oceanengine.com/home)

In Fig. 3, the SOM neural network is used to standardize and competitively learn multi-source indicator data. Subsequently, it is used as the initial parameter for K-means, and the distance between each sample and the center is iteratively calculated, followed by re-clustering and updating the center. The final output is a stable cluster center, and the villages are divided into five categories: urban-rural integration, agglomeration and improvement, characteristic protection, rectification and improvement, and relocation and consolidation. The study employs the Davidson Boulding (DB) index to evaluate the clustering performance of the SOM-K-means fusion algorithm. The formula for the DB index is given by Eq. (8).

$$DB = \frac{1}{m} \sum_{i=1}^{m} \max_{i \neq j} \left[\left(avg(C_i) + avg(C_j) \right) / D(\eta_i, \eta_j) \right] \tag{8}$$

In Eq. (8), DB is the final score of the DB index, $avg(C_i)$ is the average distance from all sample points in the ith cluster to its center point, and $D(\eta_i, \eta_j)$ is the distance between the center point of the ith cluster and the center point of the jth cluster.

The optimal number of clusters, K, was determined to be 5 based on a comprehensive evaluation that combined the elbow method, silhouette coefficient, Davies Bouldin (DB) index, and practical domain constraints. The elbow point of the within-cluster sum of squares, the maximum average silhouette width, and the minimum DB index all consistently indicated that K=5 yields the most compact and well-separated clusters. Furthermore, this number aligns with the five typical village development categories recognized in Chinese rural planning policies (urban-rural integration, agglomeration and improvement, characteristic protection, rectification and improvement, relocation and consolidation), ensuring the classification results are both statistically optimal and policy-relevant.

2.2. Design of Village Development Planning Strategies Based on Combination Weighting

Although the SOM-K-means method can classify village types, its results are greatly affected by data quality and are difficult to directly guide policies. Therefore, it is particularly crucial to establish a scientific indicator system for village development that not only provides a data foundation for classification but also effectively links classification results to planning strategies. To further eliminate the differences in scale and magnitude between indicators, it is essential to standardize the original data using the method shown in Eq. (9).

$$\begin{cases} Z_i^+ = \left(S_i - S_{\min}\right) / \left(S_{\max} - S_{\min}\right) \\ Z_i^- = \left(S_{\max} - S_i\right) / \left(S_{\max} - S_{\min}\right) \end{cases} \tag{9}$$

In Eq. (9), $Z_i^+$is the standardized positive indicator value,$S_i$is the original value of the ith village on the indicator,$S_{max}$is the max value of all villages on the indicator,$S_{min}$is the min value of all villages on the indicator,Z_i^- and is the standarized negative indicator value. The study uses a combination of AHP and EWM to determine the final weights, incorporating subjective judgments of the importance of indicators based on expert experience, as well as objective distribution patterns of the data itself, to ensure that the evaluation results are more scientific and reliable (Ji et al., 2023; Liu et al., 2023). The importance ranking process of village development indicators based on AHP and EWM is shown in Fig. 4.

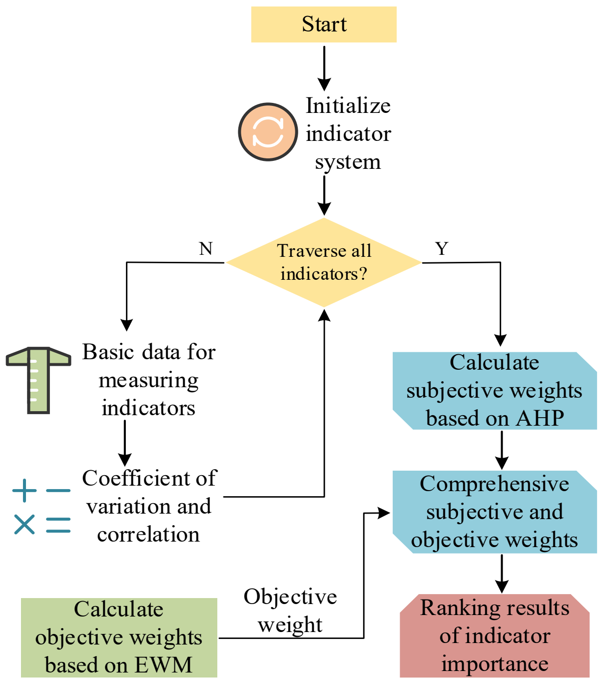


Fig. 4. Importance ranking process of village development indicators (https://iconpark.oceanengine.com/home)

In Fig. 4, the process first establishes a complete evaluation index system and collects basic data. Subsequently, AHP is applied to calculate SWs, and EWM to calculate OWs. Finally, by integrating SW and OW with a combined weighting model, comprehensive importance coefficients for each indicator are obtained, and the final importance ranking is generated from these coefficients. Due to the logarithmic calculation required by the EWM, a standardized value of 0 yields undefined logarithms; to maintain the objectivity and stability of the data, a small translation of the standardized data is also necessary. The translation formula is shown in Eq. (10).

$$s_{ij} = \beta + z_{ij}^{'} \tag{10}$$

In Eq. (10), $s_{ij}$is the final data value after translation processing,$\beta$is a very small positive number, and $z_{ij}^{'}$is the standardized indicator value.

