Metric-Driven Similarity Indices: Redefining Spectral Distance Comparisons in Hyperspectral Data

Jungkwon Kim[1]
ELROILAB Inc.
Seoul, Republic of Korea
jkkim@elroilab.com

Jihoon Jung[1]
ELROILAB Inc.
Seoul, Republic of Korea
dltkr0210@elroilab.com

Jungi Lee
ELROILAB Inc.
Seoul, Republic of Korea
ganbbang12@elroilab.com

Kwangsun Yoo
ELROILAB Inc.
Seoul, Republic of Korea
yks@elroilab.com

Seok-Joo Byun*
ELROILAB Inc.
Seoul, Republic of Korea
sjbyun@elroilab.com

***Abstract*—Hyperspectral images (HSIs) offer rich spectral details but pose challenges in analyzing spectral vector distances due to high dimensionality and inter-class similarity. Existing distance metrics, while effective in specific cases, often fail to provide consistent comparisons across different tasks due to varying scales. This study proposes novel similarity score indices that normalize metrics onto a unified scale, ensuring fair, interpretable comparisons tailored to the unique properties of HSIs. Our evaluations on public datasets reveal the indices' ability to improve accuracy and reliability in spectral similarity assessments, addressing key challenges in HSI analysis.**

***Index Terms*—Hyperspectral Images, Distance Metrics, Similarity**

I. Introduction

Hyperspectral images have emerged as a powerful analytical tool across diverse fields, including remote sensing [14], [21], environmental monitoring [20], medical diagnostics [3], and food inspections [19] by capturing a continuous spectrum for each pixel across hundreds of spectral bands. This rich spectral data enables highly precise material discrimination and composition analysis through unique spectral signatures. However, hyperspectral datasets' high dimensionality and inherent characteristics pose significant challenges in data analysis. Spectral signatures from different classes can sometimes appear remarkably similar, while spectra within the same class often exhibit considerable variability or inconsistency. These factors necessitate developing appropriate measures that effectively capture class-specific spectral characteristics, such as shape and scale, to ensure accurate analysis [22]–[25].

Many types of distance metrics—such as Manhattan and Euclidean distances [7] and cosine similarity—are used to quantify dissimilarity between spectral vectors. However, the choice of metric significantly impacts the resulting distance distributions, often making it difficult to directly compare metrics or determine the most effective one for a given analysis. Various composite metrics have been proposed to resolve these challenges, which often combine well-known metrics through multiplication or as composite functions to capture more nuanced similarities or dissimilarities [6], [12], [16], [17]. While such combined metrics can sometimes improve accuracy in specific applications, their increased computational complexity may result in diminishing returns, especially when applied to large-scale hyperspectral data. This trade-off between accuracy and computational cost becomes particularly critical in AI model training, where efficiency is as important as precision. Furthermore, these combined metrics often fail to address the issue of different scales across individual metrics, which can exacerbate the difficulty of fair comparisons and metric selection.

These two challenges, differences in scale and computational demands, restrict the applicability of HSI analysis. For example, while Keshava [9] proposed a Spectral Angle Mapper (SAM)-based band selection method, their analysis focused on SAM and Euclidean distance metrics, which limits its ability to improve discrimination by leveraging a broader range of distance metrics for hyperspectral data analysis. Additionally, while Deborah, Richard, and Hardeberg [4] analyzed various metric functions categorized by their properties, their study did not account for the complexities of public hyperspectral datasets, which often exhibit significant inter-class similarity and intra-class variability [5], [18], [26]. This high variability in spectral signatures within and between classes makes similarity measurement particularly challenging.

In addition to these issues, the type and domain of hyperspectral data strongly influence metric performance [2], [15]. Datasets from natural landscapes, urban settings, or biomedical samples each exhibit unique spectral profiles and data distri-

This research was supported by RS-2023-00304845.
[1]These authors contributed equally
*Corresponding author

butions, often require suitable distance metrics that capture relevant similarities best. Metrics that are sensitive to fine spectral variations may perform well in vegetation and land cover discrimination but underperform in urban environments, where clear separations between classes are more valuable [11]. This underscores the importance of metric selection aligning with hyperspectral datasets' specific characteristics.

