IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 September 2023, accepted 1 October 2023, date of publication 4 October 2023, date of current version 11 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3321894

RESEARCH ARTICLE

Clustering Analysis of Integrated Rural Land for Three Industries Using Deep Learning and Artificial Intelligence

QIAN HUANG[1,2], HAIBIN XIA[2,3], AND ZHANCHENG ZHANG[1]

[1]College of Art and Design, Guangdong Polytechnic, Foshan 528000, China
[2]Doctor of Philosophy in Education/Arts, University of Perpetual Help System DALTA, Las Piñas, Manila 1740, Philippines
[3]Department of Digital Media Art, College of Arts, East China Jiaotong University, Nanchang 330013, China

Corresponding author: Zhancheng Zhang (zcz08200@gmail.com)

This work was supported by Research on the Zero-Point Paradigm of Industrial Integration Agricultural Product Market led by Art Design under Grant XJKY202267.

ABSTRACT This study employs deep learning and artificial intelligence (AI) clustering analysis techniques to evaluate the suitability of integrated rural land for three industries. Diverse datasets pertaining to rural development, encompassing land use, agricultural production, and rural tourism, are gathered and harmoniously amalgamated. An innovative land suitability assessment model, merging ResNet-50 with the k-means algorithm, is devised. Specifically, ResNet-50 is harnessed for the classification and recognition of rural land-use images, thus deriving feature vectors for each sample. These feature vectors are subsequently fed into the k-means algorithm to cluster samples with akin land-use patterns. The ensuing examination of land use composition within each cluster facilitates the evaluation of rural land's suitability for three-industry integration. Experimental scrutiny discloses that this study achieves an accuracy rate of 88.3% in rural land-use classification and recognition, outperforming alternative algorithms by at least 3.1%. Furthermore, it yields an average intersection over union (IoU) of 67.29%. Remarkably, the k-means algorithm exhibits superior clustering outcomes. Consequently, the model introduces herein demonstrated substantial enhancements in rural land-use classification and recognition accuracy, average IoU, and clustering performance. It offers an innovative tool for policymakers to advance rural industry integration, fostering economic diversification. Additionally, this model aids decision-makers in identifying prospective opportunities and challenges, thus facilitating the formulation of forward-thinking and viable rural development strategies.

INDEX TERMS Artificial intelligence, deep learning, integrated rural land for three industries, cluster, suitability evaluation.

I. INTRODUCTION

A. RESEARCH BACKGROUND AND MOTIVATIONS

As social productivity continues to advance, Chinese agriculture is undergoing a transition towards modernization. The conventional rural economic framework, primarily centered on agriculture, has encountered challenges, including uneven resource utilization and a narrow industrial focus, which fall short of meeting the evolving expectations for a quality living environment and diverse industrial requirements [12], [30].

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

Against the backdrop of population urbanization, mounting resource and environmental pressures, and the imperative of rural economic transformation and upgrading, the concept of rural tri-sector integration has emerged as a prominent and extensively discussed strategy. Its primary objectives encompass stimulating rural revitalization and achieving sustainable development.

The concept of rural tri-sector integration pertains to the organic and coordinated development of agriculture, rural industries, and rural tourism within a shared geographical space. This model harnesses the distinctive strengths of these sectors, facilitating resource sharing and complementarity, thereby augmenting the diversity and comprehensive advantages of rural development [8]. Nonetheless, during the implementation of rural tri-sector integration, a crucial challenge arises: how to judiciously manage land resources and ascertain the appropriateness of land for integrated use [16], [28]. This study explores the utilization of deep learning techniques to process extensive land-use data, enabling the extraction of crucial features pivotal for assessing integrated land use [18], [25]. Furthermore, it investigates the application of artificial intelligence (AI) clustering analysis methods to classify diverse land-use data types, thereby furnishing essential support for subsequent evaluations [23].

This study addresses several pivotal concerns. Firstly, it recognizes the imperative for rural areas to achieve economic diversification and sustainable development by integrating primary, secondary, and tertiary industries. As urban-rural disparities diminish, rural regions must amalgamate agriculture, industry, and services to achieve comprehensive industrial development. Consequently, investigating the suitability of rural land becomes paramount to facilitating the integration of these sectors. Secondly, the study acknowledges the remarkable strides made in deep learning and AI technologies, particularly in image processing, pattern recognition, and data analysis. Leveraging these advanced technologies, which excel in handling large-scale, multi-source data, is expected to enhance the precision and efficiency of land suitability assessments. Thirdly, the study recognizes the need for scientific land suitability assessment methods to guide policy formulation and strategies aimed at propelling rural economic growth. This assessment extends beyond agriculture and encompasses areas like rural tourism and the service industry. Thus, the development of a decision-support-oriented land suitability assessment model is imperative for informed decision-making by governments and stakeholders.

B. RESEARCH OBJECTIVES

The objective of this study is to assess the appropriateness of rural land for tri-sector integration employing deep learning and AI cluster analysis. By systematically collecting and processing data pertinent to rural development, coupled with the application of deep learning and cluster analysis techniques, the study endeavors to precisely discern distinct categories of integrated land, including their attributes, and gauge their suitability and developmental prospects. This endeavor serves as a scientific foundation and point of reference for rural development planning and decision-making and contributes to the advancement of rural revitalization and the realization of sustainable development goals. Furthermore, the methodologies and findings presented in this study hold substantial theoretical and practical significance in terms of optimizing land resource allocation and propelling the transformation and elevation of rural economies.

This study introduces a pioneering land suitability assessment model that merges deep learning and clustering analysis methodologies. Employing ResNet-50 for image classification and recognition, in conjunction with the k-means algorithm for clustering analysis, this model adeptly evaluates rural land's suitability for seamless integration of primary, secondary, and tertiary industries. This approach's innovation stems from applying advanced AI technologies to the domain of land assessment, significantly enhancing assessment precision and efficiency. This study presents a comprehensive and scientifically rigorous approach to evaluating land's multifaceted potential, empowering decision-makers to gain a deeper understanding of rural land's attributes and prospects. This innovative method provides invaluable support for the harmonization of rural industries and the cultivation of sustainable development. Decision-makers can utilize this model to identify latent developmental opportunities and challenges, facilitating the formulation of forward-looking and viable rural development strategies ultimately advancing the enduring prosperity of rural landscapes.

II. LITERATURE REVIEW

A. REVIEW OF RELATED RESEARCH IN THE FIELD OF RURAL LAND USE EVALUATION

In the realm of rural development and land utilization assessment, a multitude of research findings have emerged, all bearing relevance to the theme of rural tri-sector integration and suitability appraisal. Wei et al. [27] analyzed the spatiotemporal characteristics and driving forces behind land marketization in Shaanxi Province, uncovering noteworthy temporal and spatial disparities in marketization levels. Li et al. [17] pioneered the formulation of an index framework for sustainable rural development grounded in the concept of ecological livability. This framework offers a comprehensive evaluation encompassing ecological, societal, and economic factors, providing an effective means to assess the sustainability of rural development. Soleimani et al. [24] harnessed Monte Carlo simulation and sensitivity analysis to gauge groundwater quality and nitrate risk, revealing nitrate concentrations in groundwater as a hazard influenced by diverse factors. Ghayour et al. [9] leveraged machine learning algorithms to assess the performance of Sentinel-2 data in land cover/use classification, culminating in commendable accuracy and consistency. Wang et al. [26] examined the potential contributions of rural revitalization by delineating the structure of rural regional systems, thus supplying crucial reference points and guidance for rural development planning and decision-making.

Upon scrutinizing the aforementioned literature, it becomes evident that they share a common focus on rural development and land use assessment, albeit with variations in their respective emphases, methodologies, and levels of depth. Notably, there exists a dearth of all-encompassing and integrated research efforts, which in turn impedes the efficacious resolution of challenges related to the integration of the rural tri-sector. This context offers a theoretical foundation for the suitability evaluation of integrated rural land for three industries as presented in this study, bearing inherent innovation and practical applicability.