To ensure a more scientific and objective weighting system for village development indicators, it is necessary to calculate the indicators' information entropy and utility values. The specific calculation formula is shown in Eq. (11).

$$E_{xj} = -\beta \sum_{i=1}^{n} B_{ij}\, ln(B_{ij}),\ \ D_j = 1 - E_{xj} \tag{11}$$

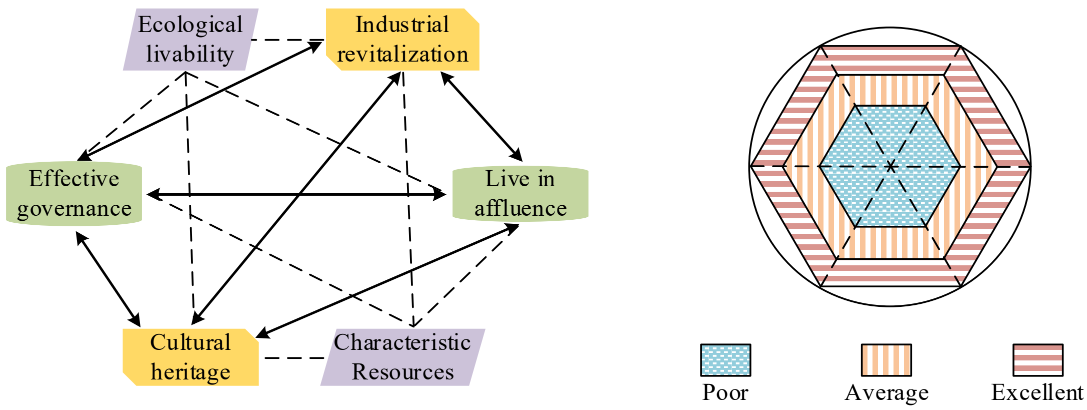
In Eq. (11), E_{xj} is the information entropy of the jth indicator, β is the normalization coefficient, and D_j is the difference coefficient of the jth indicator. To quantify the deviation between theoretical weights and expert judgments, it is necessary to conduct AHP consistency tests. The formula for calculating the maximum eigenvalue is given in Eq. (12).

$$\mu \frac{1}{r} \sum_{i=1}^{r} [(Aq)_i/q_i]_{max} \tag{12}$$

In Eq. (12), $A$is the judgment matrix,$\mu_{max}$is the max eigenvalue of the judgment matrix, and $q$is the weight vector calculated by the AHP. Afterward, the linear combination method was used to fuse the SW and OW, and the combined weight of the village development level indicator was calculated. The specific expression is shown in Eq. (13).

$$Q_z = \delta Q_i^{'} + \gamma Q_i^{''} \tag{13}$$

In Eq. (13), Q_z is the final combination weight of the ith indicator, $Q_i^{'}$ is the SW of the ith indicator calculated by the AHP, $Q_i^{''}$ is the OW of the ith indicator calculated by the EWM,$\delta$is the coefficient of the SW in the combination weight, and $\gamma$is the coefficient of the OW in the combination weight. To effectively identify the key bottlenecks constraining village development and provide a clear basis for formulating precise and differentiated strategies, the study determines a comprehensive score for village development levels based on the multi-faceted area method. The evaluation system and indicator diagram of this method are shown in Fig. 5.



(a) Evaluation system diagram

(b) Schematic diagram of indicator method

Fig. 5. Evaluation system and indicator diagram based on multi-faceted area method

As shown in Fig. 5, multiple indicator data from the village development evaluation system are standardized so that all indicator values fall within the interval of [-1, 1]. Then, with the central point as the origin, each standardized indicator value is plotted on its corresponding coordinate axis. These points are connected sequentially to form an irregular polygon. Finally, the area of this polygon is calculated. This area value represents the final score for the village's comprehensive development level. The formula for calculating the comprehensive score of the village development level is given by Eq. (14).

$$S_z = sin\frac{\varphi}{2} * (Z_a Z_b + Z_b Z_c + Z_c Z_d + Z_d Z_e + Z_e Z_a) \tag{14}$$

In Eq. (14), Z_a,Z_b,Z_c,Z_d, and Z_e are the standardized scores of villages in five different development dimensions, S_z is the comprehensive score of the village development level, and φ is the angle between the indicator axes on the radar chart. Finally, based on the above design, a research framework for village development planning strategies using the SOM-K-means classification method and combined weighting is obtained, as shown in Fig. 6.

