Review of Computer Engineering Research
2026 Vol. 13, No. 3, pp. 101-122
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v13i3.5125

Leveraging spectral information for enhanced multispectral image retrieval using deep neural network

Monali P. Mahajan[1+]
Snehal M. Kamalapur[2]

[1,2]*Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik, Savitribai Phule Pune University, Pune, Maharashtra, India.*
[1]*Email: mjmetkar@kkwagh.edu.in*
[2]*Email: smkamalapur@kkwagh.edu.in*


(+ Corresponding author)

Article History
Received: 19 January 2026
Revised: 21 April 2026
Accepted: 8 May 2026
Published: 25 August 2026

Keywords
Content-based image retrieval
Deep neural network
DenseNet
Multispectral image retrieval
Remote sensing
Similarity measures
Spectral–spatial features.

ABSTRACT

Similar multispectral images for a given query image can be retrieved from large databases of multispectral images. These images provide more detail due to multiple bands, but the high dimensionality creates a challenging, inefficient environment for retrieving similar images. This research introduces a method for multispectral image retrieval based on the DenseNet deep learning framework. DenseNet's use of dense connections between network layers enables the reuse of previously learned features and ensures that network layers support better representation of both spectral and spatial attributes of images, reducing the likelihood of important attributes being lost during data transfer through the DenseNet architecture. Deep features will be compared using two methods of measuring similarity: Euclidean distance and Cosine distance. The research finds that the DenseNet multispectral image retrieval approach provides a Mean Average Precision (MAP) score of 0.73 for the top 100 retrieved images based on Euclidean distance, which exceeds the MAP@100 score of 0.68 obtained from prior retrieval methods on the same dataset. When DenseNet is compared to state-of-the-art hash-based image retrieval methods, it consistently provides the highest MAP score for all tested code lengths, demonstrating that the DenseNet approach preserves the most critical spectral and spatial information. Overall, these results demonstrate that the DenseNet multispectral image retrieval framework is highly effective and efficient, rendering it an applicable solution for managing large multispectral image collections for actual remote sensing applications such as land use studies, environmental monitoring, or geo-spatial data management.

Contribution/Originality: This study contributes to existing research by improving multispectral image retrieval through inclusive exploitation of all spectral bands in the EuroSAT dataset within a DenseNet-based framework. It employs a novel assessment methodology by systematically comparing Euclidean and Cosine similarity measures in a multispectral retrieval context. This research is among the few that investigate full-band feature learning and benchmark against state-of-the-art hashing-based methods to demonstrate retrieval superiority.

1. INTRODUCTION

Spectral imaging refers to digital imaging techniques that acquire more detailed spectral (color) information for each pixel compared to traditional color cameras. Mahajan and Kamalapur [1] encompass the gaining of multispectral images through various bands of the electromagnetic spectrum, as shown in Figure 1 [2], including visible and non-visible wavelengths. This technique allows the capture of multiple spectral bands per pixel, presenting beyond conventional RGB (Red, Green, and Blue) imagery. It encompasses a diverse array of techniques utilizing infrared, visible, ultraviolet, and X-ray bands [1]. Depending on the number of wavelength bands captured, these images are categorized into different spectral image types. They are typically acquired using specialized ground-based or airborne platforms equipped with active or passive sensors.

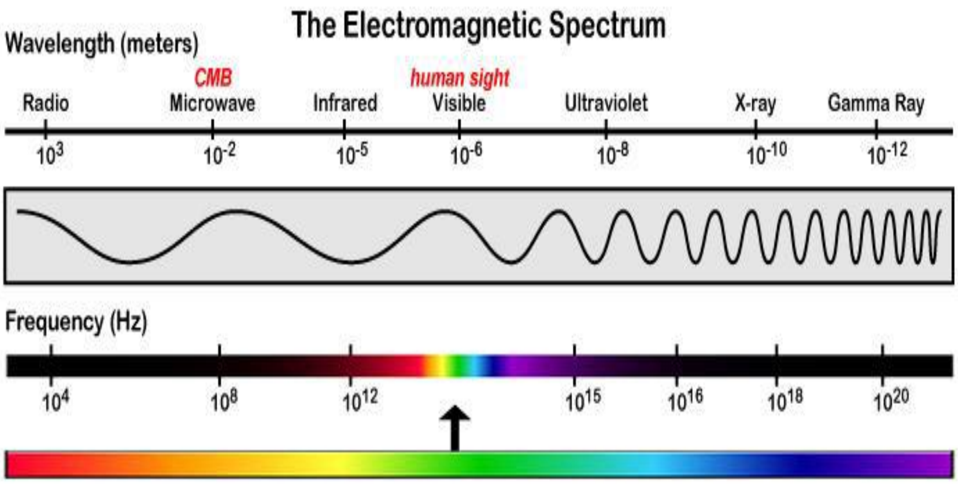

Figure 1. Electromagnetic spectrum.