B. REVIEW ON THE APPLICATION OF DEEP LEARNING IN LAND USE EVALUATION

In recent years, there has been a growing interest in the application of deep learning and artificial intelligence (AI) to land use evaluation. Deep learning technology offers robust pattern recognition capabilities, with its capacity to acquire and distill features from vast datasets through the construction of intricate neural network models. This technological advancement has garnered substantial attention from researchers. For instance, Debella-Gilo and Gjertsen [6] employed deep learning methods to map seasonal agricultural land use types, yielding results distinguished by their high accuracy and practical applicability. Likewise, Masolele et al. [19] utilized deep learning techniques to infer land use changes following deforestation, demonstrating the method's ability to accurately identify post-deforestation land use types and uncover trends in land use alterations. Castelo-Cabay et al. [4] harnessed deep learning to classify land use and land cover, ultimately showcasing its efficacy and accuracy. Zhu et al. [32] introduced a land use/land cover change detection approach tailored for high spatial resolution remote sensing images, founded on a twin global learning framework, achieving precise detection of land use/land cover changes. Boonpook et al. [2] successfully employed deep learning for the semantic segmentation of various land use and land cover types, revealing the method's capability to accurately delineate diverse land use and land cover categories.

In conclusion, these studies underscore the potential of deep learning in land use assessment, highlighting its capacity to enhance classification accuracy and detect land use changes effectively. Nevertheless, a noticeable research gap exists within the domain of rural land suitability assessment, especially concerning the context of three-industry integration. There is a dearth of methodologies that amalgamate deep learning with suitability assessment. Consequently, this study's focal point resides in evaluating rural land suitability for integrating the three industries, representing a specific and innovative research avenue. By integrating deep learning and cluster analysis, this study introduces a novel approach that offers substantial support for rural development and land planning.

C. REVIEW ON THE APPLICATION OF THE CLUSTER METHOD IN LAND USE

Cluster analysis algorithms play a pivotal role in categorizing data samples into distinct groups, effectively distinguishing similar samples from dissimilar ones. In the realm of land use evaluation, cluster analysis methods serve to identify various types and characteristics of integrated land use—a topic that has garnered significant scholarly attention. Abera et al. [1] explored the influence of clustering algorithms on ecosystem services when applied to the dynamics of land use and land cover in forest biosphere reserves. This approach effectively categorized water source protection, soil conservation, and carbon storage within ecosystems. Seaton et al. [22] leveraged cluster analysis to group soil health based on national soil indicator monitoring data. The outcomes revealed distinct cluster patterns in different regions, reflecting spatial variations in soil health. Giao et al. [10] employed remote sensing and multivariate statistical techniques to examine the relationship between land use patterns and water quality in Jiangyuan Province, Vietnam. The findings demonstrated a discernible correlation between land use patterns and water quality. Lastly, Kalisz et al. [15] delved into the utilization of land use indicators in assessing land use efficiency, concluding that judiciously selected land use indicators could effectively gauge land use efficiency.

In summary, these documents underscore the wide-ranging application of cluster analysis in land use evaluation, particularly in its capacity to discern diverse types, characteristics, and relationships within data. However, there exists a noticeable research gap in the assessment of clustering analysis algorithms for the suitability of integrated rural land use for three industries. Consequently, this study's distinctive contribution lies in the integration of deep learning with cluster analysis to assess the suitability of integrated rural land for three industries, presenting a novel and innovative approach within the realm of rural sustainable development.

D. SUMMARY

In conclusion, while the studies conducted by the aforementioned scholars have yielded valuable insights in the domain of rural development and land use assessment, they still exhibit certain limitations. Consequently, the novelty and distinctiveness of this study lie in the fusion of deep learning and artificial intelligence cluster analysis, as applied to the evaluation of the suitability of integrated rural land for three industries. By surmounting the constraints of conventional methodologies, this approach promises to deliver more precise, scientifically grounded, and objective outcomes in land use evaluation, thereby offering innovative support for rural sustainable development and land planning. The innovative significance of this method in the realm of rural development and land use assessment cannot be overstated, as it introduces a fresh perspective and solution to address practical challenges and future endeavors.

III. RURAL LAND SUITABILITY EVALUATION THROUGH THE INTEGRATION OF DEEP LEARNING AND AI CLUSTERING ALGORITHMS

A. ANALYSIS OF INTEGRATION OF THREE INDUSTRIES

The amalgamation of three industries entails the integration of agriculture, industry, and the service sector, with the optimization of land use serving as a catalyst for the harmonious development of these distinct sectors. Within the context of the modern era, the amalgamation of these rural industries has emerged as the definitive path for China to transition from traditional agriculture to the sustainable development of cutting-edge, technology-driven, and service-oriented industries. This paradigm shift has propelled the realization and implementation of China's rural revitalization strategy [31]. Figure 1 illustrates the developmental framework underpinning the integration of rural tertiary industries.

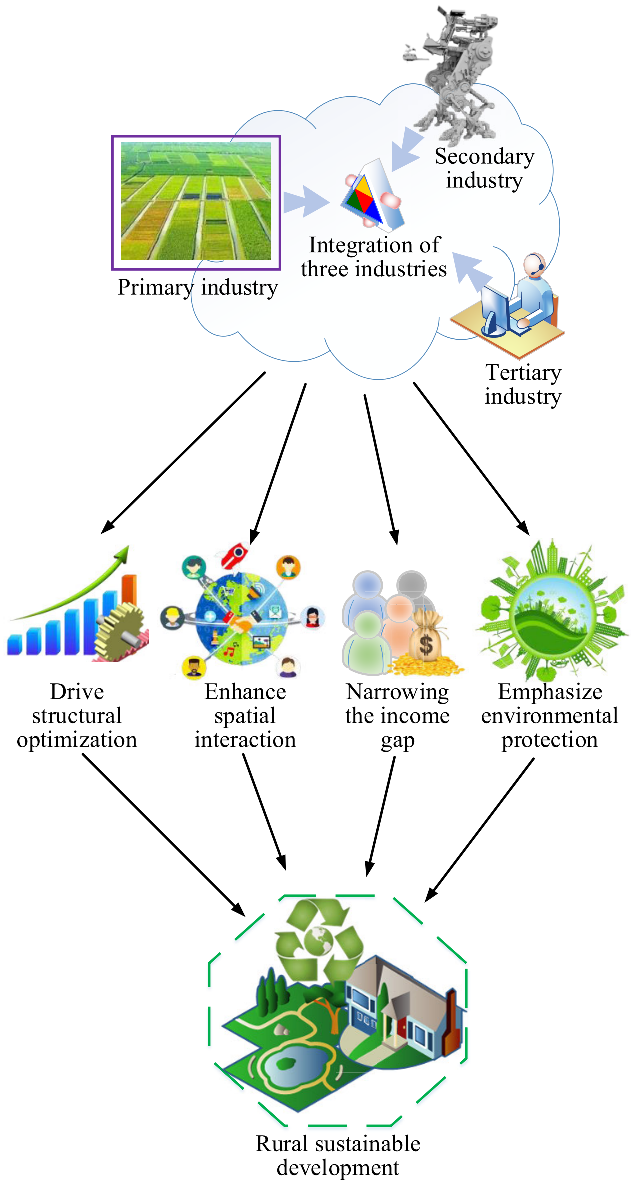

FIGURE 1. Schematic representation of the developmental framework driving the integration of rural tertiary industries.

In Figure 1, several noteworthy advantages emerge following the integration of three rural industries. Firstly, land resources can be harnessed more effectively, resulting in enhanced land use efficiency and the maximization of resources. This integration also facilitates the optimization and upgrading of the industrial structure. Secondly, it fosters the diversified development of the rural economy by encouraging synergy among industries, mitigating reliance on any single sector, and bolstering economic stability and resilience. Furthermore, the integration of different industries generates increased employment opportunities, elevates farmers' income levels, and enhances the living standards of rural residents. Consequently, it contributes to narrowing the income gap between urban and rural areas. Lastly, the amalgamation of the three industries has a positive impact on the rural environment, leading to overall improvements and creating a more appealing living environment for rural residents.

B. DEEP LEARNING AND ITS APPLICATION IN LAND TYPE IDENTIFICATION AND ANALYSIS

Deep learning is a machine learning technique that employs artificial neural network models to simulate and acquire data feature representations, enabling the recognition of intricate patterns and advanced abstractions within the data. Deep learning technology proves valuable in data analysis and prediction, providing decision-making support for assessing the suitability of integrated rural land for three industries. One commonly employed deep learning model, the Convolutional Neural Network (CNN), particularly excels in image classification tasks [7]. The utilization of CNN for land type identification is depicted in Figure 2.

