Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 2, November 2024, pp. 1309~1318
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i2.pp1309-1318

Optimizing hyperspectral classification: spectral similarity-based band selection with chord k-means

Origanti Subhash Chander Goud[1], Thogarachetti Hitendra Sarma[2], Chigarapalle Shoba Bindu[1]
[1]Department of Computer Science and Engineering, College of Engineering, Jawaharlal Nehru Technological University (JNTU), Anantapur, India
[2]Department of Information Technology, Vasavi College of Engineering, Hyderabad, India

Article Info

Article history:

Received Mar 15, 2024
Revised Jul 27, 2024
Accepted Aug 5, 2024

Keywords:

Band selection
Chord algebra
Hyperspectral image
K-means
Spectral similarity

ABSTRACT

Band selection is crucial for achieving high classification accuracy in hyperspectral image (HSI) analysis, especially when ground truth data are limited. While unsupervised algorithms are preferred in such scenarios, the effectiveness of k-means clustering depends heavily on the choice of similarity measure. This article presents a novel two-level clustering approach for band selection. In the first level, bands are clustered using k-means with various similarity measures such as Euclidean distance, spectral angle mapper (SAM), and spectral information divergence (SID). Subsequently, the second level leverages the chord metric k-means clustering to form clusters of HSI scenes upon optimal band clusters from the first level. This initial band selection reduces dimensionality and guides subsequent k-means clustering. The proposed chord-based clustering method, utilizing the chord metric, outperforms standard k-means variants, demonstrating significant improvements in accuracy. Experimental results on publicly available hyperspectral datasets confirm the effectiveness of the proposed approach as an alternative to traditional k-means algorithms, showcasing significant improvements in accuracy.

Corresponding Author:

Origanti Subhash Chander Goud
Department of Computer Science and Engineering, College of Engineering
Jawaharlal Nehru Technological University (JNTU)
Anantapur, A.P., India
Email: organtsubhash@gmail.com

1. INTRODUCTION

Hyperspectral imaging (HSI) provides more details about the objects that are captured in the scene with rich spectral information. Having large spectral information is very useful to characterize the objects and classify them efficiently. On the other hand, having more spectral bands makes results in less accuracy when using any machine-learning algorithm for classification. Hence, spectral band selection has been an important study in HSI classification [1], [2]. Mixed pixel due to low spatial resolution demand more robust band selection or dimensionality reduction techniques [3]. HSI band selection is crucial for capturing essential ground object information, reducing redundancy, and reducing computational costs. With continuous research on HSI imagery, HSI image analysis has been used in various applications such as environmental monitoring, precision agriculture, mineral exploration, and urban planning [1].

Su and Du [4] classified band selection methods into six categories, with a primary focus on ranking-based methods that prioritize bands based on spectral importance and an objective function. These methods, whether supervised or unsupervised, demonstrate a systematic and versatile nature, crucial for efficient analysis or classification tasks. While recognizing their potential, the study acknowledges that

ranking-based approaches may lead to suboptimal band subset selection (BSS), highlighting the nuanced decision-making involved in their deployment across diverse scenarios. Chang and Liu [5] propose a statistical-based band prioritization method for hyperspectral analysis, utilizing measures like variance and skewness for dimensionality reduction while preserving critical spectral information. Despite its efficiency, potential risks of inaccurate BSS impacting classification accuracy are acknowledged. In a related context, Pal and Foody [6] highlight the support vector machine (SVM) approach's superior performance in discerning relevant features, with challenges noted in unsupervised datasets. Yu *et al.* [7] introduce the linearly constrained minimum variance BSS (LCMV-BSS) algorithm, optimizing band subsets in hyperspectral data for enhanced classification accuracy and computational efficiency. Meanwhile, Imani and Ghassemian [8] present the binary coding-based feature extraction (BCFE) method, surpassing principal component analysis (PCA) and linear discriminant analysis (LDA) in dimensionality reduction for hyperspectral analysis, particularly effective in capturing discriminative features for improved classification accuracy. Together, these studies underscore the nuanced considerations and challenges in optimizing hyperspectral data analysis through advanced feature selection techniques, offering promising avenues for improved dimensionality reduction and classification accuracy.

Du and Yang [9] explored efficient band selection in hyperspectral analysis through linear regression-based (LP) and orthogonal subspace projection (OSP) methods, prioritizing bands with high prediction errors for dimensionality reduction and enhanced computational efficiency. Qian *et al.* [10] proposed a novel feature selection approach using Kullback-Leibler divergence and kurtosis-based similarity matrices, followed by Affinity Propagation clustering, improving classification accuracy by selecting informative bands and reducing dimensionality. Keshava [11] emphasized the significance of distance metrics like SAM and SID, coupled with spectral libraries, for material identification in hyperspectral data, enhancing the selection of informative bands and improving discriminative capabilities. Rajakani *et al.* [12] discussed various band selection methods, including BCM, BDM, BCC, and BDC, aiming to minimize redundancy and enhance discriminative power in hyperspectral datasets through normalized correlation-based techniques. Finally, Thenkabail [13] highlighted band selection as a dimensionality reduction technique that selectively chooses informative bands without mathematically transforming the data, facilitating efficient processing and analysis by removing redundant or irrelevant bands while retaining crucial spectral information. Overall, these studies collectively contribute to advancing band selection methodologies for improved hyperspectral analysis.

