

Research Article

Multispectral Remote Sensing Data Analysis Based on KNNLC Algorithm and Multimedia Image

Yingxin Sun 

Department of Information Engineering, Changchun College of Electronic Technology, 130000, China

Correspondence should be addressed to Yingxin Sun; 201615991118@stu.yznu.edu.cn

Received 18 February 2022; Revised 31 March 2022; Accepted 20 April 2022; Published 11 May 2022

Academic Editor: Wen Zeng

Copyright © 2022 Yingxin Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to combine multimedia imagery and multispectral remote sensing data to analyze information, preprocessing becomes a necessary part of it. It is found that the KNN algorithm is one of the classic algorithms of data mining. As one of the most important branches in the field of data analysis, it is widely used in many fields such as classification, regression, missing value filling, and machine learning. As a lazy algorithm, this method requires no prior statistical knowledge and no additional data to train description rules and is easy to implement. However, the algorithm inevitably has many problems, such as how to determine the appropriate K value, the unsatisfactory effect of data processing for some special distributions, and the unacceptable computational complexity of high-dimensional data. In order to solve these shortcomings, the researchers proposed the KNNLC algorithm. Then, taking the classification experiment as an example, through the comparison of the experimental results on different data sets, it is proved that the average level of the classification performance of the KNNLC algorithm is better than the classic KNN classification algorithm. The KNNLC algorithm shows better performance in most cases, with an accuracy rate of 2 to 5 percentage points higher. An improved algorithm is proposed for the nearest neighbor selection strategy of the traditional KNN algorithm. First, in theory, combined with the theory of sparse coding and locally constrained linear coding, the classical KNN algorithm is improved, and the KNNLC algorithm is proposed. The comparison of the experimental results on the data set proves that the average level of the KNNLC algorithm is better than the classical KNN classification algorithm in terms of classification performance.

1. Introduction

Remote sensing image fusion is a technology that combines multisource remote sensing images through advanced image processing. It makes full use of the different characteristics of a variety of data, so that the image has a higher spectral and spatial resolution at the same time, and improves the vision of the image. The effect and accuracy of image feature recognition and classification accuracy are shown in Figure 1 [1]. Remote sensing image fusion is a hot research topic in the international remote sensing community in recent years. In the method of image fusion, there are some classic algorithms, such as HIS transformation method, COS transformation method, HIS transformation method, and HSV transformation method. In recent years, with the introduction of wavelet transform into the field of image processing, image fusion methods based on wavelet transform have attracted people's

attention. The fusion of SPOT panchromatic image and multispectral image based on 2-ary and 3-ary wavelet is studied, respectively. However, these two algorithms simply replace low-resolution images with high-resolution remote sensing images for low-frequency components after wavelet decomposition, without considering the loss of image features; although the feature-based binary wavelet image is studied fusion, but without considering the resolution of the image to be fused, the fusion effect is not very good [2]. Based on the in-depth study of wavelet transform fusion method, a new fusion method is proposed, a feature-based multiband wavelet fusion. The fusion results of SPOT image and TM5, 4, 3 image, SPOT panchromatic image, and SPOT multispectral band image are given and compared with other fusion methods [3]. The experimental results show that the method in this paper has obvious advantages compared with other fusion methods. Although the KNN algorithm has a good effect on applications