In Fig. 6, first, multi-source village data are integrated and preprocessed, and then the SOM-K-means fusion algorithm is employed to achieve scientific classification of villages: SOM is used for dimensionality reduction and initial clustering, and K-means is used for precise iteration. Based on the classification results, a differentiated evaluation system is constructed, and AHP (subjective) and EWM (objective) are combined for weighting. Finally, the polygon area method is used to calculate the comprehensive development score for each village, accurately identifying strengths and weaknesses and providing a quantitative basis for formulating differentiated and precise village development planning strategies.

3. Results and Analysis

3.1. Performance Testing of Fusion Algorithms and Development Planning Strategies

To evaluate the performance of the SOM fusion algorithm and the development planning strategy for traditional village classification, the Chinese traditional village spatial distribution dataset was selected as the test data source. The software, hardware, and their parameters used for performance testing are presented in Table 1.

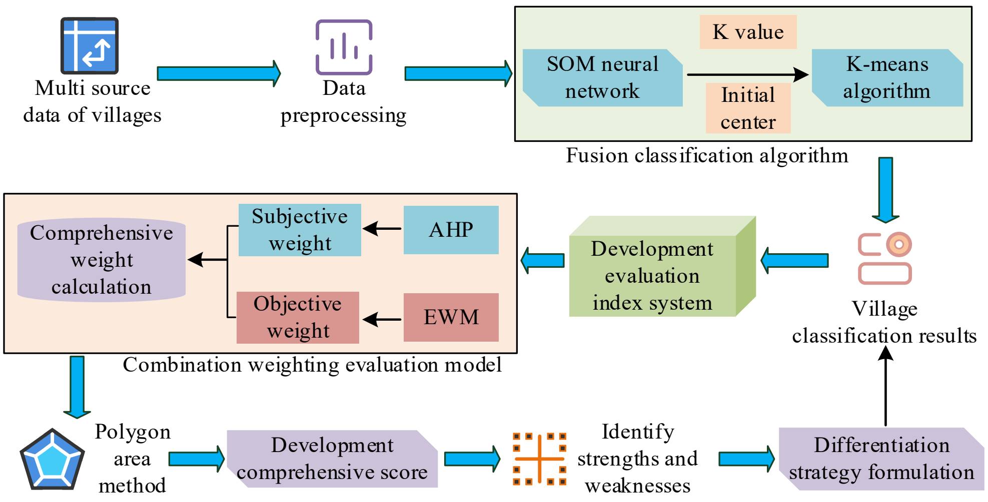


Fig. 6. Research framework for village development planning strategies (https://iconpark.oceanengine.com/home)

Table 1. The preset parameters of each component in the device

Software and hardware names	Parameter description
Programming language	Python
Geographical information system	QGIS
Statistical analysis	SPSS
Development environment	Anaconda 3
Visual analysis	VOSviewer
Central Processing Unit (CPU)	16GB
Storage	1TB

Based on the above software, hardware, and dataset, the research method was tested and referred to as the SKM-AHP-EWM method. To evaluate the robustness of the proposed method under incomplete data, missing values were artificially introduced into the dataset following a missing completely at random (MCAR) mechanism. Two common missing data handling techniques were employed: mean imputation for continuous variables and listwise deletion for samples with critical missing attributes. The results reported in Section 3.2 for different missing rates were based on the mean imputation strategy, as it preserved sample size while maintaining data integrity.

To verify the effectiveness of the proposed SKM-AHP-EWM method, the study selected two baseline models for comparison: the method combining the Gaussian Mixture Model with the Structural Equation Model (GMM-SEM) and the method combining Hierarchical Clustering with the Structural Equation Model (HC-SEM). In the GMM-SEM method, GMM was first used to perform soft clustering on the multidimensional characteristics of villages, assigning each village a probability of belonging to each cluster; then, SEM was applied to each identified cluster to assess its internal consistency and verify the rationality of the clustering structure. In the HC-SEM method, HC (using Ward's method) generated a dendrogram, and villages were classified by optimizing the height-cutting tree; subsequently, SEM was used to verify the discriminant validity and cohesion of the obtained clusters. Both baseline models were assigned the same standardized feature dataset as the proposed method and underwent the same preprocessing steps to ensure fairness in comparison. These two baselines were chosen because they are representative: GMM represents a clustering method based on probability distribution, while HC represents a clustering method based on hierarchical connectivity. By comparing these different clustering paradigms, the superiority of the SOM-K-means fusion algorithm in classification performance can be fully demonstrated. The study first compared the classification accuracy of the three methods, and the test results are shown in Fig. 7.