Source: Kot et al. [2].

Among different spectral images, multispectral images typically include up to 100 spectral bands, as shown in Figure 2. Enormous applications of multispectral images include allowing early detection in melanoma [3] detecting oil-containing dressing on salad leaves [4] identifying oral cancer [5] facilitating facial recognition [6] discerning coal and gangue [7] revealing ancient hieroglyphic text in an Egyptian stele [8] target tracking [9] conducting water stress analysis of an avocado crop [10] phenotyping and quality monitoring [11] anatomy pathology [12] etc.

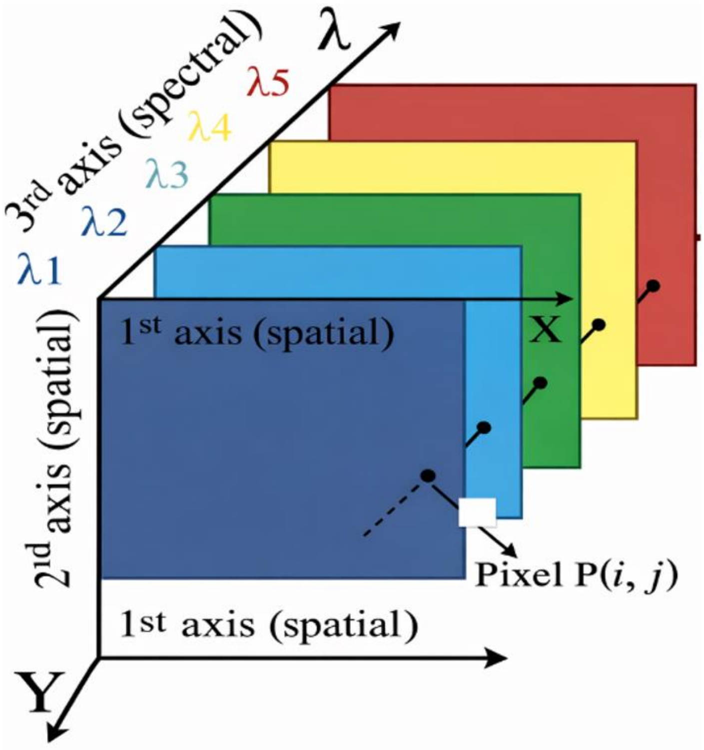

Figure 2. Multispectral image data cube.

Unlike conventional RGB images, multispectral images provide rich spectral information that allows detailed scene analysis. Spectral resolution describes how precisely a sensor can distinguish between different wavelength intervals within the electromagnetic spectrum [2]. It varies depending on the sensor used for image acquisition. For example, Band 1 of the Landsat TM sensor detects electromagnetic energy in the visible spectrum ranging from 0.45 to 0.52 micrometers, whereas Band 3 captures energy between 0.63 and 0.69 micrometers. The spectral information provided by these bands enables a comprehensive examination of objects across wavelengths, as depicted in Figure 3 [13], allowing enhanced description compared to RGB imagery.

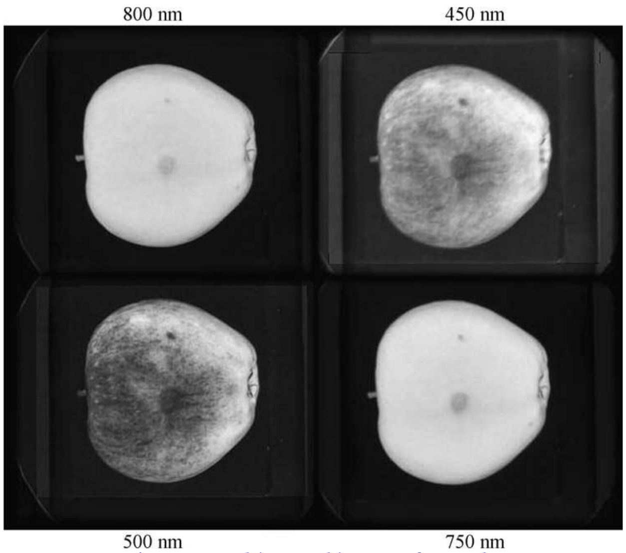

Figure 3. Multispectral images of an apple.

Source: Kleynen et al. [13].