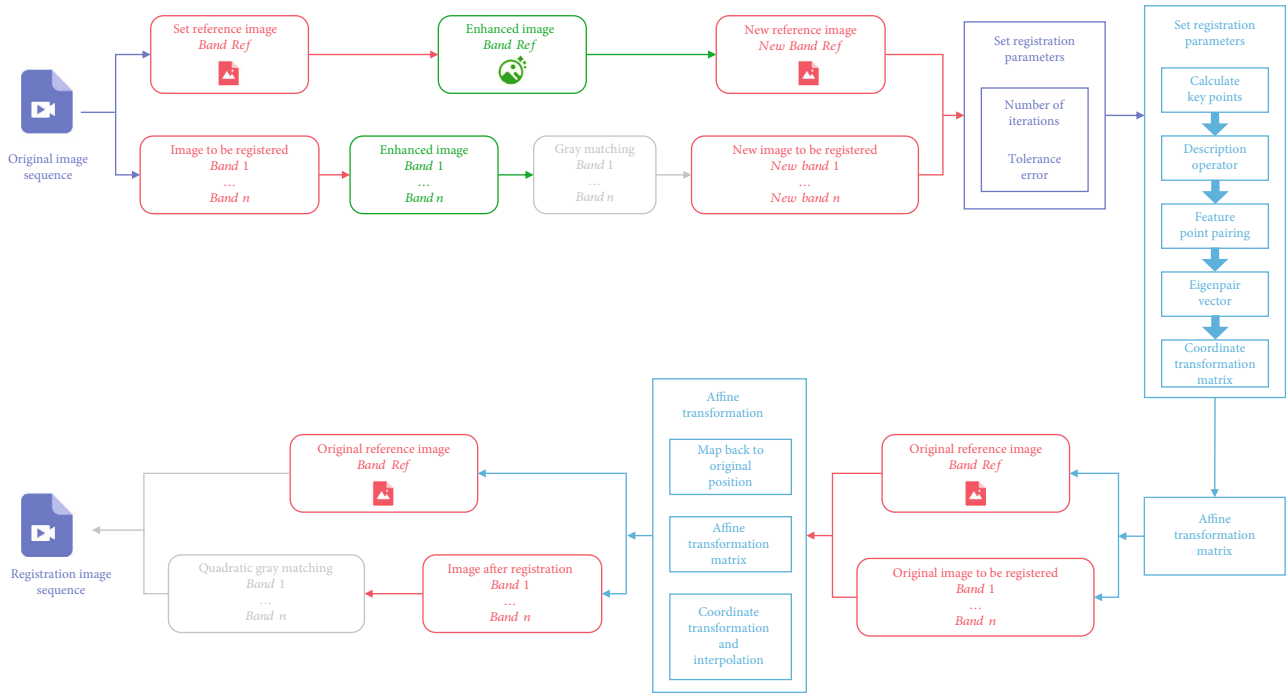


FIGURE 1: Data analysis flow chart of multispectral remote sensing image.

such as classification and prediction in many data sets, it inevitably produces many problems that need to be solved, such as the high time complexity and space complexity of the algorithm, and the K value. The processing effect is not ideal for some special distributed data and the computational complexity of high dimensional data is unacceptable. These shortcomings must be solved for a mature algorithm. Therefore, experts and scholars who are interested in this direction have done a lot of research and obtained many optimization algorithms.

2. Literature Review

Jin, R. et al. systematically analyzed the effects of traditional spectral parameters and two-band normalization and ratio vegetation index under different observation angles in estimating wheat leaf nitrogen content (LNC), thereby establishing a multiangle quantitative monitoring of wheat canopy leaf nitrogen content model [4]. Du, JH et al. found that the canopy reflectance and the coefficient of determination of 40 conventional spectral vegetation indices and LNC decreased with the increase of the observation angle, regardless of the forward or backward observation direction, and reached the maximum at -20° in the backward direction value [5]. RI-1dB and EVI-1 have the closest relationship with LNC at -20° backward and vertical angle, respectively. The areas with good correlation between the ND and SR parameters of the original spectral reflectance combination of the two bands and LNC are mainly concentrated in the blue-red light band, the green-red band, and the red-side red band combination range. This sensitive area varies with the spectrum observation. The angle is different. The new Multi-Angle Vegetation Index (MAVI), which uses the combination of sensitive spectral parameters and observation angles, can better estimate LNC. After indepen-

dent data testing at different years, the MAViR model is most sensitive to leaf nitrogen content. By systematically analyzing the angular sensitivity characteristics of different wavebands and spectral parameters, Fang, X. et al. studied the quantitative relationship between suitable characteristic parameters extracted by different spectral analysis techniques and the nitrogen content of leaves. The results show that the correlation between spectral vegetation index and leaf nitrogen content is better than that of vertical and forward observation angles in the backward observation angle. The red edge parameters mND705, GND (750, 550), NDRE, and RI-1dB are compared with LNC. The relationship is the closest, but the difference is large under different experimental factors, especially when the leaf nitrogen content is high ($>4.5\%$), the spectral parameters tend to be saturated [6]. Yang, FC and others found that the newly constructed angle insensitive parameter (AIVI) reduces the influence of different test factors. In the range of $-10^\circ \sim 40^\circ$ observation angle, AIVI can establish a unified and stable monitoring model, and it has been independently tested. According to the data test, it is the best to construct a monitoring model of wheat canopy leaf nitrogen content based on AIVI, which has strong angle adaptability [7]. Ren, J. et al. found that the inversion accuracy of wheat leaf nitrogen content based on FA-BPNN analysis was significantly higher than that of conventional spectral parameters under different observation angles [8]. Therefore, both the new vegetation index AIVI and FA-BPNN can reliably monitor the nitrogen content of wheat leaves under different experimental conditions. By comparing the relationship between various spectral analysis methods and LAI under different observation angles, the appropriate band sensitive to changes in LAI can be extracted, and the observation angle; thus, a quantitative monitoring model for wheat LAI was established. The results show that different