Fig. 7(a) indicates the classification accuracy test outcomes of three approaches, and Fig. 7(b) indicates the recall test outcomes of the three approaches. In Fig. 7(a), the classification accuracy of the SKM-AHP-EWM method was positively correlated with data volume and increased linearly with increasing data volume. As the data volume increased from 50 to 250, accuracy gradually rose from 82.1% to 92.5%. The accuracy growth trend of the GMM-SEM and HC-SEM methods was the same as that of the research method, and their accuracy reached 84.2% and 76.3%, respectively, at a data volume of 250. As shown in Fig. 7(b), the recall rate of the SKM-AHP-EWM method increased linearly with the increase of data volume. At a data volume of 50, the recall rate reached 76.3% and then increased linearly to 87.1% as the data volume increased to 250. The recall trend of the HC-SEM method was the same as that of the research method, and when the data volume reached 250, its recall rate was only 71.8%. The findings demonstrated that the research method had better classification accuracy. Afterward, the computational efficiency of the three approaches was compared with the CPU usage time and convergence iteration times as indicators. The test outcomes are shown in Fig. 8.

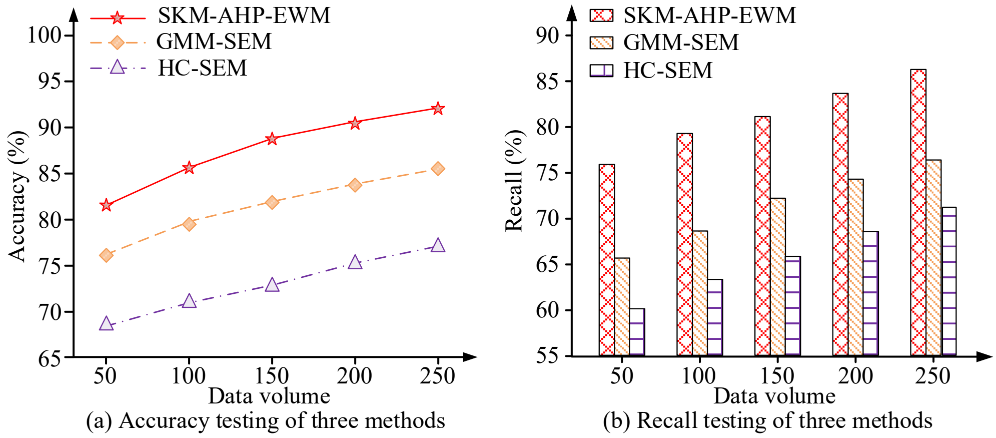


Fig. 7. Comparison of classification accuracy among three approaches

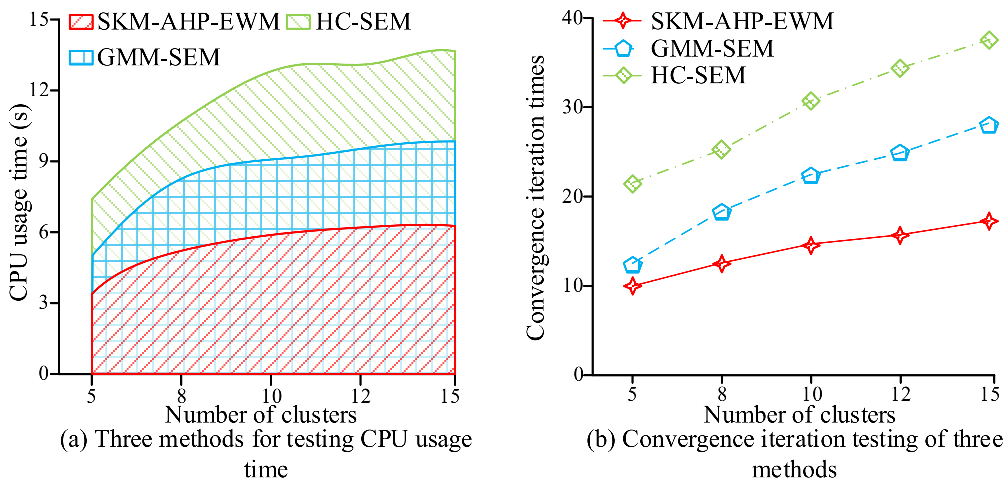
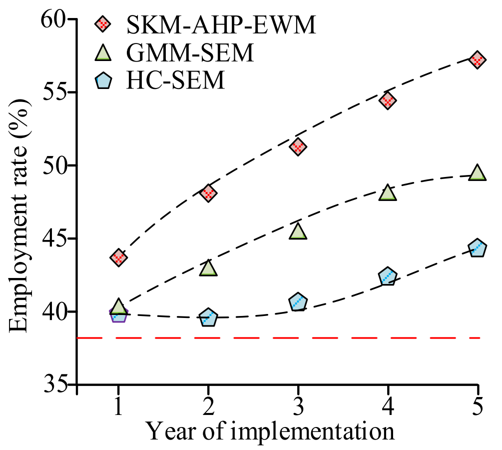