K-means clustering can be improved by effectively choosing the better similarity measure the study by Gupta and Chandra [14] shows that similarity metric has an impact on clusters formed by k-means. Gupta and Chandra [15] highlight the significant role that distance/similarity metrics play in pattern recognition tasks. They specifically mention Euclidean, Manhattan, Mahalanobis, and Minkowski metrics, emphasizing their adaptability even when dealing with binary data. In a similar vein, Konstantin and Gribov [16] delve into covariance models and explore the use of kriging and the chordal metric for analyzing multidimensional data points, with a specific focus on spatial analysis. The studies by Kapil and Chawla [17], Gupta and Chandra [15] provide valuable insights into the application of similarity/distance metrics in k-means clustering across different domains. The research emphasizes the sensitivity of the k-means algorithm to metric choices and highlights the impact of altering similarity metrics on cluster formation. These findings contribute to a better understanding of the optimal metric selection for k-means clustering, particularly in the context of online user data and IoT/multimedia applications. Further research in this area could explore additional domains and properties to ascertain the broader applicability of these insights.

Khalifa *et al.* [18] investigated the application of clustering techniques with diverse similarity/ dissimilarity measures to select compounds from a chemical drug repository. The study focuses on a clustering approach termed "dissimilarity-based compound selection (DBCS)," aiming to identify a subset of chemical molecules from a drug database through agglomerative and hierarchical clustering methods. Agglomerative clustering employed the group-average technique with various similarity measures, while hierarchical clustering utilized the Jarvis-Patrick method. In the latter, molecules were added to a cluster if their nearest neighbor lists shared common elements with the cluster.

The studies by various authors [19]-[22] provide valuable insights into the application of similarity/distance metrics in k-means clustering across different domains. The research emphasizes the sensitivity of the k-means algorithm to metric choices, highlighting the impact of altering similarity metrics on cluster formation and the trade-offs associated with traditional and innovative metrics. These findings contribute to a better understanding of the optimal metric selection for k-means clustering in various contexts, such as online user data, internet of things (IoT)/multimedia applications, high-dimensional data, text document analysis, and higher-dimensional spaces. Further research in this area could explore additional domains and properties to ascertain the broader applicability of these insights and potentially lead to enhancements in clustering algorithms for improved outcomes. The research articles by Qiao *et al.* [22],

Vijay *et al.* [23] shed light on the significance of distance metrics in deep learning and clustering analysis. Qiao *et al.* [22] emphasized the superiority of cosine similarity in classifying HSI, while Vijay *et al.* [23] highlighted the nuanced and context-dependent nature of distance metric selection in clustering problems. Their findings provide valuable insights for the effective application of distance metrics in diverse domains of deep learning and clustering analysis.

Shirkhorshidi *et al.* [24] investigated the impact of similarity metrics on distance-based clustering in higher-dimensional spaces, introducing the effective chord metric with k-means clustering. Their study highlighted k-means' versatility across diverse datasets with various metrics. Meanwhile, Ghazal *et al.* [25] extensively assessed similarity metrics on k-means performance, emphasizing the consistent superiority of the Manhattan metric in execution time across diverse datasets and cluster sizes. Both studies underscored the critical role of metric selection in optimizing k-means and suggested exploring alternative algorithms for enhanced clustering outcomes across different datasets.

Lundholm and Svensson [26] present a comprehensive overview of Clifford geometry, delving into its origins and applications. Their study rigorously explores the Clifford framework, substantiating claims with proofs, and highlights applications in cybersecurity, image processing, neural networks, and geometric fields, showcasing the efficacy of Clifford algebra. Simultaneously, Hitzer *et al.* [27] extensively explore Clifford algebra applications, emphasizing its relevance in cybersecurity for robust data analysis and in image processing for advanced techniques. The integration with neural networks introduces innovative algorithms for complex tasks in machine learning and pattern recognition. Overall, both studies underscore the versatility and significance of Clifford algebra across diverse technological domains, offering valuable insights into its potential for advancing various fields, including computer graphics, design, and robotics.

As per Dorst [28] there is a brief discussion of the chord metric detailed analysis procedure, where Dorst has mentioned chord metric and its relationship with Clifford algebra. Dorst defines the chord metric utilization which follows the Clifford algebra principle and also proves how Clifford algebra principles are being followed by the chord metric defined in (10). As per drost, the chord metric follows the Clifford principle upon scalars, vector space, bi-vector space, and tri-vector spaces which are defined as Clifford algebraic blades defined in (1).

$$1, (e1, e2, e3), \big((e1 \Lambda e2), (e1 \Lambda e3), (e2 \Lambda e3)\big), (e1 \Lambda e2 \Lambda e3) \quad (1)$$

The first segment of (1) [28] is a scalar product of values, and the second part of (*e1, e2, e3*) is a vector space representation of Clifford Algebra Blades which separate the multidimensional representation by utilizing orthogonal subspaces. The third part of equation 1 is a Bi-vector space representation separated with a combination of orthogonal subspace blades like (*e1 ^ e2*) where e1 and e2 are two orthogonal subspace vectors and the last part (*e1 ^ e2 ^ e3*) is Tri-vector orthogonal representation of a vector which has three orthogonal subspace blades of Euclidean subspace representation. As per Ruhe *et al.* [29] the Clifford algebra is used for constructing deep neural networks (DNNs), Brandstetter *et al.* [30] offer a comprehensive study on the utilization of Clifford algebra, highlighting its advantages in deep learning. The research specifically emphasizes the application of Clifford geometry in constructing neural networks, leading to improved accuracy. The authors propose Clifford neural networks as an alternative implementation for solving partial differential equations (PDEs), showcasing promising results in comparison. Drost [28], Brandstetter *et al.* [30] described geometric Clifford algebra networks (GCANs) which are based on symmetric group transformations using this geometry of Clifford. These GCANs are more suitable for places that require manipulating geometric transformations for dynamic systems. In summary, the research endeavors discussed encompass band selection, clustering with various similarity metrics, and clustering algorithms.