spectral analysis methods are more suitable for monitoring LAI (leaf area index) near the vertical angle. The spectral reflectance and the correlation between spectral parameters and LAI (leaf area index) in the backward observation direction are higher than those in the forward observation direction. Li, J. et al. found that the two-band ratio (SR) and normalized index (ND) under different observation angles did not show outstanding monitoring advantages, but the SR effect was better than the ND method [9]. Using factor analysis technology, it is found that the load of the green light band decreases with the increase of the observation angle in the first factor and increases with the increase of the observation angle in the second factor. After independent test data in different years, the wheat LAI (leaf area index) monitoring model established with the spectral parameter VI_{opt} as a variable has good test results and can be used for accurate estimation of wheat LAI (leaf area index). Torra, V. et al. analyzed and compared the saturation, angle sensitivity, and variety sensitivity of commonly used vegetation indices for estimating LAI (leaf area index). The results of wide-angle adaptability show that the accuracy of LAI (leaf area index) estimation of spectral parameters is better for erect varieties than for discrete varieties. Nonperpendicular observation angles did not significantly improve the ability of spectral parameters to estimate LAI (leaf area index) [10]. Except for EVI and TVI, the spectral parameters NDVI, SAVI, OSAVI, MSAVI, WDRVI, MTVI, and mND705 all tend to be saturated when LAI is greater than 4. KA Zweig et al. found that the angle reduction coefficient K_f constructed based on green light and near-infrared bands is closely related to LAI (leaf area index). The product of VIs and K_f effectively alleviates the saturation and variety sensitivity of LAI (leaf area index) estimation at different observation angles, and significantly improves the monitoring accuracy and adaptability of LAI (leaf area index) [11]. V., Subramaniaswam et al. found that in the above-mentioned spectral parameters, except for WDRVI, EVI, and TVI, the other spectral parameters and the K_f product established a unified monitoring model at all observation angles. The prediction model based on mND705 and OSAVI spectral parameters is more accurate and reliable [12]. The effects of various spectral processing methods to estimate chlorophyll density were analyzed by integrating spectral data from different observation angles, and a multi-angle remote sensing monitoring model for wheat leaf pigment density was established. The results show that the spectral reflectance that has a good correlation with chlorophyll density is mainly concentrated in the red edge and the near-infrared region (720-900 nm). The spectral parameters VOG1, RI-1dB, NDRE, SDR/SDB, and DD are closely related to the chlorophyll density. The sensitive bands of the normalized and ratio vegetation index of the two bands in the backward observation direction are mainly concentrated in the red area, and in the forward observation direction, they are mainly concentrated in the blue and red light areas. Research by Hu, J. et al. found that the first factor of the FA-BPNN model is mainly concentrated in the blue and red bands under different observation angles, and the second factor is mainly concentrated in the near-infrared region. Backward observation close to the vertical observation angle is beneficial to improve the prediction accuracy of chlorophyll density. The spectral parameters SDR/SDB, DD, ND (720, 760), and

ND (732, 738) are the most effective for monitoring wheat chlorophyll density [13]. K nearest neighbors (KNN for short) is an extension of the nearest neighbor method and is a lazy learning method based on instance statistical classification. As one of the classic algorithms for classifiers and machine learning, KNN's earliest related papers are nearest neighbor pattern classification published by CoverTM and HartPE of Stanford University in 1967. In order to improve the KNN algorithm, the number of nearest neighbors is set to K . The mathematical model of the theory has been developed, opening the door to various improved studies based on nearest neighbor theory.

3. Methods

3.1. KNN Algorithm. The basic idea of KNN: For an input test sample with no assigned label, first compare the features or attributes of the test sample with the corresponding features or attributes of all training samples with existing labels; then find the K nearest samples from the training samples, and then sort the labels of these training samples in descending order, and the label corresponding to the first position in the sequence is the label of the test sample. First, for a given data set, if any data in the set has a class label, then this set is called a training sample set, and the data in it is called a training sample; conversely, if the class label of the data is unknown, it is called the test sample, and the collection is called the test sample set.