Fig. 8. Comparison of computational efficiency among three approaches

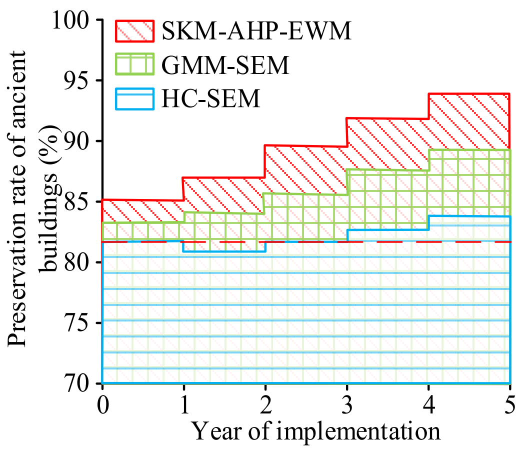
Fig. 8(a) indicates the CPU usage test outcomes of three approaches, and Fig. 8(b) shows the convergence iteration test outcomes of the three approaches. In Fig. 8(a), the CPU usage time for all three approaches was positively correlated with the number of clusters, and the SKM-AHP-EWM method optimized the initial centers through pre-clustering to reduce iteration redundancy, resulting in the slowest growth in CPU usage time. As the number of clusters increased from 5 to 15, CPU usage time gradually increased from 3.4 seconds to 6.5 seconds. The overall CPU usage time of the GMM-SEM method and the HC-SEM method was higher than that of the research method, and reached 10.2 seconds and 14.1 seconds, respectively, when the number of clusters was 15. From Fig. 8(b), the convergence iteration times of the three approaches all increased linearly with the increase of the number of clusters, while the SKM-AHP-EWM method could map high-dimensional samples to low-dimensional grids, providing an initial center close to the global optimum, so the iteration times were the least, and the growth is gentle. When the number of clusters increased from 5 to 15, the number of convergence iterations slowly increased from 10 to 16. The convergence iteration number growth trend of the HC-SEM method was the same as that of the research method, and when the number of clusters was 15, its iteration number reached 37. The findings demonstrated that the research method was more computationally efficient.

3.2. Simulation Testing of Fusion Algorithms and Development Planning Strategies

This section presents two categories of evaluations. (1) Simulation-based assessment of strategy effectiveness under hypothetical scenarios. (2) Empirical evaluations of clustering quality and robustness using the real-world village dataset. The study further validated the practical performance and effectiveness of a smart expressway evaluation system that integrated combined weighting methods, cloud models, and knowledge graphs in real-world applications. Due to the outflow of rural young adults, which leads to low initial employment rates, and the natural deterioration or improper renovation of ancient buildings resulting from a lack of systematic preservation efforts, the study compared the strategic effectiveness of the three approaches, using employment rates and ancient building preservation rates as indicators. This comparison was conducted via simulation, with initial values set to reflect typical rural conditions. The test outcomes are shown in Fig. 9.



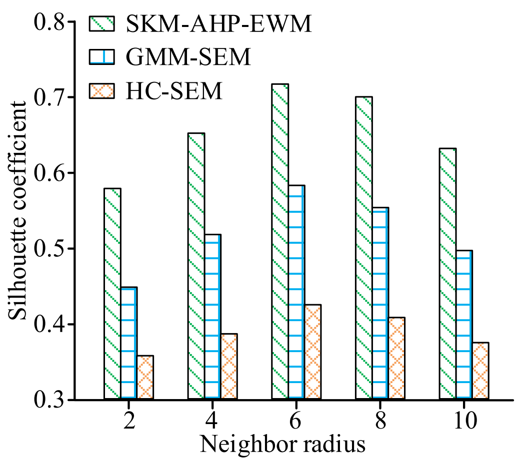
(a) Employment rate testing using three methods



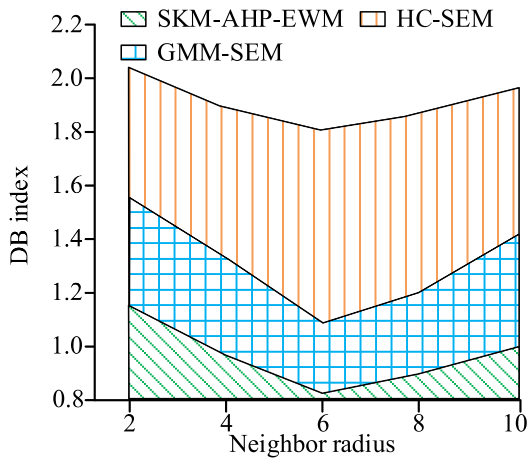
(b) Three methods for testing the preservation rate of ancient buildings