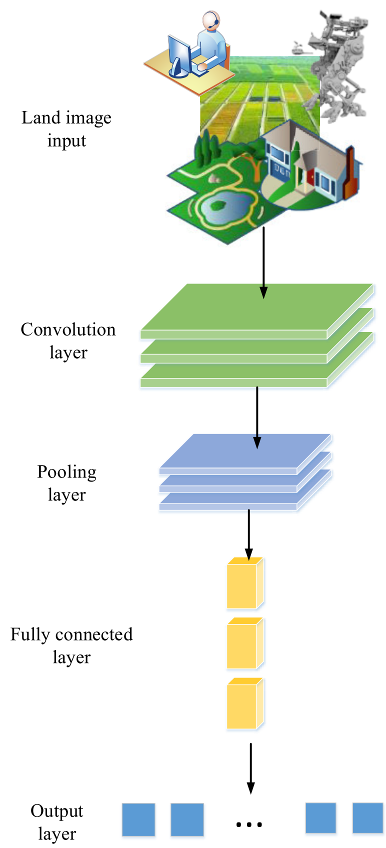

FIGURE 2. Illustration of CNN's application in land type identification.

In Figure 2, CNNs demonstrate exceptional suitability for processing data characterized by spatial structures, including images and geographic information. The convolution layer, within this framework, employs convolution operations to extract features from localized areas, while the pooling layer serves to diminish the dimensionality of the feature map while preserving key features. The process of stacking multiple convolution layers and pooling layers results in the extraction of increasingly abstract, higher-level features. This hierarchical feature extraction enables precise data classification and recognition.

Within the CNN algorithm, ResNet50, a variant of ResNet, holds significance [5]. In the domain of land type recognition, ResNet-50 offers distinct advantages, characterized by its depth, residual connections, and training on extensive image datasets. These attributes equip it to effectively capture the intricate characteristics of land, thereby enhancing classification accuracy and stability. The fusion of ResNet-50 with meticulous data preprocessing and feature extraction makes the realization of a more precise and resilient land type recognition model attainable. ResNet50 involves five down-sampling operations, with the second employing maximum pooling to halve the feature map size, while the subsequent four down-sampling stages employ a convolution step size of 2, effectively extracting land type features while progressively reducing the feature map size, as illustrated in Figure 3.

As depicted in Figure 3, when ResNet50 is employed for image feature extraction, it comprises four residual modules, each composed of a convolution layer and a pooling layer. Notably, the last two residual modules do not employ down-sampling. Within this framework, the Batch Normalization layer is commonly utilized to adjust the output data distribution from the convolution layer, thereby expediting convergence [11]. Suppose that the input of a batch at a specific neural network layer is represented as $X = [x_0, x_1 \cdots, x_n]$, where x_i denotes a rural land sample, and n signifies the batch size. To begin, the mean value μ_B of the elements within the mini-batch is computed, as indicated in Equation (1).

$$\mu_B = \frac{1}{n}\sum\nolimits_{i=1}^{n} x_i \tag{1}$$

Next, the variance σ_B^2 of the mini-batch is determined, as illustrated in Equation (2).

$$\sigma_B^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n} (x_i - \mu_B)^2 \tag{2}$$

In this manner, each element can be normalized, as depicted in Equation (3).

$$x_i' = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{3}$$

In Equation (3), ε represents a constant that prevents the denominator from being 0. After performing the aforementioned operations, the data is transformed into a normal distribution with a mean of 0 and a variance of 1, resulting in the loss of data offset. To revert to the data as it was before

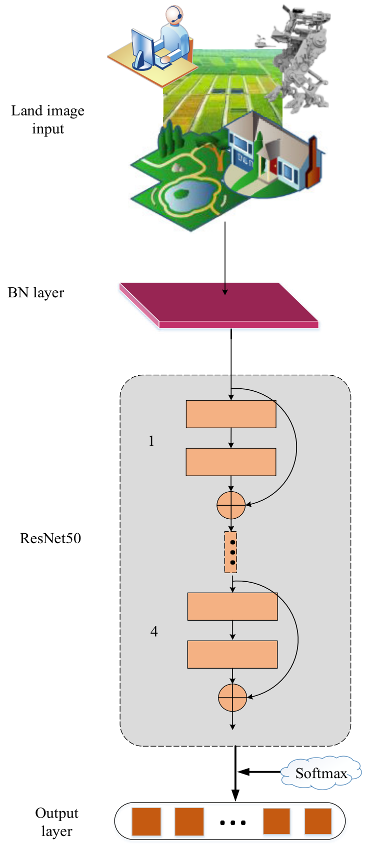

FIGURE 3. Schematic representation of ResNet50's application in land type identification.

applying Batch Normalization, an identity transformation is required, as demonstrated in Equation (4).

$$y_i = \gamma x_i' + \beta \tag{4}$$

In Equation (4), y_i represents the ultimate output of the network, while γ and β denote the variance and offset of the input data distribution, respectively. In networks lacking a Batch Normalization layer, these two values are associated with the nonlinear characteristics introduced by the preceding layer. However, following transformation, they become independent of the previous layer and instead serve as learning parameters for the current layer. This adjustment is advantageous for optimization and does not compromise network capacity. During testing, the BN operation, denoted as x_{ic}', employs unbiased estimators of the mean value, $E(x)$, and variance, $Var(x)$, recorded during each Mini-batch. The final output is designated as y_{ic}, as outlined in Equations (5) to (8).

$$E(x) = E_B(\mu_B) \tag{5}$$

$$Var(x) = \frac{m}{m-1}E_B(\sigma_B^2) \tag{6}$$

$$x_{ic}' = \frac{x_i - E(x)}{\sqrt{Var\,(x) + \varepsilon}} \tag{7}$$

$$y_{ic} = \gamma x_{ic}' + \beta \tag{8}$$

In the context of employing ResNet50 for land type identification, enhancing the model's fitting capability can be achieved by solely increasing its depth through the use of convolutional kernel residual structures. It is advisable to omit the final two down-sampling layers to reduce computational demands. Opting to remove the initial two down-sampling layers may result in minor alterations to the feature map recognition network.

When assessing the performance of a learning algorithm, typical metrics such as accuracy, precision, recall, and F1 value are commonly employed to gauge the accuracy of land type recognition. The precise equations for these metrics are presented in Equations (9) through (12).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{\sum_{i=0}^{d} p_{ii}}{\sum_{i=0}^{d}\sum_{j=0}^{d} p_{ij}} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} = \frac{\sum_{i=0}^{d} p_{ii}}{\sum_{i=0}^{d} p_{ii} + \sum_{i=0}^{d}\sum_{j=0,j\neq i}^{d} p_{ij}} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} = \frac{\sum_{i=0}^{d} p_{ii}}{\sum_{i=0}^{d} p_{ii} + \sum_{i=0}^{d}\sum_{j=0,j\neq i}^{d} p_{ji}} \tag{11}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

Here, *TP* represents the count of true positive samples correctly predicted as positive. *FP* stands for the count of negative samples inaccurately predicted as positive. *FN* denotes the count of positive samples erroneously predicted as negative. *TN* signifies the count of negative samples correctly predicted as negative. Here, d signifies the number of distinct land types classified, p_{ii} represents the count of pixels accurately classified for each land type, p_{ij} represents the count of pixels belonging to the i-th land type but classified as the j-th land type by the model (indicating false positives), and p_{ji} refers to false negative pixels.

The mean Intersection over Union (IoU) serves as another commonly employed metric. It quantifies the intersection-to-union ratio between the model's predicted area and the expected output area for a specific land type, as visually depicted in Figure 4.

As depicted in Figure 4, it represents the IoU ratio between the model's output area and the expected output area. Consider a certain land type area as an ellipse A, and the model's prediction result as a rectangle B. The degree of overlap between ellipse A and rectangle B determines the closeness of the intersection and union areas and, consequently, the proximity of the IoU value to 1. Conversely, if rectangle A and rectangle B have no overlap whatsoever, the IoU value approaches 0. A higher IoU value signifies a more accurate model classification result. The mean intersection and union ratio can be calculated as the mean IoU value across all land types, as shown in Equation (13).

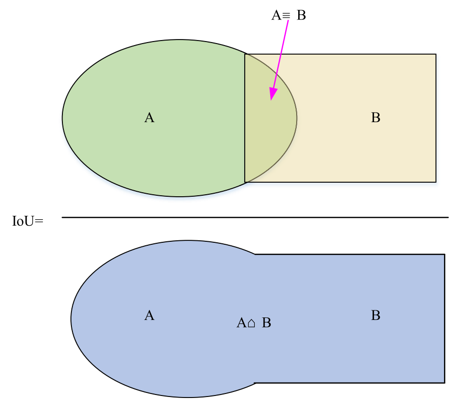

FIGURE 4. Visualisation of intersection and parallel ratio.

$$MIoU = \frac{1}{d+1}\sum_{i=0}^{d} \frac{p_{ii}}{\sum_{j=0}^{d} p_{ij} + \sum_{j=0}^{d} p_{ji} \cdot p_{ii}} \tag{13}$$

In Equation (13), d represents the number of land types being classified. p_{ii} corresponds to the number of pixels correctly classified for each land type, while p_{ij} and p_{ji} represent *FP* and *FN* samples, respectively.