The working principle of the KNN classification algorithm is to use a similarity measure to compare each attribute or feature of the test sample with the attributes or features corresponding to all training samples in the training sample set, and arrange the test samples corresponding to the similarity in descending order. According to this, the first K most similar (measured nearest) training samples (K nearest neighbors) can be found in the training set. Generally, K is selected as an integer not greater than 20. Finally, sort the number of occurrences of the class labels of the K training samples in descending order, and the label corresponding to the first place in the sequence is the class label of the test sample [14, 15]. First, for a given data set, if any data in the set has a class label, Figure 2 is a classic example of the KNN classification process. In Figure 2, the training sample includes two types of triangles (Angle) and squares (Square). For the sake of simplicity and clarity of description, we use T and S to represent their numbers, respectively. The dots in the figure are test samples, and K is the number of training samples closest to the test sample, that is, the number of nearest neighbors. When $K=3$, $T=2$, $S=1$, and $T > S$ in the small circle in the dotted line as shown in the figure. According to the principle described above, the test sample is assigned to the triangle type at this time. When $K=5$ is adjusted, $T=2$, $S=3$, and $T < S$ within the large dashed circle in the figure; at this time, the test sample label is judged to be a square. Figure 3 is a flowchart of KNN classification. The specific steps involved in calculating the similarity measurement, selecting the nearest neighbor, and classifying when executing the algorithm will be described in detail later. The flow of the algorithm is described as follows: Supposing that there are m samples in the training set, and the

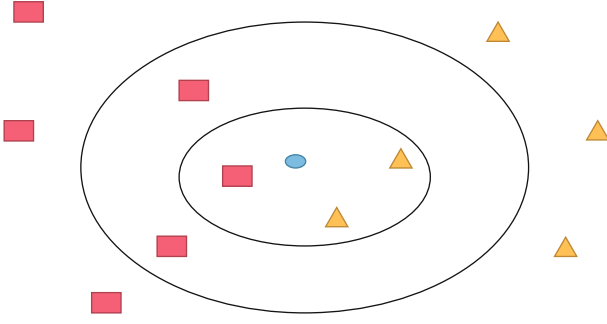


FIGURE 2: An example of KNN algorithm.

number of attributes of each sample is n , then the training set can be recorded as

$T = \{X_i = (x_{i1}, x_{i2}, \dots, x_{in}) | i = 1, 2, \dots, m\}$. The set consisting of the class label of each sample in the set can be denoted as $L = \{C_i | i = 1, 2, \dots, m\}$. The test sample set is denoted as $S = \{S_i = (s_{i1}, s_{i2}, \dots, s_{in}) | i = 1, 2, \dots, a\}$, where a is the number of test samples [16, 17]. Then, KNN classification calculation, where a is the number of test samples. The KNN classification algorithm can be described in Figure 2.

Simply put, the similarity between the test sample and each training sample in the KNN algorithm is measured by calculating the distance. For different data, using an appropriate distance metric is the premise to obtain a good data processing effect. The distance metrics commonly used in the KNN algorithm are Euclidean Distance, Manhattan Distance, MinKowsKi Distance, and Hamming Distance. Given training samples $X = (x_1, x_2, \dots, x_n)$ and test samples $S = (s_1, s_2, \dots, s_n)$, then the distance $\text{dist}(X, S)$ between them is calculated with the following formula.

See formula (1) for Euclidean distance:

$$\text{dist}(X, S) = \sqrt{\sum_{i=1}^n (x_i - s_i)^2}. \quad (1)$$

See formula (2) for Mahalanobis distance:

$$\text{dist}(X, S) = \sqrt{\sum_{i=1}^n |x_i - s_i|}. \quad (2)$$

See formula (3) for Ming's distance:

$$\text{dist}(X, S) = \left(\sum_{i=1}^n (|x_i - s_i|^q) \right)^{1/q}. \quad (3)$$

The Hamming distance is shown in formula (4):

$$\text{dist}(X, S) = \sum_{i=1}^n |x_i - s_i|. \quad (4)$$

Minmax normalization is the most common data normalization processing method. The principle of this method is to use a mapping function to project attributes or eigenvalues into the