Fig. 9. Comparison of the effectiveness of three strategies

Fig. 9 (a) indicates the employment rate test outcomes of three approaches, and Fig. 9 (b) shows the ancient building preservation rate test outcomes of the three approaches. In Fig. 9(a), with an initial employment rate of 38%, the SKM-AHP-EWM method, due to the precise classification and multidimensional weighting of the SOM fusion algorithm, could adapt well to the characteristics of industries and villages, and its employment rate continued to grow steadily. When the strategy reached its fifth year of implementation, the highest employment rate achieved by the SKM-AHP-EWM method was 58.2%. As shown in Fig. 9(b), with an initial preservation rate of 82% for ancient buildings, the SKM-AHP-EWM method could allocate protection resources through AHP-EWM weighting, resulting in a steady increase in the preservation rate of ancient buildings. When the strategy reached its fifth year of implementation, the highest preservation rate for the research method was 93.7%. The preservation rate of the HC-SEM method was unstable, and in the second year of policy implementation, its preservation rate of ancient buildings decreased to 81.2%. The findings demonstrated that the research method was more effective. To empirically assess the clustering quality of the three methods on the actual dataset, the silhouette coefficient and Davies Bouldin index were computed. The test outcomes are shown in Fig. 10.



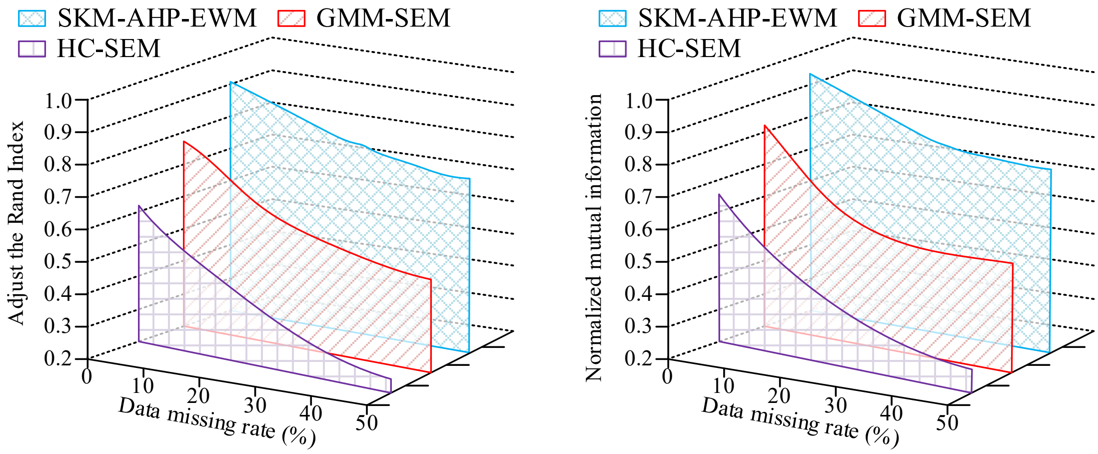
(a) Three methods of contour coefficient testing



(b) DB index testing using three methods

Fig. 10. Comparison of clustering quality among three approaches

Fig. 10(a) indicates the contour coefficient test outcomes of three approaches, and Fig. 10(b) indicates the Davies Bouldin index test outcomes of the three approaches. In Fig. 10(a), the contour coefficients of the three approaches exhibited an initial ascent followed by a descent as the neighborhood radius increased. When the neighborhood radius was 6, all three approaches reached their optimal values, and the highest contour coefficient was 0.72 for the SKM-AHP-EWM method. When the neighborhood radius was 2, the minimum contour coefficient of the HC-SEM method was 0.35. From Fig. 10(b), the Davies Bouldin indices of the three approaches all demonstrated a pattern of initial decline followed by an upswing as the neighborhood radius expanded, and the sorting indices of the SKM-AHP-EWM method were higher than those of the comparison method. When the neighborhood radius was 6, the clustering quality of the research method reached its optimum, with the lowest index of 0.82. The clustering performance of the GMM-SEM and HC-SEM methods was also optimal when the neighborhood radius was 6, with values of 1.12 and 1.84, respectively. The findings demonstrated that the research method achieved higher quality clustering. The robustness of the three approaches was empirically evaluated by artificially introducing missing data under the MCAR mechanism and applying mean imputation. The test outcomes are presented in Fig. 11.



(a) Three methods for adjusting the Rand index test

(b) Normalized mutual information testing of three methods

Fig. 11. Comparison of robustness among three approaches

Fig. 11(a) shows the adjusted Rand index test outcomes of three approaches, and Fig. 11(b) shows the normalized mutual information test outcomes of the three approaches. As shown in Fig. 11(a), the curves for the three approaches decreased with increasing data loss rate, but the SKM-AHP-EWM method showed the smoothest decline, whereas the HC-SEM method showed the steepest. When the data missing rate was 0%, the adjusted Rand index of the SKM-AHP-EWM method reached 0.91. Afterward, as the data missing rate increased to 50%, its index slowly decreased to 0.71. From Fig. 11(b), the normalized mutual information index for the three approaches was negatively correlated with the data loss rate, while the SKM-AHP-EWM method showed a slight decrease compared to the comparison method. When the data loss rate was 0%, the index reached 0.93, and at 50% data loss, the research method still maintained a normalized mutual information index of 0.68 or higher. The findings demonstrated that the research method was more robust.