C. APPLICATION ANALYSIS OF CLUSTERING ALGORITHM IN LAND FUNCTIONAL ZONING

Typically, rural land use involves a complex and diverse set of indicators. In the suitability evaluation of integrated rural land for three industries, cluster analysis plays a crucial role in classifying and identifying similar objects by establishing data similarity. When applying cluster analysis to the study of land use function zoning, it becomes possible to objectively perform land use function zoning based on the similarities among different rural land uses, thereby revealing the correlations and distinctions between different integrated land types. In this study, the primary cluster methods employed are K-means and the Gaussian mixture model. The optimal clustering scheme is then determined based on the Calinski-Harabasz (CH) coefficient and contour coefficient.

In the k-means algorithm [14], the process begins by initializing k clustering centers. Each sample is then assigned to the nearest clustering center. After this initial classification, the clustering centers are updated by calculating the centroid of the samples within each cluster. This iterative process continues until the clustering centers no longer change or the maximum number of iterations is reached. The detailed steps are depicted in Figure 5.

In Figure 5, the initial step involves determining the number of clusters (k). Subsequently, k samples are randomly chosen to serve as the initial cluster centers. The Euclidean distance between each sample and the C cluster centers is then calculated, as represented in Equation (14).

$$d_{ij} = \left\| x_i - \mu_j \right\| \tag{14}$$

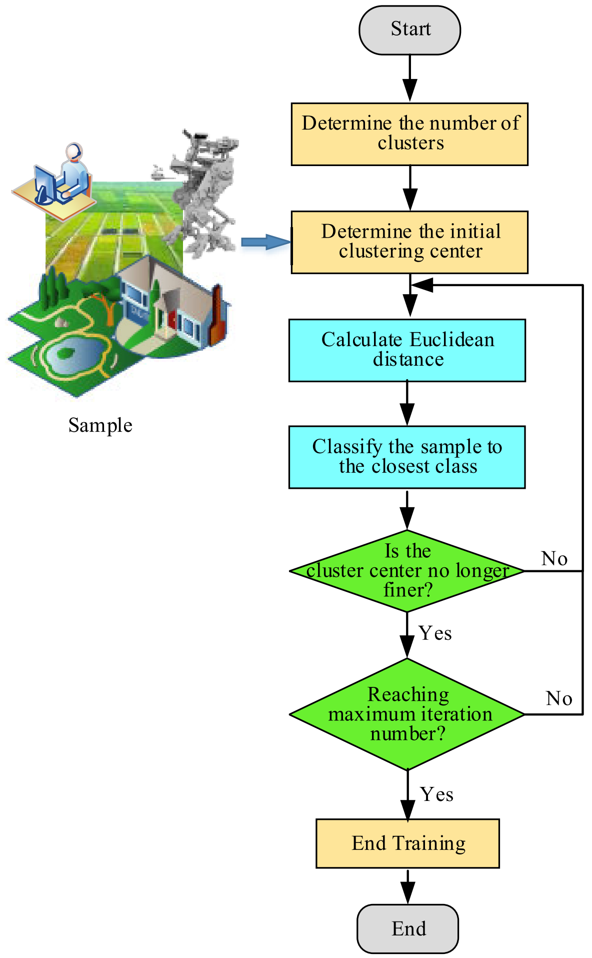

FIGURE 5. **Flow chart of land use classification under k-means algorithm.**

In Equation (14), d_{ij} denotes the Euclidean distance between the sample i and the cluster center j. x_i represents the coordinates of sample i. μ_j represents the coordinates of cluster center j. The sample is assigned to the nearest cluster, C_{λ_j}, as shown in Equation (15).

$$\lambda_j = argmin_i d_{ij} \tag{15}$$

In Equation (15), λ_j represents the cluster to which the sample is assigned, and d_{ij} refers to the Euclidean distance from sample i to cluster center j. The cluster center undergoes continuous updates, as demonstrated in Equation (16).

$$\mu_j' = \frac{1}{|C_j|}\sum\nolimits_{i\in C_j} x \tag{16}$$

Here, μ_j' refers to the centroid of the updated cluster j, and C_j signifies the j-th cluster.

The Euclidean distance of the cluster center for each sample is calculated iteratively until the cluster center remains unchanged or the maximum number of iterations is reached. The clustering result for each sample point is then returned.

While the K-means algorithm is known for its simplicity and efficiency, it has its limitations. One such limitation is the need to manually specify the number of clusters, which can be inaccurate and may lead to suboptimal results due to local optimization. To address this issue, an optimization method is employed. Initially, the number of clusters is determined using the elbow method [21]. The elbow method employs the Sum of Squares due to Error (SSE) to select the optimal number of clusters, as illustrated in Equation (17).

$$SSE = \sum\nolimits_{i=1}^{k}\sum\nolimits_{s\in w_i}|s-m_i|^2 \tag{17}$$

A smaller SSE value indicates a better clustering outcome. In Equation (17), k refers to the number of clusters, w_i refers to the i-th cluster, s rdenotesa sample in w_i, and $\mathrm{m_i}$ correspond to the centroid of the cluster w_i. Furthermore, the K-means++ algorithm is employed to initialize the cluster centers, mitigating the issue of random initialization leading to local optima, as demonstrated in Equation (18).

$$P(c_i) = \frac{Q(c_i)^2}{\sum_{i=1}^{H} Q(c_i)^2} \tag{18}$$

$P(c_i)$ represents the probability of selecting sample i as the initialization centroid, and $Q(c_i)^2$ denotes the distance between sample i and the other samples.

In the Gaussian mixture model (GMM), the statistical distribution of certain data is quantified by assuming that the sample conforms to a linear combination of k Gaussian distributions. The probability of each sample point belonging to each cluster is calculated, and the sample point is assigned to the Gaussian distribution with the highest probability of completing the clustering [3]. Let the distribution of the sample numbers be composed of k Gaussian distributions, and the mixed model of k Gaussian distributions is shown in Equation (19).

$$P(x|\theta) = \sum\nolimits_{i=1}^{k}\lambda_k\phi(x|\theta_k) \tag{19}$$

$$\sum\nolimits_{i=1}^{k}\lambda_k = 1, \quad \lambda_k \in (0,1) \tag{20}$$

In Equation (20), λ_k refers to the probability that the sample belongs to the k-th Gaussian distribution; x refers to sample data; $\phi(x|\theta_k)$ refers to the k-th Gaussian distribution, as shown in Equation (21).

$$\phi(x|\theta_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}e^{(x-\mu_k)^2/2\sigma_k^2} \tag{21}$$

σ_k^2 refers to the variance of the k-th Gaussian distribution, and μ_k refers to the expected value of the k-th Gaussian distribution. Equation (22) is shown as follows.

$$\begin{cases} n_k = \sum\nolimits_{i=1}^{N} s_{ik} \\ N = \sum\nolimits_{i=1}^{K} n_k \end{cases} \tag{22}$$

s_{ik} in Equation (22) is a hidden variable, and its equation is as follows:

$$s_{ik} = \begin{cases} 1, x_i \in \phi_k \\ 0, x_i \notin \phi_k \end{cases} \tag{23}$$

Equation (24) can be derived based on Equation (23).

$$lnP\left(x,s\mid\theta\right)=\sum\nolimits_{k=1}^{K}n_k ln\lambda_k+\sum\nolimits_{i=1}^{N}s_{ik}[-\frac{1}{2}ln2\pi -\frac{1}{2}ln\sigma_k^2-\frac{(x-\mu_k)^2}{2\sigma_k^2}] \tag{24}$$

Let s_{ik} be marked as $T(S)$, γ_{ik} as $E\left(T\right)$, and n_k as $\sum_{i=1}^{N}\gamma_{ik}$. Then, the maximum likelihood estimation is as shown in Equation (25).

$$\begin{cases} L\left(\theta\right)=\sum\nolimits_{k=1}^{K}n_k[ln\lambda_k--\frac{1}{2}ln2\pi-\frac{1}{2}ln\sigma_k^2-\frac{(x-\mu_k)^2}{2\sigma_k^2}] \\ \gamma_{ik}=\frac{\lambda_k\phi(x|\theta_k)}{\sum_{k=1}^{K}\lambda_k\phi(x|\theta_k)} \end{cases} \tag{25}$$

Let $L\left(\theta\right)$ represent the partial derivative with respect to θ_k separately, as presented in Equation (26).

$$\begin{cases} \mu_k=\frac{1}{n_k}\sum\nolimits_{i=1}^{N}\gamma_{ik}x_i \\ \sigma_k^2=\frac{1}{n_k}\sum\nolimits_{i=1}^{N}\gamma_{ik}\left(x-\mu_k\right)^2 \\ \lambda_k=\frac{n_k}{N} \end{cases} \tag{26}$$

Equations (14) and (15) represent the steps of the Expectation Maximization (EM) algorithm, which iteratively maximizes the $L(\theta)$ function to complete the training of the Gaussian mixture model [20]. The parameter updating steps for the Gaussian mixture model using the EM algorithm are illustrated in Figure 6.