Band selection methodologies, presented in [6], [8], [9], [12], provide a wide range of techniques from leveraging statistical variation to treating band selection as a ranking or regression problem, emphasizing the importance of reducing dataset dimensionality while capturing fundamental spectral information. Further, the unsupervised band selection approaches take advantage of the ability to handle unlabeled data without ground truth. Having the greater advantage of linear time complexity and ease of implementation, the k-means clustering method and its improvement with spectral similarities are revisited in the proposed work and a two-level clustering approach for band selection. The first level focuses on identifying the spectrally dissimilar bands and the second level clustering leverages the chord metric distance to further improve the accuracy of the band selection. This unsupervised approach effectively preserves spectral information, handling complex characteristics for accurate band selection and improved hyperspectral data analysis results.

2. METHOD

The key objective of the research work presented in this paper is to identify a subset of bands using an unsupervised (clustering) approach leveraging spectral similarity measure and to incorporate chord metric, improvising the classification accuracy of the traditional k-means clustering algorithm.

The proposed methodology runs in two steps:

a. Initial band selection by clustering the spectral bands using k-means with spectral similarity measure (KM-SSM).
b. Implement chord k-means (CKM) for optimal band selection.

Band selection is performed with k-means as shown in Algorithm 1 where bands selected for step 2 (CKM clustering) these chosen bands are the original represented bands of the original HSI data sets. In our proposed approach, the spectral similarity measures viz., spectral angel mapper (SAM), spectral information divergence (SID), hybrid measure with spectral angle mapper (SIDSAM), Jeffery-Matusita (JM), JM with SAM (JM-SAM), and normalised cross correlation (NCC) are used in place of Euclidean measure for spectral similarity. The mathematical formulae to compute the ED, SAM, SID, SIDSAM, JM, JM-SAM, and NCC are shown in (2)-(8). The step 2 process involves applying Algorithm 2 for band selection using (2)-(8). Then, the bands selected by Algorithm 1 are used as input for the Algorithm 2. By evaluating and fine-tuning 'r' for each set of bands, the goal is to improve the performance of the chord k-means algorithm, particularly for datasets with high dimensionality (more than 20 bands).

Algorithm 1. K-means clustering with SSM for band selection in HSI ($D_N \times_M$, k, $\pi^{(0)}$) (KM-SSM)

1. Initialize the seed cluster centers K-means++ initialization, denoted by $\mu^{(0)} = \{b_1^{(0)}, b_2^{(0)}, \ldots b_k^{(0)}\}$.
2. For each spectral band b_i in D, find its nearest spectral(l) band b_j using ED/SAM/SID/SIDSAM, and assign b_i to $C^l{}_j$.
3. Compute the updated mean spectral similarity of each cluster $C^l{}_j$.
4. Reassign each spectral band b_i in D, find its nearest spectral band $b_j{}^{(l+1)}$ using ED/SAM/SID/SIDSAM, and assign b_i to $C^l{}_j{}^{+1}$.
5. Repeat steps 2 through step 4 till convergence.
6. Identify one optimal band $b_i{}^{(0)}$ from each cluster such that $b_i{}^{(0)}$ is the most spectrally similar to the center of the cluster $C_i{}^{(0)}$.
7. Output: selected spectral bands from each cluster.

$$\mathrm{ED}\left(\mathrm{S_i}, \mathrm{S_j}\right) = \sqrt{\left\{\sum_{\{i,j=1\}}^{L}\left(\mathrm{S_i} - \mathrm{S_j}\right)^2\right\}} \tag{2}$$

$$SAM\left(S_i, S_j\right) = cos^{-1}(\theta),$$

$$\theta = \left[\left\{\frac{\sum_{i,j=1}^{L}\{S_i S_j\}}{\left\{\sqrt{\{\sum_{i=1}^{L}\{s_i^2\}\}}\right\} * \left\{\sqrt{\{\sum_{j=1}^{L}\{s_j^2\}\}}\right\}}\right\}\right] \tag{3}$$

$$SID\left(S_i, S_j\right) = \sum_{i=1}^{L} S_i * \log\left(\frac{S_i}{S_j}\right) + \sum_{i=1}^{L} S_j * \log\left(\frac{S_j}{S_i}\right) \tag{4}$$

$$SIDSAM\left(S_i, S_j\right) = SID\left(S_i, S_j\right) * \tan(SAM\,(S_i, S_j)) \tag{5}$$

$$JM\left(S_i, S_j\right) = \sqrt{\sum_{i=1}^{L}[\sqrt{S_i} - \sqrt{S_j}]^2} \tag{6}$$

$$JM - SAM\left(S_i, S_j\right) = JM\left(S_i, S_j\right) * \tan(SAM(S_i, S_j)) \tag{7}$$

$$NCC(i,j) = \frac{\sum_{x,y}[D_i(x,y) - \overline{D_i}] * [D_j(x,y) - \overline{D_i}]}{\sqrt{\sum_{x,y}(D_i(x,y) - \overline{D_i})^2} * \sqrt{\sum_{x,y}(D_j(x,y) - \overline{D_j})^2}} \tag{8}$$

$$Norm(S_i, X) = \frac{S_i - \overline{X}}{\sigma_{(x)}} \tag{9}$$

$$Cosine\_distance\left(S_i, S_j\right) = \frac{S_i, S_j}{S_i * S_j}, \Theta = \cos^{-1}(\,cosine\_distance\,)$$

$$Chord\left(S_i, S_j\right) = 2 * r * \sin\frac{\Theta}{2} \tag{10}$$

Algorithm 2. Chord k-means clustering with band selection HSI ($D_{L\times k}L = N \times M$ *Pixels, K, Lab*$^{(1)}$) (chord-KM)