4. Conclusion

To solve the problems of inaccurate classification and homogenized strategies in the protection and development of traditional villages, an innovative development planning strategy system based on the SOM fusion algorithm and combined weighting was developed. In terms of research methods, a SOM-K-means fusion algorithm that combines SOM and K-means was proposed to classify villages into five development types. By combining AHP and EWM for weighting, a village development evaluation system was constructed to identify development shortcomings and formulate precise strategies. The experimental findings demonstrated that when the data volume increased to 250, the recall rate linearly increased to 87.1%. When the number of clusters was 15, the number of iterations required to converge was 16. In practice, when the strategy's implementation reached its fifth year, the highest preservation rate for the research method was 93.7%. When the neighborhood radius was 6, the lowest exponent of the research method was 0.82. When the data missing rate was 0%, its normalized mutual information index reached 0.93. In summary, the research method outperformed the comparative method in terms of classification accuracy, computational efficiency, strategy effectiveness, and robustness. However, the SOM algorithm was limited by the preset grid structure and lacked flexibility in class boundaries. In combination with weighting, SWs relied on expert experience and might have certain biases. In the future, more flexible spectral clustering algorithms will be introduced to further enhance boundary adaptability, while using the Delphi method or group decision optimization to obtain SWs.

Author Contributions

Kai Qin contributed to conceptualization, methodology, data collection, project administration, and draft preparation. Cuixia Li contributed to software, analysis, data collection, supervision, and manuscript editing. Qian Wang contributed to analysis, supervision, project administration, and manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors used AI tools solely for language editing and readability improvement. The authors reviewed and verified all content and take full responsibility for the accuracy and integrity of the manuscript.

References

Allawi, A. H. and Al-Jazaeri, H. M. J. (2023). A new approach towards the sustainability of urban-rural integration: The development strategy for central villages in the Abbasiya District of Iraq using GIS techniques. *Regional Sustainability*, 4(1), 28-43. doi: 10.1016/j.regsus.2023.02.004

Bai, X., Fan, Y., Hou, J., and Liu, Y. (2024). A comprehensive evaluation method for reliability confidence capacity of renewable energy based on improved DEMATEL-AHP-EWM. *Journal of Electrical Engineering & Technology*, 19(3), 1205-1216. doi: 10.1007/s42835-023-01652-3

Hebbi, C. and Mamatha, H. (2023). Comprehensive Dataset Building and Recognition of Isolated Handwritten Kannada Characters Using Machine Learning Models. *Artificial Intelligence and Applications*, 1(3), 179-190. doi: 10.47852/bonviewAIA3202624

Hua, Z. and Anderson-Frey, A. K. (2022). Self-organizing maps for the classification of spatial and temporal variability of tornado-favorable parameters. *Monthly Weather Review*, 150(2), 393-407. doi: 10.1175/MWR-D-21-0168.1

Jhong, Y. D., Lin, H. P., Chen, C. S., and Jhong, B. C. (2022). Real-time neural-network-based Ensemble Typhoon Flood forecasting model with self-organizing map cluster analysis: a Case Study on the Wu River Basin in Taiwan. *Water Resources Management*, 36(9), 3221-3245. doi: 10.1007/s11269-022-03197-y

Ji, Y., Xie, H., Shi, S., He, P., Jin, N., and Wang, H. (2023). Voltage and Reactive Power Combinational Evaluation of Regional Power Grid Based on EWM-AHP-BP Neural Network. *Journal of System Simulation*, 35(4), 843-852. doi: 10.16182/j.issn1004731x.joss.21-1329

Li, D., Wang, Y., and Lu, L. (2023). Sustainable livelihood effect of integrated rural tourism development under the goal of common prosperity: Based on a case study of Lujia village, Anji county, Zhejiang province. *Journal of Natural Resources*, 38(2), 511-528. doi: 10.31497/zrzyxb.20230215

Liew, S. J., Ng, S. C., Mustapa, M. Z., Usop, Z., Fauthan, M. A., Mahalil, K. B., and Tan, C. O. (2023). Colour sorting of red oak, yellow poplar and maple veneers using self-organizing map: comparisons between different camera genres. *European Journal of Wood and Wood Products*, 81(3), 777-789. doi: 10.1007/s00107-022-01900-9