Finally, the optimal number of clusters is determined using the Bayesian Information Criterion (BIC). The penalization equation for this criterion is presented in Equation (27).

$$BIC=Kln\left(N\right)-2ln(L) \tag{27}$$

In Equation (27), K represents the number of models fitted in the Gaussian mixture model, N denotes the number of samples, and L signifies the likelihood function. The determination of the number of clusters relies on the BIC value, with the k value corresponding to the minimum BIC selected as the optimal number of clusters in the Gaussian mixture model.

Furthermore, the optimal clustering algorithm is chosen based on internal clustering performance evaluation metrics, including the CH coefficient and contour coefficient.

The CH coefficient evaluates the clustering effectiveness by quantifying the cohesion within clusters and the separation among clusters. A higher value indicates a more favorable clustering outcome. Its formulation is presented in Equation (28).

$$CH(n_c)=\frac{\sum_{i=1_c^n}n_i d^2(X_i,X_M)(n_c-1)}{\sum_{i=1_c^n}\sum_{x\in X_i}d^2(x,X_i)/(n-n_c)} \tag{28}$$

In Equation (28), X_M represents the center of the sample data X. n_c denotes the total number of data points contained in category X_i. D stands for the distance between data point X_i and cluster center X_M.

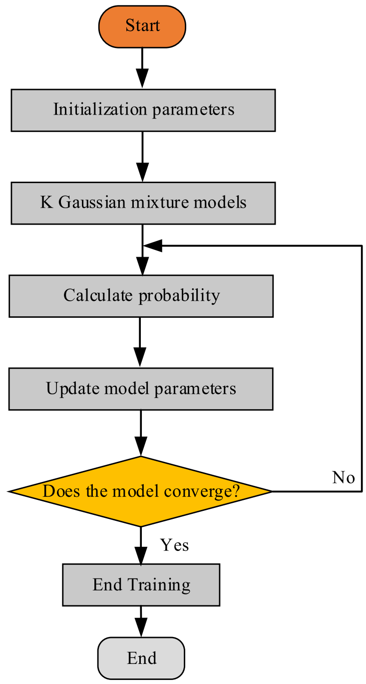

FIGURE 6. Flow chart of updating parameters of Gaussian mixture model by the EM algorithm.

The contour coefficient combines the two elements of intra-cluster aggregation and inter-cluster separation to evaluate the clustering effect, as shown in Equation (29).

$$\bar{S}=\frac{1}{n}\sum\nolimits_{i=1}^{n}(\frac{b\left(i\right)-a(i)}{max[a\left(i\right),b(i)]}) \tag{29}$$

In Equation (29), $\bar{S}$ refers to the contour coefficient, with values ranging from -1 to 1. A larger value indicates a better clustering result. When the value is negative, it suggests that the sample clustering might be incorrect.a(i) represents intra-cluster dissimilarity, while $b\left(i\right)$ signifies inter-cluster dissimilarity.

D. CONSTRUCTION AND ANALYSIS OF SUITABILITY EVALUATION MODEL FOR RURAL LAND INTEGRATION OF TERTIARY INDUSTRY BASED ON DEEP LEARNING FUSION CLUSTER ANALYSIS

This study introduces an advanced method that amalgamates deep learning and clustering analysis techniques. The deep learning model, ResNet-50, is harnessed to extract intricate abstract features. Subsequently, clustering analysis is deployed to categorize analogous land regions into coherent clusters. The selection of the optimal clustering scheme relies on the evaluation of C and silhouette coefficients. An improved land suitability assessment methodology is presented, markedly augmenting the precision and comprehensive appraisal of rural land concerning its viability for the integration of the three industries. The model for evaluating

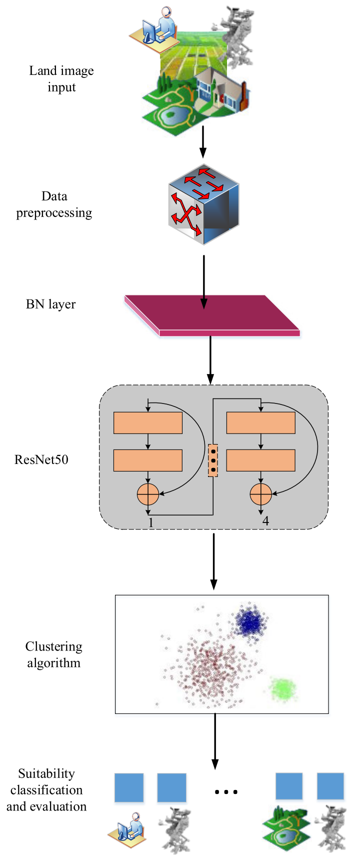

FIGURE 7. **Schematic diagram of the rural land suitability evaluation model based on the fusion of deep learning and cluster analysis.**

the suitability of rural land for tertiary industry integration, which is based on the fusion of deep learning and cluster analysis, is illustrated in Figure 7.

In Figure 7, this model begins with a series of data preprocessing and feature learning steps. Initially, diverse data from rural areas, encompassing land use, agricultural production, and rural tourism, is gathered. These data undergo a series of operations including cleaning, denoising, and normalization, to ensure data quality and consistency. Subsequently, a ResNet-50 model is employed for training and feature learning. In the process of transfer learning, the pre-trained ResNet-50 model is loaded and trained on a substantial image dataset, allowing it to acquire advanced features from images. The model is then tailored to align with specific task requirements. To preserve the learned feature representation capability of the pre-trained model, specific lower-level convolution layers are frozen, preventing their weights from being updated during training and retaining the features they initially acquired from the source data. Ultimately, through training on new layers, the model gradually adapts to the task at hand, acquiring the ability to map the pre-trained feature representation to the specific rural land suitability evaluation task. ResNet-50 excels at learning advanced abstract features within the data, including various land use types, levels of agricultural production, and rural tourism resources. This enables the model to more accurately capture regional

```
Start
Input: Land image data
Output: Comprehensive evaluation results of land
suitability
# Data preparation and preprocessing
# Combine data
# Data cleaning, noise reduction, normalization, etc.
# Split data into training and testing sets
# Build ResNet-50 model
   def build_resnet50_model()
   # Compile and train ResNet-50 model
   # Use ResNet-50 for feature extraction
# Perform clustering analysis
# Conduct suitability assessment
   def suitability_assessment(data, clusters):
   suitability_scores = []
   for cluster_id in range(num_clusters)
End
```

FIGURE 8. **Model algorithm flow.**

characteristics and lays a robust foundation for subsequent evaluation processes.

Secondly, the process involves feature fusion and cluster analysis. Here, the advanced features extracted from the ResNet-50 model are integrated with cluster analysis. This method classifies similar plots or areas into the same cluster through cluster analysis techniques by utilizing the features derived from the deep learning model as input. This integration allows for a comprehensive consideration of various factors and illuminates both the relationships and distinctions among different categories of integrated land. Cluster analysis plays a pivotal role in uncovering potential laws and patterns, thereby enhancing the comprehensiveness and accuracy of land suitability evaluations.

Thirdly, the process involves comprehensive evaluation and result interpretation. The clusters generated through cluster analysis serve as the classification of land types for rural three-industry integration, facilitating a comprehensive assessment of land suitability. By comparing and analyzing the characteristics of different clusters in alignment with existing rural development and land use policies, a deeper understanding of the suitability of each type is achieved, thereby furnishing decision-makers with a scientifically grounded basis for their choices. This capacity for comprehensive analysis enhances the credibility and practicality of the evaluation results.

The core algorithm flow of this model is depicted in Figure 8.

IV. EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

A. DATASETS COLLECTION

The data sources for this study are diverse and encompass various categories, as outlined in Table 1.

After acquiring the data, this study preprocesses it and then partitions it into a training set and a test set based on the data

TABLE 1. Data sources of rural tertiary industry integration.