To address the challenge of comparing different metrics, this study proposes a similarity score index tailored to each metric. This index enables more consistent and interpretable comparisons across metrics by aligning to the uniform range of scores. By focusing on spectral characteristics specific to hyperspectral data, this approach ensures a more accurate evaluation of similarity. Moreover, we emphasize the importance of selecting distance metrics that are adapted to the particularities of hyperspectral datasets, as this choice is crucial for achieving optimal results across diverse application tasks. Given the complexities outlined above, the proposed similarity score indices bridge the gap between theoretical properties and practical applications, offering a more interpretable and standardized framework for hyperspectral image analysis.

This study presents a comprehensive evaluation of distance metrics in HSI analysis, aiming to address the following key questions: (1) How do different distance metrics perform in hyperspectral image analysis, and how do they capture spectral similarities across datasets with varying properties? (2) How can distance metrics be reformulated to enable more consistent and interpretable comparisons in hyperspectral image analysis? (3) What is the most suitable distance metric index for hyperspectral image analysis?

By analyzing and comparing a range of distance metrics across multiple types of hyperspectral datasets, we explore the potential for a more standardized framework to assess the effectiveness of various distance metrics in hyperspectral imaging, focusing on the proposed indices that best capture spectral dissimilarities and similarities in specific dataset contexts. Our contributions are summarized as follows.

- We systematically evaluate numerous distance metrics in HSI analysis, assessing their ability to capture relevant spectral similarities and considering their sensitivity to dataset-specific properties.
- We introduce a similarity score index tailored to each distance metric, enabling a more consistent and interpretable comparison.
- We propose the most suitable index based on the basic distance metric by analyzing carefully implemented experiments and taking diverse features of hyperspectral data into account.

II. Characteristic and Trade-offs of Common Distance Metrics

Distance metrics are fundamental in HSI analysis for quantifying spectral dissimilarity between pixels or spectral vectors. Widely used metrics, including Cosine similarity, Manhattan, Euclidean, Chebyshev, and others like Canberra, Chi-square, and Jeffrey distances, each have distinct properties that impact inter-class separability, a critical factor for tasks like classification and clustering. This section introduces the characteristics, strengths, and weaknesses of these metrics, emphasizing their suitability for different analytical scenarios and the trade-offs associated with their application.

A. Metrics in Hyperspectral Contexts

Cosine Similarity: Cosine similarity measures the angular difference between spectral vectors, making it insensitive to magnitude and effective for datasets where the shape of spectral curves matters more than intensity. Spectral Angle Mapper (SAM) [10], derived from cosine similarity, enhances this approach by introducing rotational invariance, focusing solely on angular relationships. While SAM performs well under varying lighting conditions, it may neglect amplitude differences crucial in fields like agriculture and environmental monitoring. Its reliance on trigonometric computations also creates challenges in large-scale or high-dimensional datasets. To mitigate these issues, hybrid methods like combining Spectral Information Divergence with angular metrics, proposed by Chang [1], integrate shape-based similarity with statistical divergence, improving performance across diverse applications while retaining the core benefits of cosine similarity.

Manhattan Distance: The computational simplicity and robustness to outliers of the Manhattan distance make it particularly useful in noisy datasets or those non-uniform distributions, where extreme values might influence other metrics. However, its reliance on absolute differences limits its sensitivity to subtle spectral variations, which can be crucial in datasets requiring fine-grained separability, such as those containing agricultural classes with overlapping spectral signatures.

Euclidean Distance: Euclidean distance is one of the most widely used metrics due to its intuitive geometric interpretation, representing the straight-line distance between spectral vectors. It is effective in datasets with well-separated classes and well-distributed spectral characteristics. However, it is sensitive to scaling and high-dimensional noise, which can lead to inflated distances in hyperspectral data. Preprocessing techniques, such as normalization or dimensionality reduction, are often necessary to ensure reliable performance.

Chebyshev Distance: The Chebyshev metric is particularly suited for identifying large discrepancies and outliers in the data, which can be advantageous for datasets with distinct, large separations. However, it tends to overlook smaller but meaningful spectral differences, which can reduce its effectiveness in cases where the classes are spectrally similar, such as in datasets with fine spectral variations between classes.