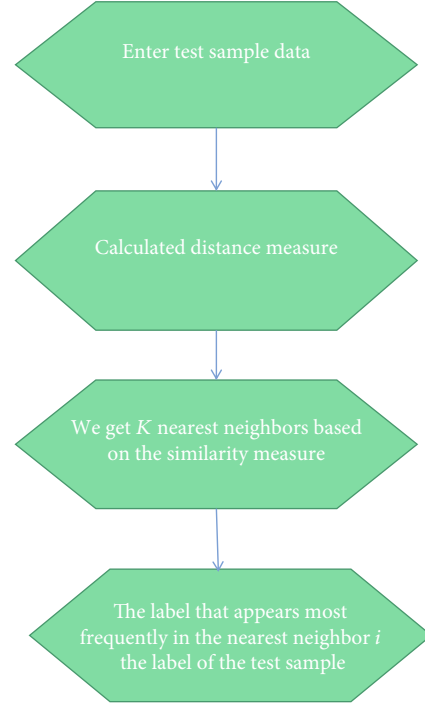


FIGURE 3: KNN algorithm flow chart.

[0,1] interval. Minmax normalization can be expressed by the following formula (5):

$$a' = \frac{a - \min_F}{\max_F - \min_F}. \quad (5)$$

Among them, a is the original value, a' is the value mapped in the [0,1] interval, and \min_F and \max_F are the lower and upper bounds of the values belonging to the same attribute or feature, respectively. There are many commonly used normalization methods, but the basic principles are similar. For example, the form introduced below is shown in formula (6):

$$a'_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{m+1} \sum_{j=1}^n a_{ij}^2}}. \quad (6)$$

The rules of this voting mechanism will also be used in this study. The mathematical representation of the most frequent rule is shown in Equation (7):

$$C_S = \arg \max_{x \in N} \sum I(v = C_x). \quad (7)$$

Since attributes naturally have such distribution characteristics, in the algorithm, in order to make the processing results better reflect the objective facts, weights are generally assigned to each attribute. In the KNN algorithm, the weighting formula of the label can be used as formula (8) form representation:

$$C_s = \arg \max_{x \in N} \sum w_i \times I(v = C_x). \quad (8)$$

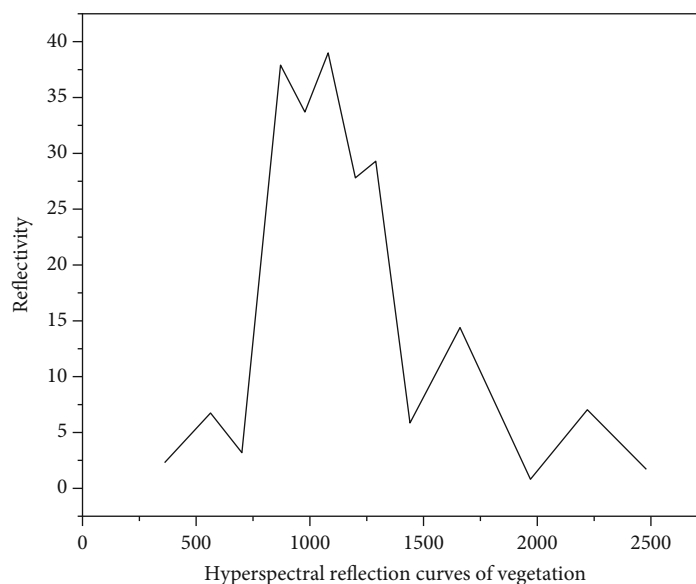


FIGURE 4: The main peak and valley characteristics of the vegetation hyperspectral reflectance curve.

Among the many weighting methods, the weighting method based on the reciprocal similarity is the most classic, and it is also used consistently by the KNN algorithm. Its mathematical expression is shown in Equation (9):

$$w_i = \frac{1}{\text{dist}(X, S)^P}. \quad (9)$$

Among them, P is the weighted power exponent, usually taken as $P=2$, see formula (10):

$$w_i = \frac{1}{\text{dist}(X, S)^2}. \quad (10)$$

3.2. Data Analysis and Utilization. A Locally Constrained Linear Coding Based on Related Research (Locality-constrained Linear Coding, LLC) to improve the classic KNN algorithm. Multiangle remote sensing data has the characteristics of rich information and large amount of data. The selection of spectral absorption characteristic parameters, sensitive angles, and optimal calculation methods are important issues in the study of hyperspectral remote sensing. In addition to the data processing methods commonly used in conventional vertical remote sensing, this study also adopted normalization (ND), ratio (SR), and neural network (BP) analysis methods. The specific methods are as follows: Figure 4 shows typical reflections of vegetation spectral curves, which form absorption valleys at 455 (blue light), 680 (red light), 980, 1200, and 1468 nm, and reflection peaks at 550 (green light), 1090, 1285, 1685, and 2200 nm [18, 19].