1. Use reduced bands from Algorithm 1 for each method of (2) to (7) so that the data set is represented with the bands of the methods as $D_{L\times k} = \{b_1, b_2, \ldots, b_k\}$
2. Normalise the reduced $D_{L\times k}$ with Z-score normalisation as (9)
3. For $D_L\times_k$, randomlyinitialise the seeds as $\mu^{(0)} = \{p_1, p_2, \ldots, p_K\}$
4. For each point in $D_{L\times k}$ find its nearest point from $\mu^{(0)}$ seeds using chord metric in equation 10 and group the points according to the nearest seed as clusters C^1_j.
5. Compute the updated mean of each cluster C^1_j.
6. Reassign each point in $D_{L\times k}$, find its nearest seed updated in step 5.
7. Repeat steps 5 and step 6 till convergence and label each point according to cluster C^1_j.
8. Final labels of each point in $D_{L\times k}$ is compared to ground truth for accuracies.

Output: the labels of step 7 are $Lab^{(1)} = \{l_1, l_2, \ldots, l_k\}$

Here in (2) to (7) Si and Sj are the pixel vectors of an HSI image, wherein it holds the condition that i,j< L where L is the number of bands present in an HSI. In (8) Di (x,y), Dj (x,y) represents (x,y) indexed pixel value of an HSI band, and $\bar{D}$ represents the mean of the band pixels of HSI.

In (10) represents the chord metric that follows the Clifford Algebra principles as described by Dorst [28] where the principle of the chord is applied by slicing the multidimensional Sphere as shown in Figure 1. Where the chord metric is a slice of the sphere with a certain radius that variation of radius tries to differentiate the clusters formed by k-means with non-convex shapes. In Figure 1 (10) 'r' is the radius of the sphere which is sliced at a part of the sphere and 'Θ' is the angle formed by the radius of the sphere which is sliced, Figure 2 represents the quaternion representation of Clifford algebra of the sliced sphere where e1, e2, e3 are of (10).

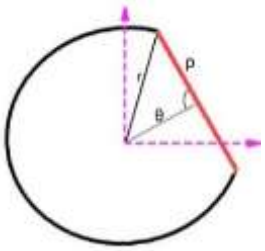

Figure 1. Chord metric of (10)

To enhance the approach, it's crucial to consider the effectiveness of both the 1 and 2 algorithms, as well as the suitability of the equations for band selection. Additionally, the dataset's characteristics will play a key role in determining the selected bands' relevance and the approach's overall success.

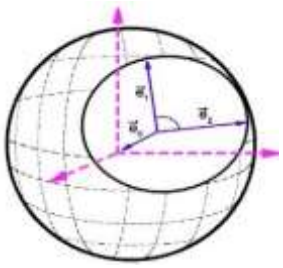

Figure 2. Quaternions representation of Clifford algebra [28]

3. RESULTS AND DISCUSSION

3.1. Datasets

Indian Pines were collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwestern Indiana, USA. The dataset contains 145×145 pixels, with 224 spectral bands covering the wavelength range from 0.4 to 2.5 micrometers, and has a spatial resolution of 20 meters per pixel in 16 different classes. Salinas was collected by the AVIRIS sensor over the Salinas Valley in California, USA. The dataset contains 512×217 pixels, with 224 spectral bands covering the wavelength

range from 0.4 to 2.5 micrometers, and has a spatial resolution of 3.7 meters per pixel. the ground truth contains 16 different classes.

Pavia Univeristy hyperspectral data was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. It has 610*610 pixels and the ground truth contains 9 classes. (Provided by Prof. Paolo Gamba). Pavia center contains 610 spectral bands and covers an area of approximately 610×340 pixels, with each pixel representing a square of 1.3 meters. This hyperspectral dataset has nine classes, which represent different land-cover types in the city of Pavia.

3.2. Results

This section represents the utilization of real-time HSI datasets in applying the proposed methodology. The results are presented in Tables 1 and 2 as the overall accuracy and kappa coefficient, where Algorithm 2 is used with a chord metric ranging from 2 to 3. The highlighted values indicate the highest accuracy achieved using the respective method described in each column.