Data	Belonging industry	Data type	Source
Land use data	First industry, second industry and third industry	Land use type, land use area, land use change, etc.	National land use data network (http://www.dsac.cn/)
Agricultural production data	First industry	Crop planting area, agricultural product output, agricultural input and income, etc.	National Bureau of Statistics (http://www.stats.gov.cn/)
Industrial production data	Second industry	The number, scale, output value, land area and pollution discharge of industrial enterprises, etc.	National Bureau of Statistics (http://www.stats.gov.cn/)
Rural tourism data	Third industry	The number of tourist attractions, tourists, tourism income, etc.	National Tourism Administration (http://www.ctaweb.org.cn/)

type, maintaining a 7:3 ratio between the training and test sets.

B. EXPERIMENTAL ENVIRONMENT

All experiments are conducted on an Ubuntu 18.04 system with an Intel 9900K CPU, featuring 8 cores and 1G threads, with a maximum frequency of SGHz. The server is equipped with 32GB of memory. An NVIDIA RTX 2080TI GPU with 11GB of video memory is employed. The deep learning framework utilized for these experiments is TensorFlow, along with essential Python packages and tool libraries such as NumPy.

C. PARAMETERS SETTING

In the training process, the following hyper-parameters are employed: a weight decay of 0.00001, a momentum of 0.9, and an initial learning rate of 0.15. A Cosine Annealing strategy is adopted to modulate the learning rate as the number

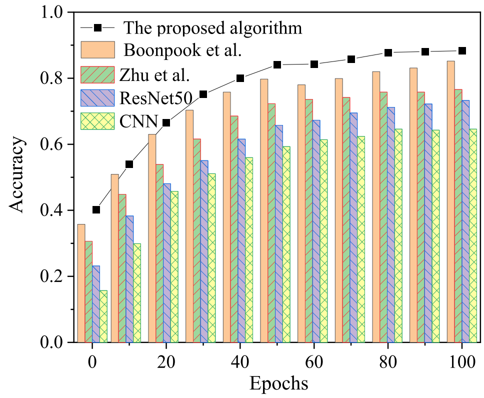

FIGURE 9. Accuracy results of rural land use classification recognition under each algorithm.

of consecutive cycles increased. At the end of each cycle, the learning rate is eventually reduced to 0.00001, and this training process is repeated 100 times. These parameter settings allow the training process to commence with a high initial learning rate, facilitating rapid convergence aided by momentum. Subsequently, the learning rate is gradually reduced using the Cosine Annealing strategy, enhancing the model's stability and generalization capacity in the later stages of training. This adjustment contributes to the model's improved adaptation to the data, resulting in superior performance. The optimization algorithm used in the model is Adam, the activation function is ReLU, and the batch size is 64. The model consists of a total of 49 convolutional layers and 1 fully connected layer.

D. PERFORMANCE EVALUATION

In the initial stage of the analysis, several algorithms are employed, including the model algorithm proposed in this study, ResNet50, CNN, Zhu et al. [32], and Boonpook et al. [2]. Rural land use classification and identification accuracy are evaluated using various metrics, including accuracy, precision, recall, and F1 value. Figures 9 through 12 illustrate the outcomes of these assessments.

Figures 9 to 12 present a detailed analysis of rural land use classification and identification accuracy using the model algorithm developed in this study, ResNet50, CNN, Zhu et al. [32], and Boonpook et al. [2]. These analyses consider metrics such as accuracy, precision, recall, and F1 value score. The results reveal that the rural land use classification and identification accuracy achieved by the model algorithm in this study is notably high, reaching 88.3%. This performance surpasses that of the other model algorithms, with at least a 3.1% advantage. The order of accuracy from highest to lowest is as follows: the model algorithm in this study > Boonpook et al. [2]> Zhu et al. [32] > RESNET 50 > CNN. Furthermore, when examining precision, recall, and F1 values, it becomes evident that the research model algorithm for rural land use classification and identification is superior. This superiority may be attributed to the combination of

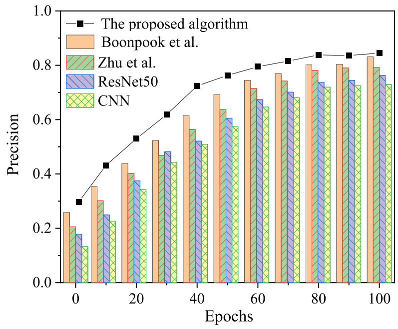

FIGURE 10. Precision results of rural land use classification recognition under each algorithm.

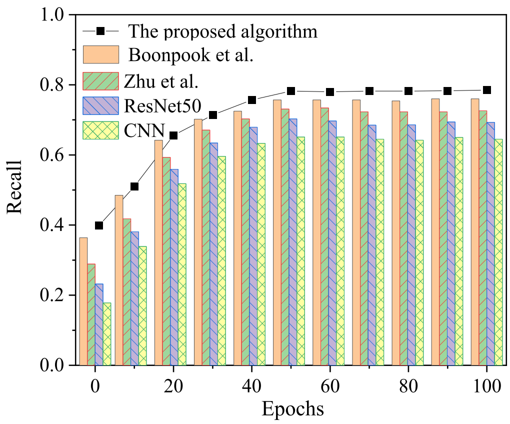

FIGURE 11. Recall results of rural land use classification recognition under each algorithm.

ResNet-50 and the K-means algorithm, which enhances the model's ability to accurately capture land characteristics and classify similar plots or areas, ultimately improving accuracy and the overall analytical capability for land suitability evaluation. Consequently, this study's rural land suitability evaluation model based on deep learning fusion cluster analysis excels in rural land use classification and identification accuracy.

Furthermore, a comparison of the average intersection and union ratio for each algorithm is illustrated in Figure 13.

Figure 13 provides insight into the mean IoU ratio of each algorithm. Analysis reveals a common trend of initial increase followed by stabilization in the average intersection ratio as the training period progresses. Notably, when the training period reaches 100, the model algorithm introduced in this study attains a substantial mean IoU of 67.29%. In contrast, the highest mean IoU achieved by other model algorithms peaks at 64.29%, significantly lower than that of the algorithm reported here. Consequently, a comprehensive analysis of experimental results consistently highlights the superior accuracy of rural land use classification and identification achieved by the model algorithm presented in this study.

Furthermore, to enhance the clustering performance of the model, a comparative analysis of the CH coefficient and contour coefficient between the k-means algorithm and GMM algorithm is presented in Figure 14.

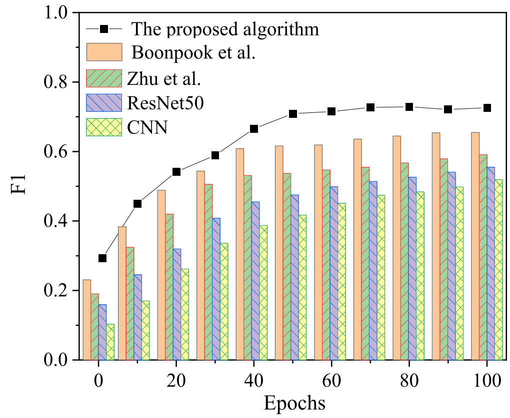

FIGURE 12. F1 value results of rural land use classification and identification under each algorithm.

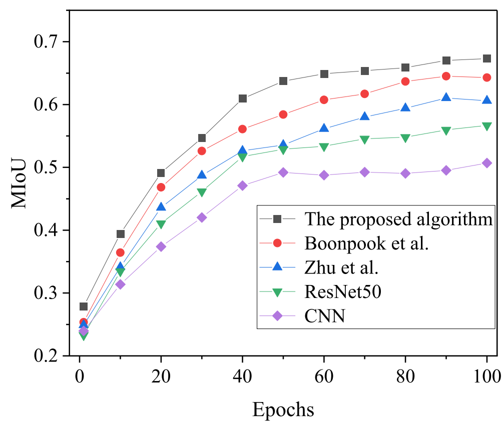

FIGURE 13. Mean IoU results of each algorithm.

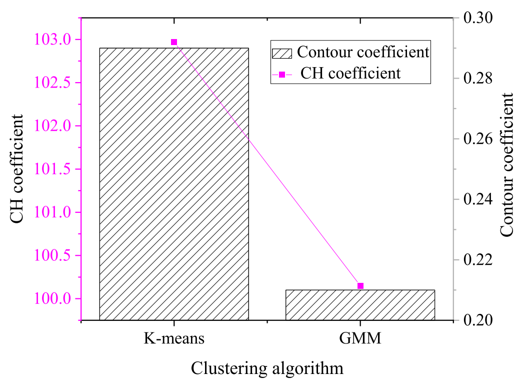

FIGURE 14. Comparison of scores of different clustering algorithms.