Factor analysis is a statistical method for extracting common factors from multiple variables for the purpose of dimensionality reduction. Factor analysis can make factor variables more interpretable through rotation, as shown in Figure 5. Use SPSS software to perform factor analysis on the standardized spectrum data, select the critical factor

numbers whose cumulative contribution rate exceeds 99%, and output the factor data. The BPNN model is provided by Matlab's Neural Network Toolbox. The network is divided into an input layer, a hidden layer, and an output layer. In this study, the input vector is I and the learning goal is T . The input layer is a comprehensive factor with a large contribution rate obtained after factor analysis: the number of neurons in the middle layer is the number of comprehensive factors, and the activation function of the middle hidden layer is "TANSIG"; the neurons in the output layer are 1, and the activation function is "PURELN"; the training function uses the TRAINLM function [20, 21].

4. Results and Analysis

In the classification experiment, we select 4 typical sample sets from the UCI data set as experimental materials. The basic situation of the sample data is shown in Table 1. Among them, Australian and Magic are data sets with only two labels, respectively, and the experiments on them belong to the binary classification experiment, while the experiments on the two data sets of Mpgdata and Ins belong to the case of multiple classification [21–23]. In the four data sets, Iris is a common data set prone to overfitting. The sample is quoted only to show the feasibility of the algorithm and the degree of improvement compared with the more classic algorithms. Its classification effect is obviously difficult to reach in practice. However, the amount of data in the Magic data set is relatively large. Although it is not as good as the big data standard, it is still not general. From this, the potential of the improved algorithm in processing high-capacity data sets compared to traditional algorithms can be seen.

For algorithm classification performance, regardless of the sample size and the number of label classes, the establishment of the corresponding confusion matrix (confusion matrix) is a relatively common and objective evaluation method [24, 25]. The following briefly introduces the confusion matrix. In the

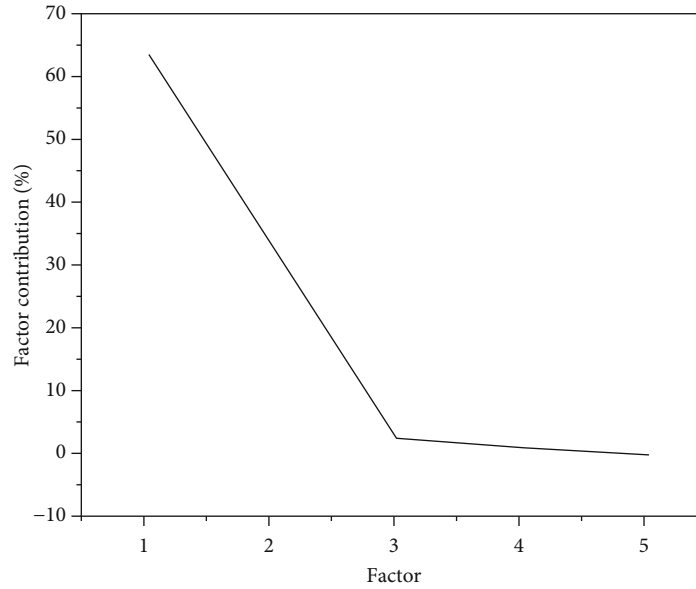


FIGURE 5: Factor contribution degree after t400-900nm band rotation.

TABLE 1: Basic situation of experimental materials.

Data set	Number of instances	Number of attributes
Inis	150	4
Mpgdata	390	7
Australian	690	14
Magic	19020	10

TABLE 2: Confusion matrix.

Predicted class Actual class	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TABLE 3: Accuracy comparison of two classification algorithms (mean \pm standard deviation).

Data set	KNNLC	KNN
Iris	0.9667 \pm 0.0470	0.9400 \pm 0.0857
Mpg data	0.8000 \pm 0.0551	0.7769 \pm 0.0745
Australian	0.8464 \pm 0.0445	0.7928 \pm 0.0305
Magic	0.8220 \pm 0.0055	0.8002 \pm 0.0088

classification problem, due to the difference between the predicted result and the objective facts, the final result has four possibilities, namely, TP (true positive), FP (false positive), FN (false negative), and TN (true negative). They, respectively, represent the four possibilities of pairing between the sample and the label, and the corresponding confusion matrix is shown in Table 2.