Table 1. Overall accuracy of Pavia center

R	ED	SID	JM	JM-SAM	NCC	SAM	SID-SAM
2	**0.97**	0.95	0.97	0.98	0.97	0.97	0.95
2.05	0.97	0.95	0.96	0.98	0.96	0.97	0.94
2.1	0.97	0.97	0.97	0.98	0.97	0.97	0.97
2.15	0.96	0.94	0.97	0.98	0.97	0.97	0.96
2.2	0.97	0.95	0.97	0.98	**0.97**	0.97	0.97
2.25	0.97	0.97	0.97	0.97	0.97	0.97	0.95
2.3	0.96	0.96	0.97	0.98	0.97	**0.98**	0.95
2.35	0.97	0.97	0.96	0.98	0.97	0.98	0.97
2.4	0.97	0.95	0.97	0.98	0.97	0.97	0.97
2.45	0.96	0.96	0.97	0.98	0.96	0.97	0.95
2.5	0.95	0.96	0.96	0.98	0.97	0.97	0.95
2.55	0.96	**0.97**	0.97	0.98	0.97	0.97	0.97
2.6	0.97	0.97	0.97	**0.98**	0.97	0.95	0.96
2.65	0.97	0.97	0.97	0.98	0.97	0.95	0.97
2.7	0.97	0.95	0.97	0.98	0.96	0.97	0.97
2.75	0.97	0.94	0.96	0.98	0.97	0.97	0.95
2.8	0.96	0.97	0.97	0.98	0.96	0.97	0.97
2.85	0.97	0.97	0.97	0.98	0.97	0.97	0.97
2.9	0.97	0.97	0.96	0.96	0.94	0.96	0.95
2.95	0.97	0.97	0.97	0.97	0.97	0.95	0.95

Table 2. Kappa coefficient of Pavia center

R	ED	SID	JM	JM-SAM	NCC	SAM	SID-SAM
2	**0.91**	0.86	0.91	0.93	0.90	0.91	0.86
2.05	0.91	0.86	0.88	0.93	0.86	0.92	0.83
2.1	0.91	0.91	0.91	0.94	0.91	0.91	0.90
2.15	0.87	0.83	0.91	0.93	0.91	0.91	0.89
2.2	0.91	0.86	0.91	0.93	**0.91**	0.92	0.90
2.25	0.91	0.91	0.91	0.91	0.91	0.91	0.86
2.3	0.87	0.89	0.91	0.93	0.91	**0.93**	0.86
2.35	0.91	0.90	0.87	0.93	0.91	0.93	0.91
2.4	0.91	0.86	0.90	0.93	0.91	0.91	0.91
2.45	0.87	0.89	0.91	0.93	0.87	0.91	0.86
2.5	0.86	0.89	0.88	0.93	0.91	0.92	0.86
2.55	0.87	**0.91**	0.91	0.93	0.90	0.91	0.91
2.6	0.91	0.91	0.91	**0.94**	0.91	0.86	0.89
2.65	0.91	0.91	0.91	0.93	0.91	0.86	0.90
2.7	0.90	0.86	**0.91**	0.93	0.87	0.91	0.91
2.75	0.91	0.82	0.89	0.93	0.90	0.91	0.86
2.8	0.87	0.90	0.91	0.93	0.86	0.92	**0.91**
2.85	0.91	0.91	0.91	0.93	0.91	0.92	0.91
2.9	0.91	0.91	0.87	0.88	0.83	0.89	0.86
2.95	0.91	0.91	0.91	0.91	0.91	0.84	0.86

For instance, Table 1 represents the scene of Pavia center and the column labeled ‘R’ corresponds to the range of ‘r’ values in (9). Additionally, the columns which are implementation of Algorithm 1 KM-SSM band selection methods which are defined as follows:

- ED: represents the ‘Euclidean distance’ method applied in (2).
- SAM: denotes the ‘Spectral angle mapper’ as defined in (3).

- SID: denotes the 'Spectral information divergence' as defined in (4).
- SIDSAM: denotes the 'Spectral information divergence' with 'spectral angle mapper' as defined in (5).
- JM: denotes the 'Jeffereys Matusita' as defined in (6).
- JM-SAM': denotes the 'Jeffereys Matusita' with 'spectral angle mapper' as defined in (7).
- NCC: denotes the 'normalised cross co-relation' in (8).

Figure 3 represents the graphical representation of Algorithms 1 and 2 and the workflow of the complete proposed method. Algorithms 1 and 2 are applied one after the other, and the result of Algorithm 1 is given as input to Algorithm 2. The ground truth and classification images of the Pavia center after the proposed work are represented in Figures 3 to 5. Tables 1 and 2 provide insights into the accuracy achieved by the proposed method, with columns like ED, SID, representing various techniques for band selection from Algorithm 1. Each column in the tables corresponds to the accuracy achieved by applying these band selection methods with chord metric-based clustering by Algorithm 2. The first column, denoted as R, signifies the chord metric radius, varying between 2 and 3 with an interval of 0.05. Notably, the best accuracy varies for each band subset generated by Algorithm 1. JM-SAM stands out with the highest accuracy, indicated by a kappa value of 0.94 in Table 2. In comparison to other seeding mechanisms like Random-Seeding and Kmeans++, "chord k-means with spectral similarity measures-based band selection (CKM-SSB)" demonstrates superior performance. Specifically, for the JM-SAM band subset, the Kappa values are 0.722 for random seeding, 0.694 for KM++, and notably higher at 0.942 for CKM-SSB, highlighting its efficacy in achieving improved accuracy.

The KM-SSM-Chord-KM algorithm exhibits superior performance compared to traditional k-means methods, namely KM-RS and KM++, as demonstrated on the Pavia center dataset. With a Kappa value of 0.942, KM-SSM-Chord-KM outperforms other k-means approaches (Kappa values around 0.72) shown in Table 3 when coupled with various band selection strategies such as SID and JM. The results are consistent across different datasets, highlighting the algorithm's robustness and efficacy. The research suggests that KM-SSM-Chord-KM holds promise for HSI analysis, potentially surpassing current state-of-the-art clustering algorithms. Further exploration into its generalizability and avenues for future research is warranted. The result of Algorithm 1 is represented in Table 4 where the bands represent from 0 to the number of bands present in the bands set, band 0 represents the first band of the data set.