Canberra Distance: The Canberra distance [13] is highly sensitive to small relative differences between spectral values. This property is advantageous for distinguishing between spectral vectors with subtle differences, which is often the case in datasets with high spectral overlap, such as agricultural datasets. However, its sensitivity to small differences can also amplify noise, particularly in low-intensity spectral bands, leading to inconsistent results in noisy datasets.

Chi-square Distance: The chi-square distance emphasizes proportional differences, making it effective for datasets with uneven spectral distributions, where smaller spectral values play a significant role in class separability. This metric is sensitive to discrepancies in smaller values, which is beneficial when dealing with spectral distributions that are not uniform. However, like the Canberra distance, it is prone to amplifying noise, particularly in datasets with low signal-to-noise ratios, and may require careful preprocessing.

Jeffrey Distance: The Jeffrey distance [8], derived from information theory, is based on the Kullback-Leibler (KL) divergence which is defined as

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

where $P(i), Q(i)$ are discrete probability distributions. Then the Jeffrey distance is a symmetrical form of KL divergence, defined as $D_{\mathrm{Jef}}(P,Q) = D_{\mathrm{KL}}(P\|Q) + D_{\mathrm{KL}}(Q\|P)$, and is particularly effective for capturing dissimilarities in spectral distributions. It is useful in applications where statistical differences are critical, such as in the comparison of materials with varying spectral properties. However, its computational intensity and sensitivity to small probabilities can make it challenging to apply in practice, especially when dealing with large datasets or noisy data.

III. Methodologies

This section introduces the definitions of various metrics discussed in Section II, and based on these, we propose carefully formulated similarity score indices. Additionally, we describe the datasets used for the analysis.

A. Similarity Score Indices induced from the Distance Metrics

Distance metrics serve as the foundation for proposing similarity score indices to measure spectral similarity and dissimilarity in a unified manner across various metrics. To begin with, let $\mathbf{x} = ([x_i]_{i=1}^n)^T \in \mathbb{R}^n$ be a vector in n-dimensional Euclidean space. Then, we shall give the basic definitions of a metric space on Euclidean spaces for the sake of completeness.

Definition 3.1: Let $\mathfrak{M} \subset \mathbb{R}^n$ be a metric space with a distance function d satisfying the following properties:

- Non-negativity : For all $\mathbf{x}, \mathbf{y} \in \mathfrak{M}$, $0 \le d(\mathbf{x}, \mathbf{y}) < \infty$.
- Reflexivity : $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$.
- Symmetry : $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathfrak{M}$.
- Triangle inequality : $d(\mathbf{x}, \mathbf{y}) \le d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathfrak{M}$.

It is well known that the ℓ^p space is a metric space equipped with the norm defined as

$$\|\mathbf{x}\|_{\ell^p} = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad p \ge 1. \tag{1}$$

Throughout this paper, $\|\cdot\|$ without specification stands for the ℓ^2 norm.

Let $S^i, S^j \in \mathbb{R}^\lambda$ be spectrum data obtained across λ number of bands in classes i and j with s_k^i represents k-th data in S^i. Then, we give the formulae for the metrics as follows:

$$d_{\cos}(S^i, S^j) = \frac{S^i \cdot S^j}{\|S^i\|\|S^j\|}, \tag{2}$$

$$d_{\ell^p}(S^i, S^j) = \|S^i - S^j\|_{\ell^p}, \tag{3}$$

$$d_{\mathrm{Can}}(S^i, S^j) = \sum_k \frac{|s_k^i - s_k^j|}{s_k^i + s_k^j}, \tag{4}$$

$$d_{\chi^2}(S^i, S^j) = \frac{1}{2} \sum_k \frac{(s_k^i - s_k^j)^2}{s_k^i + s_k^j}, \tag{5}$$

$$d_{\mathrm{Jef}}(S^i, S^j) = \sum_k \left(s_k^i \log \frac{s_k^i}{s_k^j} + s_k^j \log \frac{s_k^j}{s_k^i} \right). \tag{6}$$

Depending on the metric, the interpretation of the distance values varies. For instance, in ℓ^p distances (e.g., Manhattan ($p = 1$), Euclidean ($p = 2$), and Chebyshev ($p = \infty$) distances), larger values indicate greater dissimilarity, while in distances like cosine similarity, larger values indicate higher similarity. Metrics also differ in the scale of their distance values, as can be seen in (2)−(6). This variability calls for a standardized similarity score that aligns with a common scale, ensuring more consistent and meaningful comparisons.