Liu, F., Wang, F., Ding, Y., and Yang, S. (2022). SOM-based binary coding for single sample face recognition. *Journal of Ambient Intelligence and Humanized Computing*, 13(12), 5861-5871. doi: 10.1007/s12652-021-03255-0

Liu, W., Cui, W., Chen, M., Hu, Q., and Song, Z. (2023). Operation safety risk assessment of water distribution networks based on the combined weighting method (CWM). *KSCE Journal of Civil Engineerin*g, 27(5), 2116-2130. doi: 10.1007/s12205-023-1552-4

Meng, C., Song, Y., Ji, J., Jia, Z., Zhou, Z., Gao, P., and Liu, S. (2022). Automatic classification of rural building characteristics using deep learning methods on oblique photography. *Building Simulation*, 15(6), 1161-1174. doi: 10.1007/s12273-021-0872-x

Mohammadi, M. and Mobarakeh, M. I. (2022). An integrated clustering algorithm based on firefly algorithm and self-organized neural network. *Progress in Artificial Intelligence*, 11(3), 207-217. doi: 10.1007/s13748-022-00275-5

Ouzerbane, Z., Loulida, S., El Hmaidi, A., Essahlaoui, A., Boughalem, M., Ousmana, H., and Berrada, M. (2022). Study of the salinity of groundwater in the HAHA syncline by the Kohonen self-organized classification (Essaouira, Morocco). *Environmental Science and Pollution Research*, 29(9), 13592-13611. doi: 10.1007/s11356-021-16598-0

Sakkari, M., Hamdi, M., Elmannai, H., AlGarni, A., and Zaied, M. (2022). Feature extraction-based deep self-organizing map. *Circuits, Systems, and Signal Processing*, 41(5), 2802-2824. doi: 10.1007/s00034-021-01914-3

Tissot, H. (2025). FormulAI: Designing rule-based datasets for interpretable and challenging machine learning tasks. *Artificial Intelligence and Applications*, 3(1), 72-82. doi: 10.47852/bonviewAIA42021781

Wang, S., Wang, J., Shen, W., and Wu, H. (2023). The evaluation of tourism service facilities in Chinese traditional villages based on the living protection concept: theoretical framework and empirical case study. *Journal of Asian Architecture and Building Engineering*, 22(1), 14-31. doi: 10.1080/13467581.2021.2007109

Wen, Z. (2023). Evaluation of heterogeneity in tectonically deformed coal reservoirs based on the analytic hierarchy process-entropy weight method coupling model: a case study. *ACS Omega*, 8(40), 36700-36709. doi: 10.1021/acsomega.3c02764

Wiyono, S. H., Subianto, A., and Nuhman, N. (2023). Sustainable ecotourism development and community empowerment: A case study of the center for environmental education in seloliman village, Indonesia. *Society*, 11(2), 310-328. doi: 10.33019/society.v11i2.528

Xiao, M., Luo, S., and Yang, S. (2025). Synergizing Technology and Tradition: A Pathway to Intelligent Village Governance and Sustainable Rural Development. *Journal of the Knowledge Economy*, 16(1), 1768-1823. doi: 10.1007/s13132-024-01937-6

Zhang, X., Liu, Y., Lin, Y., Liao, Q., and Li, Y. (2024). Uv-sam: Adapting segment anything model for urban village identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20), 22520-22528. doi: 10.1609/aaai.v38i20.30260

Kai Qin obtained his master's degree in architecture from Beijing Jiaotong University in 2021. Currently, he serves as an architecture teacher at the College of Architecture and Engineering of Huanghuai University. He has participated in numerous rural planning and design projects and practices. He is a registered architect of the People's Republic of China. He has published several academic research papers on rural topics. He is currently studying rural village layout, rural system planning, and the renovation and transformation of traditional rural buildings.

Cuixia Li obtained her master's degree in veterinary medicine from the Chinese Academy of Agricultural Sciences in 2018. Currently, she serves as a professional teacher of animal husbandry and veterinary medicine at the College of Animal Husbandry Engineering, Zhumadian Agricultural Engineering Vocational College. She is primarily engaged in teaching and scientific research in animal husbandry and veterinary medicine, focusing on agricultural technology, rural animal husbandry development and related fields. She has participated in a number of education and teaching reform projects, published numerous academic papers on animal husbandry and veterinary medicine, and has long been committed to research on promoting veterinary technology and rural industry development.

Qian Wang obtained her master's degree in urban and Rural Planning from Kunming University of Science and Technology in 2019. Currently, she serves as an urban and rural planning teacher at the Architecture and Engineering College of Huanghuai University. She has participated in numerous rural planning and design projects and practices, including traditional village protection planning and rural renovation projects. She has published several academic research papers and a book on rural planning, smart cities, and heritage protection. She is currently engaged in the study of rural settlement protection, digital technology applications in urban-rural planning, and heritage site conservation and development.