Figure 6 demonstrates the comparative impact of the proposed algorithm method against traditional mechanisms like k-means with standard random seeding and k-means++ seeding. It reveals a marginal decrease in accuracy when employing traditional methods, whereas CKM notably enhances dataset accuracy. Interestingly, CKM consistently outperforms traditional methods across various metrics, including ED, SID, SAM, NCC, JM, and hybrid methods like SID-SAM and JM-SAM. This improvement is evident even with metric variations in the k-means algorithm.

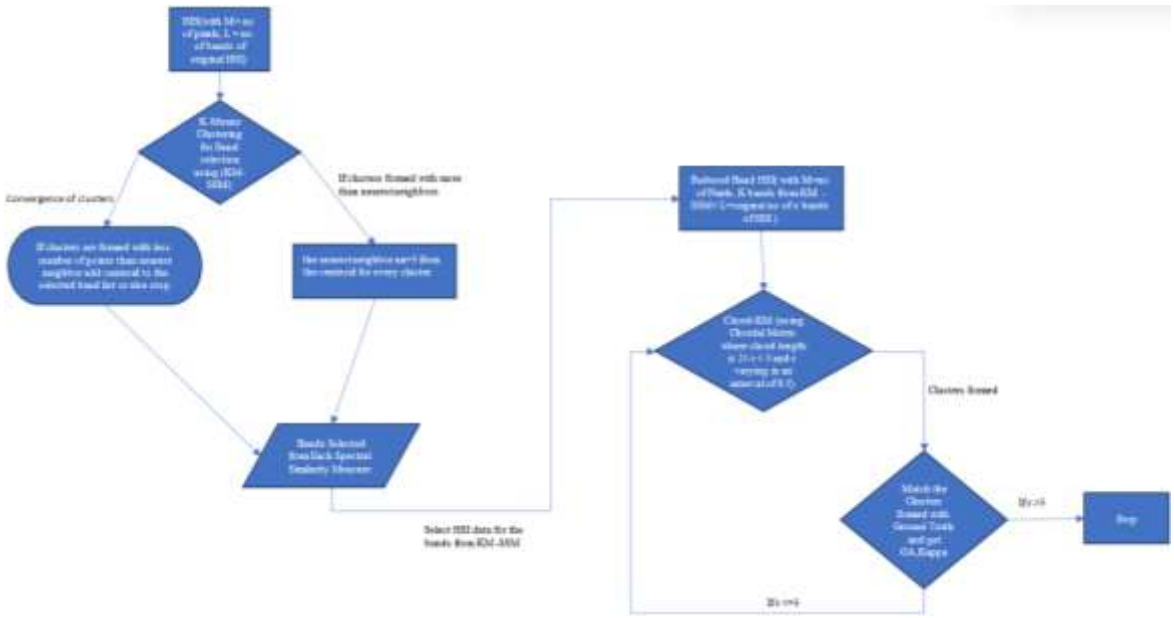

Figure 3. Flowchart of proposed work

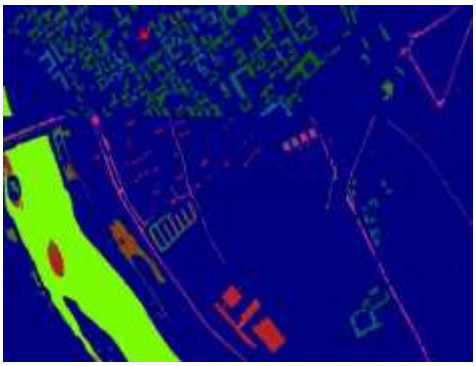

Figure 4. Classification map of Pavia center by the proposed method with r = '2.6'

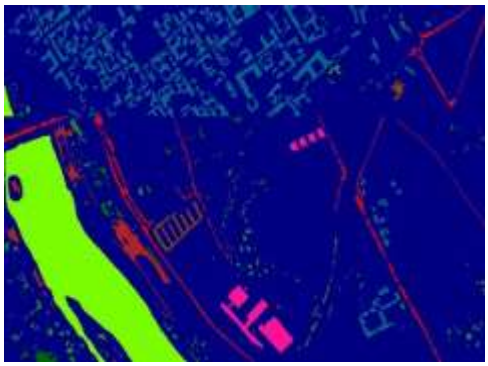

Figure 5. Ground truth map of Pavia center

Table 3. Kappa of Pavia center data set

Method	KM+RS	KM++	SSB-CKM
ED	0.544	0.725	0.914
SID	0.558	0.689	0.909
SAM	0.628	0.701	0.932
SIDSAM	0.558	0.689	0.907
NCC	0.622	0.686	0.912
JM	0.480	0.722	0.914
JM-SAM	0.722	0.694	**0.942**