Now, we introduce similarity score indices mapping distance values to a common range $[0, 1]$. For metrics other than cosine similarity, the similarity score $\mathcal{S}$ is defined as:

$$\mathcal{S} = 1 - \frac{d}{d_{\max}}. \tag{7}$$

Here, d represents the computed distance, and $d_{\max}$ denotes the properly designed maximum distance across all spectrum bands. For cosine similarity, the similarity score is identical to the computed distance, $\mathcal{S}_{\cos}(S^i, S^j) = d_{\cos}(S^i, S^j)$. The complete definitions of similarity score indices for the remaining metrics (3)−(6) are defined as follows:

$$\mathcal{S}_{\mathrm{Area}}(S^i, S^j) = 1 - \frac{\|S^i - S^j\|_{\ell^1}}{\lambda \cdot \max_k\{s_k^i, s_k^j\}}, \tag{8}$$

$$\mathcal{S}_{\ell^p}(S^i, S^j) = 1 - \frac{\|S^i - S^j\|_{\ell^p}}{\|\max\{s_k^i, s_k^j\}\|_{\ell^p}}, \tag{9}$$

$$\mathcal{S}_{\mathrm{Can}}(S^i, S^j) = 1 - \frac{d_{\mathrm{Can}}(S^i, S^j)}{\lambda \cdot \max_k \left\{ \frac{s_k^i}{s_k^i + s_k^j}, \frac{s_k^j}{s_k^i + s_k^j} \right\}}, \tag{10}$$

$$\mathcal{S}_{\chi^2}(S^i, S^j) = 1 - \frac{d_{\chi^2}(S^i, S^j)}{\lambda/2 \cdot \max_k \left\{ \frac{(s_k^i)^2}{s_k^i + s_k^j}, \frac{(s_k^j)^2}{s_k^i + s_k^j} \right\}}, \tag{11}$$

$$\mathcal{S}_{\mathrm{Jef}}(S^i, S^j) = 1 - \frac{d_{\mathrm{Jef}}(S^i, S^j)}{\lambda \cdot \max_k \left\{ s_k^i \log \frac{s_k^i}{s_k^j} + s_k^j \log \frac{s_k^j}{s_k^i} \right\}}. \tag{12}$$

We provide a formal definition of $\max$ in the indices above. Let $\max\{a, b\}$ denote the greater of the two values a and b, and let $\max_k\{a_k\}$ represent the maximum value in the sequence $\{a_k\}$ over the index k. By combining these two, we see $\max_k\{a_k, b_k\} = \max_k\{\max\{a_k, b_k\}\}$. In (8), as a

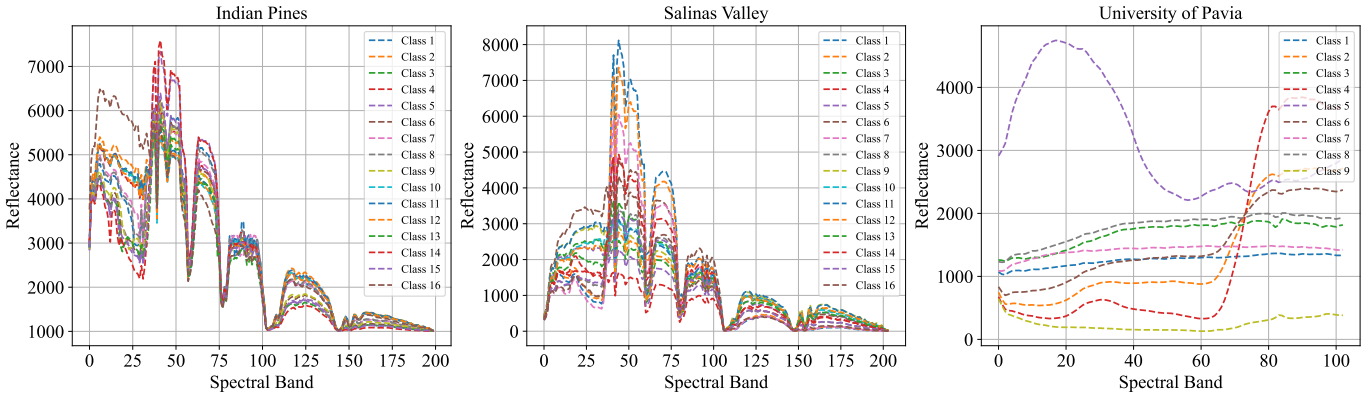

Fig. 1: Mean spectral distributions of three hyperspectral datasets – Indian Pines, Salinas Valley, and University of Pavia.