Table 2 shows a quantitative comparison of the two algorithms on the four data sets. The use of the mean \pm standard

deviation makes the results more accurate. For classification algorithms, accuracy is often one of the most concerned evaluation criteria. Compared with the classic KNN classification algorithm, the KNNLC algorithm shows better performance in most cases, with an accuracy rate of 2 to 5 percentage points higher [26]. Absolutely, the classic algorithm is classic because of its extensive effectiveness. For data with different distributions, the adaptability of classic algorithms may be more common. For the classification of Mpgdata data sets, sometimes the effect of improved algorithms is not satisfying. It shows that the KNNLC algorithm is sensitive to specific distributed data and needs further improvement. The results on high-capacity sample data sets show that to a certain extent, the KNNLC algorithm may be more suitable for processing high-dimensional data, and its classification accuracy and stability are significantly better than the classic KNN algorithm. On the whole, in the classification problem, the KNNLC classification algorithm has better performance than the classic KNN classification algorithm, and the higher the data dimension, the more obvious this advantage. In fact, this is because KNNLC uses local coding to obtain neighbor samples, which meets the expectations of theoretical research, as shown in Table 3.

Table 3 shows the quantitative comparison of the experiments of the two algorithms on the four data sets, respectively, using the form of mean \pm standard deviation to make the results more accurate. For classification algorithms, accuracy is often one of the most concerned evaluation criteria. As can be seen from the above table, compared to the classic KNN classification algorithm, the KNNLC algorithm shows better performance in most cases, with an accuracy rate of 2 to 5 percentage points higher.

5. Conclusion

Research on the improvement of KNN algorithm and its application in the field of image processing. This article mainly