Table 4. Bands selected using the proposed method for Pavia center

Data set	Method	Bands selected from SSB methods
Pavia center	ED	0/ 1/ 2/ 3/ 4/ 5/ 6/ 7/ 16/ 17/ 18/ 28/ 29/ 30/ 42/ 43/ 44/ 69/ 70/ 71/ 83/ 84/ 85
	SAM	0/ 1/ 2/ 3/ 11/ 12/ 13/ 25/ 26/ 27/ 38/ 39/ 40/ 55/ 56/ 57/ 72/ 73/ 74/ 83/ 84/ 85
	SID	0/ 1/ 2/ 3/ 4/ 5/ 6/ 8/ 9/ 10/ 24/ 25/ 26/ 41/ 42/ 43/ 68/ 69/ 70/ 83/ 84/ 85
	SID-SAM	0/ 1/ 2/ 3/ 4/ 5/ 6/ 8/ 9/ 10/ 24/ 25/ 26/ 41/ 42/ 43/ 68/ 69/ 70/ 83/ 84/ 85
	NCC	0/1/2/3/4/5/6/15/16/17/31/32/33/52/53/54/72/73/74/83/84/85
	JM	0/ 1/ 2/ 3/ 4/ 5/ 6/ 7/ 8/ 9/ 10/ 11/ 12/ 17/ 18/ 19/ 26/ 27/ 28/ 36/ 37/ 38/ 54/ 55/ 56/ 71/ 72/ 73/ 83/ 84/ 85
	JM-SAM	2/ 3/ 11/ 14/ 21/ 23/ 26/ 27/ 29/ 32/ 33/ 34/ 36/ 39/ 40/ 41/ 42/ 43/ 44/ 45/ 46/ 47/ 48/ 49/ 50/ 56/ 57/ 62/ 63/ 65/ 66/ 67/ 70/ 71/ 72/ 74/ 78/ 80/ 84/ 85/ 86/ 98

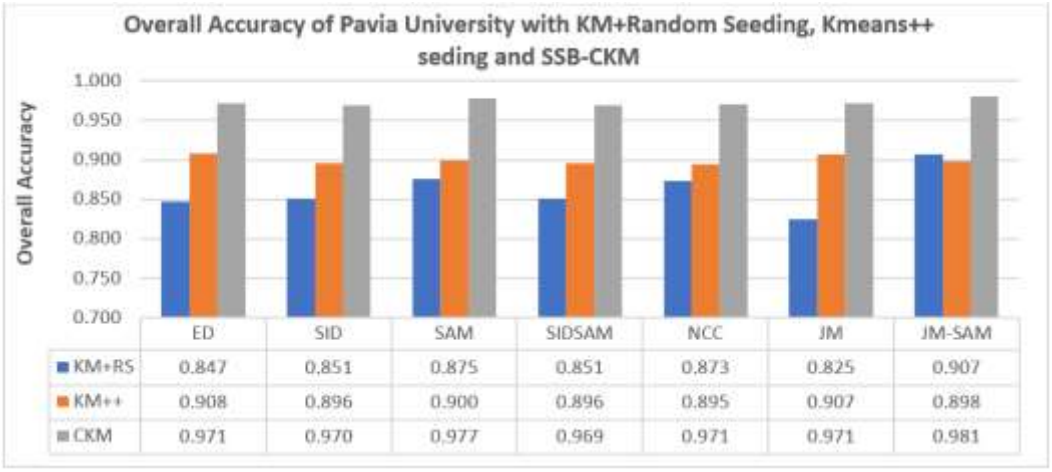

Figure 6. Comparison of the proposed method with different k-means variants with overall accuracy

4. CONCLUSION

The proposed approach demonstrates a notable enhancement over conventional k-means clustering techniques employing random seeding and k-means++ seeding. Specifically, k-means-SSM clustering is

executed for feature selection. It becomes evident that incorporating DR into the chord k-means framework contributes to improved results. The proposed chord metric-based k-means (chord-k-means) clustering finding the exact 'r' value which yields the best accuracy is another problem perspective. This strategic integration of DR enhances the algorithm's ability to extract relevant features, leading to a more refined and effective clustering process. Furthermore, exploring the impact of different seeding methods on the execution of chord k-means provides promising results.

REFERENCES

[1] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, Dec. 2017, doi: 10.1109/MGRS.2017.2762087.

[2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, Jun. 2013, doi: 10.1109/MGRS.2013.2244672.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002, doi: 10.1109/79.974718.

[4] W. Sun and Q. Du, "Hyperspectral band selection: a review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, Jun. 2019, doi: 10.1109/MGRS.2019.2911100.

[5] C. I. Chang and K. H. Liu, "Progressive band selection of spectral unmixing for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2002–2017, Apr. 2014, doi: 10.1109/TGRS.2013.2257604.

[6] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2297–2307, May 2010, doi: 10.1109/TGRS.2009.2039484.

[7] C. Yu, M. Song, and C. I. Chang, "Band subset selection for hyperspectral image classification," *Remote Sensing*, vol. 10, no. 1, p. 113, Jan. 2018, doi: 10.3390/rs10010113.

[8] M. Imani and H. Ghassemian, "An unsupervised feature extractionl method for classification of hyperspectral images," *22nd Iranian Conference on Electrical Engineering, ICEE 2014*, pp. 1389–1394, 2014, doi: 10.1109/IranianCEE.2014.6999750.

[9] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, Oct. 2008, doi: 10.1109/LGRS.2008.2000619.

[10] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Computer Vision*, vol. 3, no. 4, pp. 213–222, 2009, doi: 10.1049/iet-cvi.2009.0034.

[11] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1552–1565, Jul. 2004, doi: 10.1109/TGRS.2004.830549.

[12] M. Rajakani, T. E. D., and S. Rajesh, "Band selection for hyperspectral image classification using statistical features," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 1133–1139, 2018, Available: 59.pdf (acadpubl.eu).

[13] P. S. Thenkabail, "Hyperspectral data processing: algorithm design and analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 6, pp. 441–442, Jun. 2015, doi: 10.14358/pers.81.6.441.

[14] M. K. Gupta and P. Chandra, "Effects of similarity/distance metrics on k-means algorithm with respect to its applications in IoT and multimedia: a review," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37007–37032, Nov. 2022, doi: 10.1007/s11042-021-11255-7.