variation derived from ℓ^p distances, we additionally define the relative area index, offering an alternative perspective for interpreting spectral data through the lens of its relative area. Adopting these similarity score indices, we establish a foundation for a unified framework in HSI analysis, eliminating biases introduced by varying metric scales.

B. Datasets

Proposed similarity score indices are evaluated on three publicly available hyperspectral datasets, each presenting unique challenges that will highlight the strengths and weaknesses of the distance measures in different scenarios.

Indian Pines (IP) Dataset: Captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, IP contains 224 spectral bands and covers agricultural land with 16 classes. This dataset is characterized by subtle spectral differences between classes, particularly among crop types, making it challenging to distinguish classes based on their spectral similarity.

Salinas Valley (SV) Dataset: This dataset was also captured by a 224-band AVIRIS sensor over Salinas Valley, California, with a high spatial resolution (3.7-meter pixels). The area covered comprises 512×217 pixels and 20 water absorption bands were discarded ($[108-112], [154-167], 224$). It includes vegetables, bare soils, and vineyard fields classified into 16 classes.

University of Pavia (UP) Dataset: Acquired using the Reflective Optics System Imaging Spectrometer (ROSIS-3) sensor, the UP is composed of 610×340 pixels with 103 spectral bands. The image is categorized into 9 classes, consisting of $42,776$ labeled samples, representing various land cover types. This dataset is useful for testing how well distance metrics handle data where clear separations between classes are generally more evident, particularly in urban environments.

IV. Experiments and Results

In this section, we investigate the similarity score indices induced from the various distance metrics for distinguishing between classes in HSI datasets. Our analysis focuses on determining which score index best captures class similarity and dissimilarity among classes, considering the diverse characteristics of datasets. Specifically, we focus on inter-class spectral separability by analyzing representative profiles based on the mean spectral signatures of each class.

A. Experimental Setup

For each class within each dataset, we first derive a mean spectrum by averaging the spectral signatures across all pixels in that class. This mean spectrum serves as a simplified but representative profile for each class, allowing us to examine spectral similarity at the class level. Next, we make use of the similarity score indices in Section III-A to compute similarity scores between the class mean spectra of all classes within each dataset. These scores provide a comprehensive view of how well each index separates different classes. After that, we analyze the inter-class similarity scores to assess the general effectiveness of different indices across all datasets. Our goal is to identify one index that performs best according to the dataset-specific characteristics.

B. Evaluation Criteria

To evaluate the effectiveness of each similarity score index, we employ the following criteria:

Mean and Variance of Similarity Scores: The average similarity score across all inter-class pairs indicates an index's overall effectiveness, with lower means reflecting better class distinction. High variance suggests sensitivity to subtle spectral differences, useful for datasets with overlapping classes, while low variance may indicate limited adaptability to nuanced distinctions in closely related classes.

Boxplot Visualization: Boxplots visually summarize the performance of each index by displaying key statistics such as the interquartile range (IQR), quartiles (Q1, Q2, Q3), and outliers. These plots complement the mean and variance of similarity scores which will be given in Table I, by providing a clear visual distinction of score distributions, making it easier to identify indices with consistently low medians (Q2) or wider variability (longer IQR). This approach highlights comparative trends across datasets, aiding in the interpretation of the distinguishing capability of each index.