discusses and improves the traditional KNN algorithm and the mean filtering algorithm from the perspective of the nearest neighbor selection strategy. Therefore, in the process of obtaining the neighbors, the neighbors with higher similarity can be captured. At the same time, based on this idea, the template selection of the mean filter is regarded as the neighbor selection, and the membership function is combined to obtain a more effective filter template, and experiments have proved the advantages of the improved algorithm [27]. Aiming at the problem that traditional KNN algorithm is sensitive to data distribution, combined with sparse coding and LLC theory, using the nearest neighbor selection strategy of local coding, KNNLC algorithm is proposed, which improves the effect of classic KNN algorithm. Through experiments on multiple representative data sets, it is proved that the KNNLC algorithm has great advantages and potential compared with the classic KNN algorithm in classification performance. The KNNLC algorithm shows better performance in most cases, with an accuracy rate of 2 to 5 percentage points higher.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Y. Liang, C. Sun, J. Jiang, X. Liu, and Y. Xie, "An efficiency-improved clustering algorithm based on KNN under ultradense network," *IEEE Access*, vol. 8, pp. 43796–43805, 2020.
- [2] H. Guo, Y. Li, Y. Li, L. Xiao, and J. Li, "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 176–193, 2016.
- [3] K. Wang, X. Yu, Q. Xiong, Q. Zhu, and L. Zhao, "Learning to improve WLAN indoor positioning accuracy based on DBSCAN-KRF algorithm from RSS fingerprint data," *IEEE Access*, vol. 7, pp. 72308–72315, 2019.
- [4] R. Jin, Y. Wang, C. Niu, and W. Song, "A novel high precision and low consumption indoor positioning algorithm for internet of things," *IEEE Access*, vol. 7, pp. 86874–86883, 2019.
- [5] J. H. Du, "Automatic text classification algorithm based on gauss improved convolutional neural network," *Journal of Computational Science*, vol. 21, pp. 195–200, 2017.
- [6] X. Fang, Z. Jiang, L. Nan, and L. Chen, "Optimal weighted k-nearest neighbour algorithm for wireless sensor network fingerprint localisation in noisy environment," *IET Communications*, vol. 12, no. 10, pp. 1171–1177, 2018.
- [7] F. C. Yang, B. Tseng, C. Y. Lin, Y. J. Yu, A. Linacre, and C. I. Lee, "Population inference based on mitochondrial dna control region data by the nearest neighbors algorithm," *International Journal of Legal Medicine*, vol. 135, no. 4, pp. 1191–1199, 2021.
- [8] J. Ren, Y. Wang, C. Niu, W. Song, and S. Huang, "A novel clustering algorithm for wi-fi indoor positioning," *IEEE Access*, vol. 7, pp. 122428–122434, 2019.
- [9] J. Li, "An improved K-nearest neighbor algorithm using tree structure and pruning technology," *Intelligent Automation and Soft Computing*, vol. 25, no. 1, pp. 1–15, 2019.
- [10] V. Torra, Y. Narukawa, A. Honda, and S. Inoue, "Modeling decisions for artificial intelligence," *14th International Conference, MDAI 2017*, 2017, pp. 77–88, Kitakyushu, Japan, 2017.
- [11] K. A. Zweig, O. Deussen, and T. D. Krafft, "Algorithmen und meinungsbildung," *Informatik-Spektrum*, vol. 40, no. 4, pp. 318–326, 2017.
- [12] V. Subramaniaswam and R. Logesh, "Adaptive KNN based recommender system through mining of user preferences," *Wireless Personal Communications*, vol. 97, no. 2, pp. 2229–2247, 2017.
- [13] J. Hu, Y. Li, W. X. Yan, J. Y. Yang, H. B. Shen, and D. J. Yu, "KNN-based dynamic query-driven sample rescaling strategy for class imbalance learning," *Neurocomputing*, vol. 191, pp. 363–373, 2016.
- [14] S. Yin, H. Zheng, S. Xu, H. Rong, and N. Zhang, "A text classification algorithm based on feature library projection," *Journal of Central South University*, vol. 48, no. 7, pp. 1782–1789, 2017.
- [15] C. Yücelba, "A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 2, pp. 511–524, 2021.
- [16] Z. Qin, T. Kirubarajan, and Y. Liang, "Application of an efficient graph-based partitioning algorithm for extended target tracking using gm-phd filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 6, pp. 4451–4466, 2020.
- [17] H. Sun, X. Liu, Q. Deng, W. Jiang, and Y. Ha, "Efficient FPGA implementation of K-nearest-neighbor search algorithm for 3D LIDAR localization and mapping in smart vehicles," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 67, no. 9, pp. 1644–1648, 2020.
- [18] N. Bhaskar and P. M. Kumar, "Optimal processing of nearest-neighbor user queries in crowdsourcing based on the whale optimization algorithm," *Soft Computing*, vol. 24, no. 17, pp. 13037–13050, 2020.
- [19] D. A. Adeniyi, Z. Wei, and Y. Yang, "Personalised news filtering and recommendation system using Chi-square statistics-based K-nearest neighbour ((SB)-s-2-KNN) model," *Enterprise Information Systems*, vol. 11, no. 6-10, pp. 1283–1316, 2017.
- [20] H. J. Cho and J. Chae, "A safe exit algorithm for moving k nearest neighbor queries in directed and dynamic spatial networks," *Journal of Information Science and Engineering*, vol. 32, no. 4, pp. 969–993, 2016.
- [21] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Sensitivity of pansharpening methods to temporal and instrumental changes between multispectral and panchromatic data sets," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 55, no. 1, pp. 308–319, 2017.
- [22] U. K. Ghosh, K. K. Naik, and M. P. Kesari, "Digital image processing of multispectral aster imagery for delineation of alteration and related clay minerals in sakoli belt: Maharashtra – a case study," *Journal of the Geological Society of India*, vol. 88, no. 4, pp. 464–470, 2016.
- [23] Q. Xie, X. Chen, L. Li, K. Rao, L. Tao, and C. Ma, "Image fusion based on kernel estimation and data envelopment analysis," *International Journal of Information Technology & Decision Making*, vol. 18, no. 2, pp. 487–515, 2019.

- [24] L. Chen, X. Zhang, and H. Ma, "Sparse representation over shared coefficients in multispectral pansharpener," *Tsinghua Science and Technology*, vol. 23, no. 3, pp. 315–322, 2018.
- [25] C. Han, H. Zhang, C. Gao, C. Jiang, N. Sang, and L. Zhang, "A remote sensing image fusion method based on the analysis sparse model," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, vol. 9, no. 1, pp. 439–453, 2016.
- [26] D. Stevic, I. Hut, N. Dojcinovic, and J. Jokovic, "Automated identification of land cover type using multispectral satellite images," *Energy & Buildings*, vol. 115, pp. 131–137, 2016.
- [27] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: a meta-analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, no. 3, pp. 309–322, 2021.