[15] M. K. Gupta and P. Chandra, "An empirical evaluation of k-means clustering algorithm using different distance/similarity metrics," in *Lecture Notes in Electrical Engineering*, vol. 605, 2020, pp. 884–892.

[16] K. Krivoruchko and A. Gribov, "Distance metrics for data interpolation over large areas on Earth's surface," *Spatial Statistics*, vol. 35, p. 100396, Mar. 2020, doi: 10.1016/j.spasta.2019.100396.

[17] S. Kapil and M. Chawla, "Performance evaluation of K-means clustering algorithm with various distance metrics," in *1st IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems, ICPEICES 2016*, Jul. 2017, pp. 1–4, doi: 10.1109/ICPEICES.2016.7853264.

[18] A. Al-Khalifa, M. Haranczyk, and J. Holliday, "Comparison of nonbinary similarity coefficients for similarity searching, clustering and compound selection," *Journal of Chemical Information and Modeling*, vol. 49, no. 5, pp. 1193–1201, May 2009, doi: 10.1021/ci8004644.

[19] X. Gu, P. P. Angelov, D. Kangin, and J. C. Principe, "A new type of distance metric and its use for clustering," *Evolving Systems*, vol. 8, no. 3, pp. 167–177, Sep. 2017, doi: 10.1007/s12530-017-9195-7.

[20] A. Singh, A. Yadav, and A. Rana, "K-means with three different distance metrics," *International Journal of Computer Applications*, vol. 67, no. 10, pp. 13–17, Apr. 2013, doi: 10.5120/11430-6785.

[21] F. A. Allah, W. I. Grosky, and D. Aboutajdine, "Document clustering based on diffusion maps and a comparison of the k-means performances in various spaces," in *Proceedings - IEEE Symposium on Computers and Communications*, Jul. 2008, pp. 579–584, doi: 10.1109/ISCC.2008.4625693.

[22] X. Qiao, H. Wu, S. K. Roy, and W. Huang, "Hyperspectral image classification based on 3d sharpened cosine similarity operation," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2023, vol. 2023-July, pp. 7669–7672, doi: 10.1109/IGARSS52108.2023.10281949.

[23] K. Vijay, K. C. Jitender, and K. Dinesh, "Performance evaluation of distance metrics in the clustering algorithms," *INFOCOMP Journal of Computer Science*, vol. 13, pp. 38–52, 2014, [Online]. Available: https://infocomp.dcc.ufla.br/index.php/INFOCOMP/article/view/21.

[24] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Ying Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS ONE*, vol. 10, no. 12, p. e0144059, Dec. 2015, doi: 10.1371/journal.pone.0144059.

[25] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[26] D. Lundholm and L. Svensson, "Clifford algebra, geometric algebra, and applications," *arXiv preprint arXiv:0907.5356*, 2009, [Online]. Available: http://arxiv.org/abs/0907.5356.

[27] E. Hitzer, T. Nitta, and Y. Kuroe, "Applications of Clifford's geometric algebra," *Advances in Applied Clifford Algebras*, vol. 23, no. 2, pp. 377–404, Jun. 2013, doi: 10.1007/s00006-013-0378-4.

[28] L. Dorst, "Optimal combination of orientation measurements under angle, axis and chord metrics," in *SEMA SIMAI Springer Series*, vol. 13, 2021, pp. 47–88.
[29] D. Ruhe, J. K. Gupta, S. de Keninck, M. Welling, and J. Brandstetter, "Geometric Clifford algebra networks," *Proceedings of Machine Learning Research*, vol. 202, pp. 29306–29337, 2023.
[30] J. Brandstetter, R. van den Berg, M. Welling, and J. K. Gupta, "Clifford neural layers for PDE modeling," *arXiv preprint arXiv:2209.04934*, 2022, [Online]. Available: http://arxiv.org/abs/2209.04934.

BIOGRAPHIES OF AUTHORS

Origanti Subhash Chander Goud is a research scholar at the Department of Computer Science and Engineering, JNTU College of Engineering Ananthapur, Jawaharlal Nehru Technological University Ananthapur, Andhra Pradesh, India. He is pursuing a Ph.D. in Computer Science and Engineering at JNTUA. He holds a Masters's degree in Computer Engineering from JNTU Hyderabad and Masters's degree in Computer Application from Osmania University. His research areas are hyperspectral image analysis, pattern recognition, and machine learning. He can be contacted at email: organtsubhash@gmail.com.

Dr. Thogarachetti Hitendra Sarma obtained Ph.D. in machine learning from JNT University, Anantapur, Andhra Pradesh, India in the year 2013. He is a recipient of the Teachers Associateship for Research Excellence (TARE) grant by SERB-DST Govt. of India. He has published over 25 articles in peer-reviewed Journals and reputed international conferences like IJCNN, CEC, PReMI and others. He delivered an invitation at FSDM -2017 in Taiwan. He is a senior member of IEEE. His research areas include machine learning, hyperspectral image processing, and data Mining. He can be contacted at email: t.hitendrasarma@gmail.com.

Prof. Dr. Chigarapalle Shoba Bindu Ph.D. in CSE from JNTUA, Anantapuramu, Andhra Pradesh. She is currently working as a Professor in the Department of CSE, JNTUA College of Engineering, Ananthapuramu. Her research areas include computer networks, network security, machine learning, and cloud computing. She can be contacted at email: shobabindhu@gmail.com.