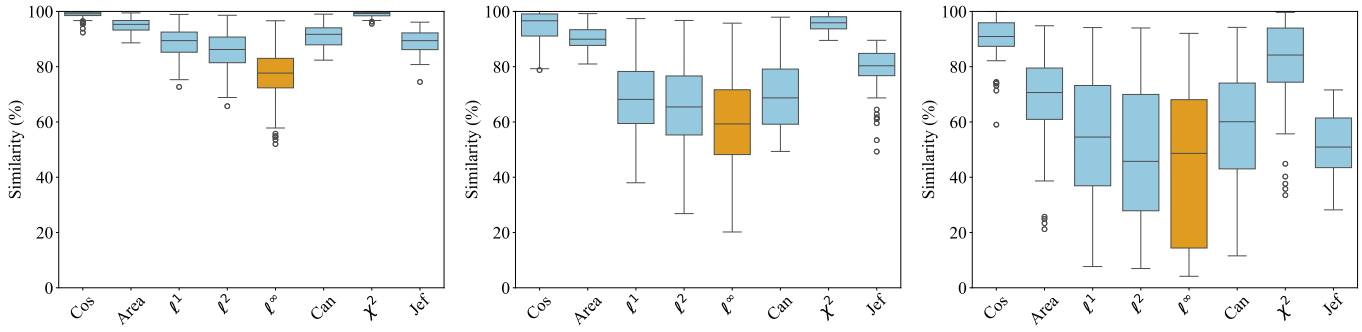

Fig. 2: Boxplots of similarity scores for IP, SV, and UP datasets across different similarity score indices.

C. Results

As shown in Fig. 1, the distributions of classwise mean spectra are notably similar between the IP and SV datasets, whereas the UP dataset exhibits a distinctly different distribution. Based on this, it can be hypothesized that the index most suitable for the IP and SV is likely to coincide, while a different index may perform better for distinguishing classes in the UP dataset.

However, the results presented in Table I reveal a different trend. Class-averaged similarity scores show that the index defined in (9) with $p = \infty$ (Chebyshev index) yields the lowest similarity scores across all three datasets. On the other hand, the index derived from cosine similarity uniformly achieves the highest or second-highest average similarity scores, regardless of the dataset, reflecting its limited ability to distinguish between different classes. These observations can be further clarified and supported by the boxplots Fig. 2, which visualize the distribution of similarity scores across datasets and indices. The Chebyshev index shows the lowest Q2 scores across IP and SV datasets, indicating its superior inter-class separability. This result aligns with the result in Table I, where the Chebyshev index achieved the lowest average similarity scores. Having a broader IQR than any other indices, the Chebyshev index's variability shows its adaptability to diverse spectral characteristics, making it particularly effective in datasets with subtle spectral distinctions, such as IP and SV. In contrast, indices with narrower IQRs may struggle with datasets that have more subtle spectral variations.

The boxplots in the left and middle of Fig. 2 show that the similarity score distributions for IP and SV are closely aligned, reinforcing the earlier observation of similar mean spectra. In contrast, the UP dataset exhibits broader distributions, reflecting its higher inter-class variability. Especially, the Chebyshev index again demonstrates superior performance due to its broader IQR despite similar Q2 values to the ℓ^2 based index for the UP dataset. In the left panel of Fig. 2, there are some outliers of the Chebyshev index which do not harm our assertion since they are close to the Q3.

In conclusion, the Chebyshev index is the best one that captures subtle spectral differences between classes, making it more effective in datasets with intricate class separability and further justifying its position as the most reliable index for hyperspectral analysis.

TABLE I: Average similarity scores across metrics and datasets

Metric	IP	SV	UP
Cosine	98.75 ± 1.38	94.05 ± 6.17	89.31 ± 9.62
Area	95.18 ± 2.42	90.54 ± 3.70	66.58 ± 19.09
ℓ^1	89.00 ± 5.70	68.33 ± 13.70	50.86 ± 24.22
ℓ^2	86.00 ± 7.19	65.19 ± 15.38	48.30 ± 24.94
ℓ^∞	**78.08** ± 9.84	**59.56** ± 16.95	**44.06** ± 26.91
Canberra	91.43 ± 4.29	69.72 ± 12.43	55.80 ± 21.74
χ^2	98.99 ± 0.88	95.83 ± 2.58	78.43 ± 19.08
Jeffrey	89.04 ± 4.06	79.43 ± 7.56	51.19 ± 12.39