Upon comparing the clustering scores of each algorithm, depicted in Figure 14, it is evident that the k-means algorithm attains scores of 102.97 for the CH coefficient and 0.29 for the contour coefficient, while the GMM algorithm achieves scores of 100.15 for the CH coefficient and 0.21 for the contour coefficient. Clearly, the k-means algorithm outperforms in terms of both CH coefficient and contour coefficient, signifying its superior clustering efficacy. Hence, this study selects the clustering results generated by the k-means algorithm as the foundation for classifying rural three-industry integration land.

E. DISCUSSION

This study reveals that the algorithm developed in this study achieves an impressive accuracy of 88.3% in rural land use classification and identification, significantly outperforming other algorithms in this regard. Additionally, the mean IoU reaches 67.29%, indicating the superior feature identification accuracy of this study's model algorithm for integrated rural land for three industries when compared to Zhu et al. [32] and Boonpoke et al. [2]. Furthermore, in the analysis of clustering algorithms, the k-means algorithm demonstrates a strong clustering effect, notably surpassing the clustering score of GMM, consistent with Jiang and Beck [13].

The model reported here offers a versatile array of practical applications. Firstly, it enhances the accuracy of rural land use classification, enabling a more precise understanding of land use conditions. This, in turn, provides essential data support for urban planners and decision-makers, facilitating the formulation of more effective development strategies. Secondly, it allows for more targeted planning of rural industrial land by delineating the types and directions of industrial development in rural areas. This precise planning can identify the specific types of industrial development needed in rural areas, thereby guiding future industrial development in different regions [29]. Such guidance promotes industrial revitalization in various locales, contributing to the realization of sustainable rural development. In conclusion, the practical application of this research model holds significant potential for rural development planning and land use management, offering strategic support and guidance for the sustainable development of rural areas.

This guide offers a systematic approach for novice researchers, particularly graduate students, aiming to apply cutting-edge methods in unfamiliar fields or scenarios. The proposed steps and recommendations encompass the following key aspects:

1) Acquaintance with Current Methods: Beginners should embark on a comprehensive exploration of the prevailing methodologies and techniques. This entails mastering deep learning models (e.g., ResNet-50), familiarizing themselves with clustering analysis algorithms (e.g., K-means and Gaussian Mixture Models), and grasping the foundational principles of land suitability assessment. This proficiency can be cultivated through an extensive review of pertinent literature, engagement in online courses and tutorials, and hands-on programming and experimentation.
2) Data Gathering and Organization: Graduate students should undertake the collection of pertinent data pertinent to their new area of interest or research inquiry. This may encompass geographical information system (GIS) data, remote sensing images, rural economic statistics, and other relevant sources. Vigilance in preserving data quality and consistency is paramount; hence, data preprocessing and cleansing are indispensable stages.
3) Data Preprocessing and Feature Engineering: Novices must become proficient in the art of data preprocessing and feature engineering. These activities encompass tasks such as data cleansing, noise mitigation, normalization, and the extraction of salient features from raw datasets. The caliber of feature engineering directly impacts the model's efficacy.
4) Model Selection and Training: Depending on the research query, neophytes should elect an apt deep learning model, potentially ResNet-50 or an alternative well-suited model. Thereafter, they should engage in model training using their proprietary data, which may necessitate access to a computer equipped with GPU acceleration. The refinement of the model's hyperparameters to attain peak performance is also a pivotal endeavor.
5) Clustering Analysis and Interpretation: Following the acquisition of predictions from the deep learning model, beginners can amalgamate these outcomes with clustering analysis techniques. This synergy aids in the revelation of latent patterns and cohorts within the data, further elucidating the model's prognostications.
6) Results Comprehension and Documentation: Lastly, newcomers must interpret their research findings and compile an exhaustive report or scholarly paper. This entails bridging the model's outputs with real-world contexts, scrutinizing the import of the results, and deliberating on the study's constraints and prospective avenues for further research.

This study outlines key avenues for advancing land suitability assessment methods, focusing on improving model performance, multimodal data fusion, interpretability studies, and cross-domain applications. Future research should prioritize enhancing deep learning models' performance in new scenarios. This involves optimizing model architectures and leveraging data augmentation techniques to increase their effectiveness. Research efforts should explore techniques for seamlessly integrating these multimodal data into unified models in cases involving diverse data types (e.g., images, geographic, economic). This integration can boost prediction accuracy and offer a more holistic view of complex landscapes. Addressing the interpretability of deep learning models is paramount. Research should delve into methods to elucidate model decisions and establish meaningful links between interpretability findings and real-world contexts. This direction aims to make models more transparent and accountable. Extending these methods to different domains (e.g., urban planning, environmental protection, natural resource management) holds significant promise. Adapting these techniques to new scenarios across diverse fields can lead to innovative solutions for multifaceted challenges.

By following these methods and recommendations, beginners can acquire the skills to adapt existing techniques to novel fields or scenarios and explore promising avenues for future research. Furthermore, this practice encourages fellow scholars to engage actively and contribute to the advancement of pertinent domains.

V. CONCLUSION

A. RESEARCH CONTRIBUTION

This study introduces deep learning and clustering algorithms and presents an innovative rural land suitability evaluation model based on deep learning fusion clustering analysis. The novelty of this study lies in the synergistic integration of deep learning and cluster analysis, which yields more precise and comprehensive support for rural sustainable development. Experimental analysis reveals that the accuracy of rural land classification achieved by the model algorithm presented in this study reaches 88.3%. Additionally, the mean IoU attains 67.29%, demonstrating clear superiority over other algorithms. The utilization of the k-means algorithm for clustering land suitability enhances its utility, providing robust support and guidance for the sustainable development of rural areas.

The ResNet-50 model employed in this analysis exhibits several strengths and limitations. ResNet-50 is a deep CNN model characterized by its substantial depth, which facilitates the extraction of high-level abstract features from data. It excels at recognizing diverse land-use patterns, agricultural production levels, and rural tourism resources, thereby enhancing its feature extraction capabilities for land characteristics. ResNet-50 adopts a transfer learning approach, initially pre-trained on extensive image datasets and subsequently fine-tuned for specific tasks. This strategy expedites model convergence, diminishes the demand for copious labeled data, and enhances the model's generalization prowess. Parameter freezing is applied to the lower convolutional layers of ResNet-50, ensuring that these layers retain their original feature representations and thereby aiding the model's adaptation to the specific task. In scenarios involving data with spatial structures, such as image-based land suitability assessment and land-use classification, ResNet-50 is renowned for its high performance. However, it also presents several limitations. Notably, deep learning models, including ResNet-50, necessitate substantial volumes of labeled data, which can be a limiting factor when such data is scarce. Furthermore, the computational resources required for training and inference are considerable, often mandating high-performance computing equipment and hardware acceleration. The numerous hyperparameters inherent to deep learning models, including learning rates, weight decay, and iteration counts, demand meticulous tuning and optimization to achieve peak performance. Inappropriate parameter configurations can lead to reduced model efficacy or training instability. Lastly, deep learning models tend to be regarded as black-box models, hindering their interpretability. In fields like land suitability assessment, where interpretability is critical, additional efforts may be required to elucidate the model's decision-making processes.

B. FUTURE WORKS AND RESEARCH LIMITATIONS

Nonetheless, this study has certain limitations. For instance, while it utilizes multi-source data as the foundation, it lacks detailed explanations regarding the specific processing methods for distinct types of data, including remote sensing images, land use data, and tourism data. To address this, future research could focus on expanding the scale and diversity of datasets and integrating additional data sources, such as geographic information system data and socio-economic data. This expansion and integration would serve to enhance the accuracy and comprehensiveness of the suitability evaluation for integrated rural land for three industries.

REFERENCES

[1] W. Abera, L. Tamene, T. Kassawmar, K. Mulatu, H. Kassa, L. Verchot, and M. Quintero, "Impacts of land use and land cover dynamics on ecosystem services in the Yayo coffee forest biosphere reserve, southwestern Ethiopia," *Ecosystem Services*, vol. 50, Aug. 2021, Art. no. 101338.

[2] W. Boonpook, Y. Tan, A. Nardkulpat, K. Torsri, P. Torteeka, P. Kamsing, U. Sawangwit, J. Pena, and M. Jainaen, "Deep learning semantic segmentation for land use and land cover types using Landsat 8 imagery," *ISPRS Int. J. Geo-Information*, vol. 12, no. 1, p. 14, Jan. 2023.