V. Discussion

For certain pairs of classes used to measure similarity scores, the Chebyshev index does not always yield the lowest score. For instance, the Jeffrey similarity index (12) produces a lower score than the Chebyshev index for the IP dataset when comparing class 11 with classes 2, 3, 10, and 11. To further validate our assertions in Section IV, we analyze the rank of the Chebyshev index and compare differences in similarity scores using the following two inequalities:

- Optimal Admissibility (O/A):

$$\begin{aligned}&|\mathcal{S}_1(S^i, S^j) - \mathcal{S}_{\ell^\infty}(S^i, S^j)| \\ &\qquad < |\mathcal{S}_{\ell^\infty}(S^i, S^j) - \mathcal{S}_3(S^i, S^j)|,\end{aligned} \tag{13}$$

- Sub-optimal Admissibility (So/A):

$$\begin{aligned}&|\mathcal{S}_2(S^i, S^j) - \mathcal{S}_{\ell^\infty}(S^i, S^j)| \\ &\qquad < |\mathcal{S}_{\mathrm{Q2}}(S^i, S^j) - \mathcal{S}_{\ell^\infty}(S^i, S^j)|,\end{aligned} \tag{14}$$

where $\mathcal{S}_t(S^i, S^j)$ represents the score of the t-th lowest similarity index computed with the mean spectra of classes i and j, and $\mathcal{S}_{\ell^\infty}$ denotes the Chebyshev index score. Also, $\mathcal{S}_{\mathrm{Q2}}(S^i, S^j)$ is computed by the average of $\mathcal{S}_4$ and $\mathcal{S}_5$ since there are eight indices.

In (13), O/A indicates that the Chebyshev index achieves the second-lowest score for the class pair (i, j) with a smaller difference from the lowest index than from the third-lowest index. This supports the distinguishing capability of the Chebyshev index as near optimal. Additionally using (14),

So/A signifies that the Chebyshev index either achieves the second-lowest score for the class pair (i, j) or it has a smaller difference from the second-lowest index than from the $\mathcal{S}_{Q2}$ when it does not score at least the second-lowest. This justifies the distinguishing capability of the Chebyshev index as sub-optimal.

Using these definitions, the Chebyshev index is classified as at least O/A for $93.33\%(112/120)$ of cases, including pairs where the index achieves the lowest score. Furthermore, all the 120 cases are classified as at least So/A in the IP dataset. Similarly, for the SV and UP datasets, the Chebyshev index is at least O/A for $81.67\%(98/120)$ and $58.33\%(21/36)$ of cases, and at least So/A for $93.33\%(112/120)$ and $77.78\%(28/36)$ of cases, respectively.

VI. Conclusion

This study addressed the critical challenge of evaluating distance metrics in hyperspectral image (HSI) analysis, where high dimensionality and complex spectral properties hinder effective metric selection. By introducing novel similarity score indices that normalize metrics to a unified scale, we enabled consistent and interpretable comparisons across datasets, improving inter-class separability and intra-class consistency. Our experiments on benchmark HSI datasets demonstrated the robustness of the proposed framework, with the Chebyshev-based index consistently achieving superior performance. This approach offers a scalable solution that balances computational efficiency with analytical precision, making it suitable for large-scale AI applications. Future work will explore integrating these indices into advanced machine learning models for tasks such as anomaly detection and domain adaptation, while also addressing robustness under noisy or incomplete data. This research sets the stage for more reliable and interpretable metric selection in hyperspectral image analysis.

References

[1] C. Chang, "Spectral information divergence for hyperspectral image analysis," IEEE Int. Geosci. Remote Sens. Symp., vol. 1, pp. 509-511, 1999.

[2] C. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," IEEE Trans. Inf. Theory, vol. 46, no. 5, pp. 1927-1932, Aug. 2000.

[3] R. Cui, H. Yu, T. Xu, X. Xing, X. Cao, K. Yan, and J. Chen, "Deep learning in medical hyperspectral images: A review," Sensors, vol. 22, no. 24, Dec. 2022.

[4] H. Deborah, N. Richard, and J. Y. Hardeberg, "A comprehensive evaluation of spectral distance functions and metrics for hyperspectral image processing," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 8, no. 6, pp. 3224-3234, Jun. 2015.

[5] L. Drumetz, M. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," IEEE Trans. Image Process., vol. 25, no. 8, pp. 3890-3905, Aug. 2016.

[6] Y. Du, C. Chang, H. Ren, C. Chang, J. O. Jensen, and F. M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," Opt. Eng., vol. 43, no. 8, pp. 1777-1786, Aug. 2004.

[7] J. C. Gower, "Properties of Euclidean and non-Euclidean distance matrices," Linear Algebra Appl., vol. 67, pp. 81-97, Jun. 1985.

[8] H. Jeffreys, "An invariant form for the prior probability in estimation problems," Proc. R. Soc. A., vol. 186, pp. 453-461, Sep. 1946.

[9] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 7, pp. 1552-1565, Jul. 2004.

[10] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data," Remote Sens. Environ., vol. 44, no. 2, pp. 145-163, 1993.

[11] C. Kumar, S. Chatterjee, T. Oommen, and A. Guha, "New effective spectral matching measures for hyperspectral data analysis," Int. J. Remote Sens., vol. 42, no. 11, pp. 4126-4156, Mar. 2021.

[12] M. N. Kumar, M. V. R. Seshasai, K. S. V. Prasad, V. Kamala, K. V. Ramana, R. S. Dwivedi, and P. S. Roy, "A new hybrid spectral similarity measure for discrimination among Vigna species,", Int. J. Remote Sens., vol. 32, no. 14, pp. 4041-4053, 2011.

[13] G. N. Lance and W. T. Williams, "Computer programs for hierarchical polythetic classification("Similarity analyses")," Comput. J., vol. 9, pp. 60-64, May. 1966.

[14] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," Remote Sens., vol. 12, no. 16, Aug. 2020.

[15] F. D. van der Meer, and H. M. A. van der Werff, F. J. A. van Ruitenbeek, and C. A. Hecker, W. H. Bakker, M. F. Noomen, M. van der Meijde, E. J. M. Carranza, J. B. de Smeth, and T. Woldai, "Multi- and hyperspectral geologic remote sensing: A review," Int. J. Appl. Earth Obs. Geoinf., vol. 14, no. 1, pp. 112-128, 2012.

[16] R. R. Nidamanuri and B. Zbell, "Use of field reflectance data for crop mapping using airborne hyperspectral image," ISPRS J. Photogramm. Remote Sens., vol. 66, no. 5, pp. 683-691, Sep. 2011.

[17] S. Padma and S. Sanjeevi, "Jeffries Matusita based mixed-measure for improved spectral matching in hyperspectral image analysis," Int. J. Appl. Earth Obs. Geoinf., vol. 32, pp. 138-151, Oct. 2014.

[18] C. Revel, Y. Deville, V. Achard, X. Briottet, and C. Weber, "Inertia-constrained pixel-by-pixel nonnegative matrix factorisation: A hyperspectral unmixing method dealing with intra-class variability," Remote Sens., vol. 10, no. 11, Oct. 2018.

[19] D. Saha and A. Manickavasagan, "Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review," Curr. res. nutr. food sci., vol. 4, pp. 28-44, 2021.

[20] M. B. Stuart, A. J. S. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems," Sensors, vol. 19, no. 14, Jul. 2019.

[21] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," IEEE Geosci. Remote Sens. Mag., vol. 10, no. 1, pp. 64-90, Mar. 2022.

[22] W. Sun and Q. Du, "Hyperspectral band selection: A review," IEEE Geosci. Remote Sens. Mag., vol. 7, no. 2, pp. 118-139, Jun. 2019.

[23] P. Uddin and A. Mamun, and A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," IETE Tech. Rev., vol. 38, no. 4, pp. 377-396, Mar. 2021.

[24] B. Yang, H. Li, and Z. Guo, "Learning a deep similarity network for hyperspectral image classification," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 14, pp. 1482-1496, Nov. 2020.

[25] A. Zehtabian and H. Ghassemian, "Automatic object-based hyperspectral image classification using complex diffusions and a new distance metric," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 9, pp. 4106-4114, Jul. 2016.

[26] S. Zhang, Z. Chen, D. Wang, and Z. J. Wang, "Cross-domain few-shot contrastive learning for hyperspectral images classification," IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1-5, Dec. 2022.