[3] J. Cao, L. Chen, D. Hu, F. Dong, T. Jiang, W. Gao, and F. Gao, "Unsupervised eye blink artifact detection from EEG with Gaussian mixture model," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 2895–2905, Aug. 2021.

[4] M. Castelo-Cabay, J. A. Piedra-Fernandez, and R. Ayala, "Deep learning for land use and land cover classification from the Ecuadorian Paramo," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 1001–1017, Dec. 2022.

[5] G. Cecili, P. De Fioravante, P. Dichicco, L. Congedo, M. Marchetti, and M. Munafò, "Land cover mapping with convolutional neural networks using Sentinel-2 images: Case study of Rome," *Land*, vol. 12, no. 4, p. 879, Apr. 2023.

[6] M. Debella-Gilo and A. K. Gjertsen, "Mapping seasonal agricultural land use types using deep learning on Sentinel-2 image time series," *Remote Sens.*, vol. 13, no. 2, p. 289, Jan. 2021.

[7] R. Garg, A. Kumar, N. Bansal, M. Prateek, and S. Kumar, "Semantic segmentation of PolSAR image data using advanced deep learning model," *Sci. Rep.*, vol. 11, no. 1, pp. 1–18, Jul. 2021.

[8] H. Ge, B. Li, D. Tang, H. Xu, and V. Boamah, "Research on digital inclusive finance promoting the integration of rural three-industry," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, p. 3363, Mar. 2022.

[9] L. Ghayour, A. Neshat, S. Paryani, H. Shahabi, A. Shirzadi, W. Chen, N. Al-Ansari, M. Geertsema, M. P. Amiri, M. Gholamnia, J. Dou, and A. Ahmad, "Performance evaluation of Sentinel-2 and Landsat 8 OLI data for land cover/use classification using a comparison between machine learning algorithms," *Remote Sens.*, vol. 13, no. 7, p. 1349, Apr. 2021.

[10] N. T. Giao, N. V. Cong, and H. T. H. Nhien, "Using remote sensing and multivariate statistics in analyzing the relationship between land use pattern and water quality in Tien Giang Province, Vietnam," *Water*, vol. 13, no. 8, p. 1093, Apr. 2021.

[11] J. Groth, K. Hermans, C. Wiederkehr, E. Kassa, and J. Thober, "Investigating environment-related migration processes in Ethiopia—A participatory Bayesian network," *Ecosystems People*, vol. 17, no. 1, pp. 128–147, Jan. 2021.

[12] Z. Jia, Y. Jiao, W. Zhang, and Z. Chen, "Rural tourism competitiveness and development mode, a case study from Chinese township scale using integrated multi-source data," *Sustainability*, vol. 14, no. 7, p. 4147, Mar. 2022.

[13] X. Jiang and M. Beck, "Visual attention during scene viewing–eye tracking discovery with K-means and Gaussian mixture model," *J. Vis.*, vol. 22, no. 14, p. 3631, Dec. 2022.

[14] J. Jing, S. Ke, T. Li, and T. Wang, "Energy method of geophysical logging lithology based on K-means dynamic clustering analysis," *Environ. Technol. Innov.*, vol. 23, Aug. 2021, Art. no. 101534.

[15] B. Kalisz, K. Żuk-Gołaszewska, W. Radawiec, and J. Gołaszewski, "Land use indicators in the context of land use efficiency," *Sustainability*, vol. 15, no. 2, p. 1106, Jan. 2023.

[16] X. Leng and G. Tong, "The digital economy empowers the sustainable development of China's agriculture-related industries," *Sustainability*, vol. 14, no. 17, p. 10967, Sep. 2022.

[17] X. Li, H. Yang, J. Jia, Y. Shen, and J. Liu, "Index system of sustainable rural development based on the concept of ecological livability," *Environ. Impact Assessment Rev.*, vol. 86, Jan. 2021, Art. no. 106478.

[18] M. Liu, B. Fu, S. Xie, H. He, F. Lan, Y. Li, P. Lou, and D. Fan, "Comparison of multi-source satellite images for classifying Marsh vegetation using DeepLabV3 plus deep learning algorithm," *Ecological Indicators*, vol. 125, Jun. 2021, Art. no. 107562.

[19] R. N. Masolele, V. De Sy, M. Herold, D. Marcos, J. Verbesselt, F. Gieseke, A. G. Mullissa, and C. Martius, "Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series," *Remote Sens. Environ.*, vol. 264, Oct. 2021, Art. no. 112600.

[20] R. Orellana, R. Carvajal, P. Escárate, and J. C. Agüero, "On the uncertainty identification for linear dynamic systems using stochastic embedding approach with Gaussian mixture models," *Sensors*, vol. 21, no. 11, p. 3837, Jun. 2021.

[21] G. Racolte, A. Marques, L. Scalco, L. Tonietto, D. C. Zanotta, C. L. Cazarin, L. Gonzaga, and M. R. Veronez, "Spherical K-means and elbow method optimizations with Fisher statistics for 3D stochastic DFN from virtual outcrop models," *IEEE Access*, vol. 10, pp. 63723–63735, 2022.

[22] F. M. Seaton, G. Barrett, A. Burden, S. Creer, E. Fitos, A. Garbutt, R. I. Griffiths, P. Henrys, D. L. Jones, P. Keenan, A. Keith, I. Lebron, L. Maskell, M. G. Pereira, S. Reinsch, S. M. Smart, B. Williams, B. A. Emmett, and D. A. Robinson, "Soil health cluster analysis based on national monitoring of soil indicators," *Eur. J. Soil Sci.*, vol. 72, no. 6, pp. 2414–2429, Nov. 2021.

[23] A. Shirazy, A. Hezarkhani, A. Shirazi, S. Khakmardan, and R. Rooki, "K-means clustering and general regression neural network methods for copper mineralization probability in Chahar-Farsakh, Iran," *Türkiye Jeoloji Bülteni/Geological Bull. Turkey*, vol. 65, no. 1, pp. 79–92, 2022.

[24] H. Soleimani, O. Nasri, M. Ghoochani, A. Azhdarpoor, M. Dehghani, M. Radfard, M. Darvishmotevalli, V. Oskoei, and M. Heydari, "Groundwater quality evaluation and risk assessment of nitrate using Monte Carlo simulation and sensitivity analysis in rural areas of Divandarreh county, kurdistan Province, Iran," *Int. J. Environ. Anal. Chem.*, vol. 102, no. 10, pp. 2213–2231, Aug. 2022.

[25] D. Wang, W. Cao, F. Zhang, Z. Li, S. Xu, and X. Wu, "A review of deep learning in multiscale agricultural sensing," *Remote Sens.*, vol. 14, no. 3, p. 559, Jan. 2022.

[26] J. Wang, L. Qu, Y. Li, and W. Feng, "Identifying the structure of rural regional system and implications for rural revitalization: A case study of Yanchi county in northern China," *Land Use Policy*, vol. 124, Jan. 2023, Art. no. 106436.

[27] X. Wei, N. Wang, P. Luo, J. Yang, J. Zhang, and K. Lin, "Spatiotemporal assessment of land marketization and its driving forces for sustainable urban–rural development in Shaanxi province in China," *Sustainability*, vol. 13, no. 14, p. 7755, Jul. 2021.

[28] Q. Wen, J. Li, J. Ding, and J. Wang, "Evolutionary process and mechanism of population hollowing out in rural villages in the farming-pastoral ecotone of northern China: A case study of Yanchi county, Ningxia," *Land Use Policy*, vol. 125, Feb. 2023, Art. no. 106506.

[29] Q. Yanbo, J. Guanghui, M. Wenqiu, and L. Zitong, "How does the rural settlement transition contribute to shaping sustainable rural development? Evidence from Shandong," *J. Rural Stud.*, vol. 82, pp. 279–293, Feb. 2021.

[30] J. Yang, R. Yang, M.-H. Chen, C.-H. Su, Y. Zhi, and J. Xi, "Effects of rural revitalization on rural tourism," *J. Hospitality Tourism Manage.*, vol. 47, pp. 35–45, Jun. 2021.

[31] Y. Zheng, J. Tan, Y. Huang, and Z. Wang, "The governance path of urban–rural integration in changing urban–rural relationships in the metropolitan area: A case study of Wuhan, China," *Land*, vol. 11, no. 8, p. 1334, Aug. 2022.

[32] Q. Zhu, X. Guo, W. Deng, S. Shi, Q. Guan, Y. Zhong, L. Zhang, and D. Li, "Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 184, pp. 63–78, Feb. 2022.