Large numbers of multispectral images can be produced from applications such as earth observation, precision agriculture, military and defense, medical imaging, archaeology, and urban planning. To find matching multi-channel images, the process of multispectral image retrieval [14] requires searching through a vast database to find images that fit a given query image. A typical multispectral imaging system can be seen in the example of multispectral image retrieval systems, as shown in Figure 4.

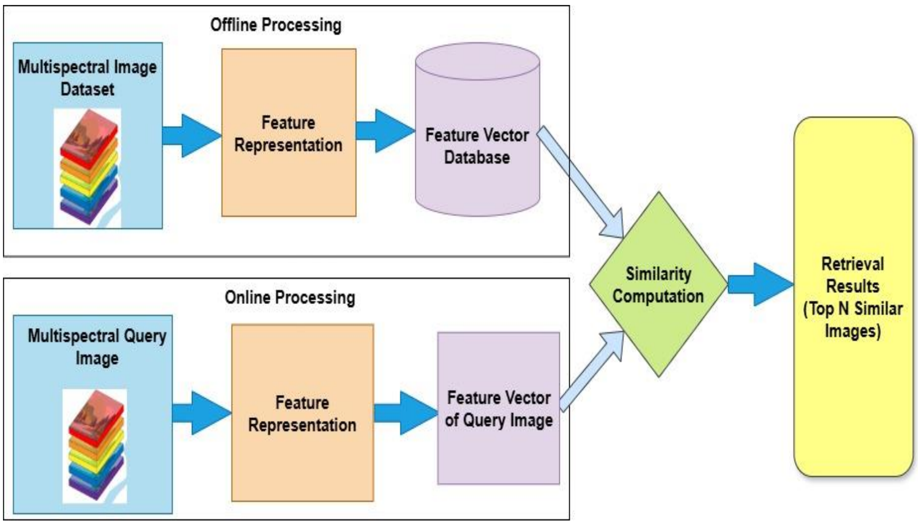

Figure 4. Multispectral image retrieval process.

As shown in Figure 4, the process begins with an archive containing stored multispectral images captured using dedicated sensors. The subsequent step involves extracting and representing image features using various techniques.

These feature representations form the basis for similarity comparison. The query image undergoes similar feature extraction, and its feature vector is compared with those of database images using different distance metrics to measure similarity. Based on similarity scores, the most relevant images are retrieved [14].

Numerous approaches have been proposed for feature representation of multispectral images, including texture-based feature representation [15], scale-invariant and Gabor descriptors [16], spectral features [17], generative adversarial networks (GAN) [18], Multiview Graph Convolutional Hashing (MGCH) [19], and Fast Recurrent Convolutional Neural Networks (FRCNN) [20]. After feature extraction, similarity computation is performed using distance measures such as Euclidean distance [15, 17], Manhattan distance [16], Chi-square distance, etc.

The intricate and high-dimensional nature of spectral data makes retrieving similar images from multispectral datasets particularly challenging. Therefore, a system that considers all spectral bands in multispectral images and effectively retrieves images matching a query image is necessary.

The key contributions of this article are summarized as follows:

1. As multispectral images contain rich spectral information, the proposed system considers all spectral bands for feature extraction, improving retrieval performance.
2. A deep neural network is applied for feature extraction to obtain comprehensive multispectral representations.
3. The proposed system employs Euclidean and Cosine distances as similarity metrics for measuring image similarity.
4. A comparative analysis of retrieval performance using Euclidean and Cosine distance metrics is performed.
5. To achieve state-of-the-art retrieval performance, image features and their latent relationships are learned and optimized jointly.
6. Experimentation is conducted on the EuroSAT dataset for multispectral image retrieval to validate the proposed deep neural network method. Precision, Recall, and Mean Average Precision (MAP) are used as performance evaluation metrics.

Section 2 of the paper focuses on related works for multispectral image retrieval. The proposed work on multispectral image retrieval is presented in Section 3. Section 4 discusses results and analysis, followed by a comparison with state-of-the-art methods in Section 5. The conclusion is summarized in Section 6.

2. RELATED WORK

A variety of methods have been proposed for retrieving similar images for a given query image in multispectral images by considering spectral information. Features of multispectral images can be represented using various techniques, and similarity between images is typically measured through different distance metrics. This segment focuses on the study of image retrieval methods for multispectral images, which remain challenging due to the complexity of spectral information in multispectral datasets.

2.1. Texture-Based Feature Representation Methods

Early research in multispectral image retrieval primarily relied on handcrafted texture descriptors to represent spectral–spatial characteristics. Traditional texture representations such as Fourier power spectrum, spatial co-occurrence matrices, wavelets, and Gabor filters were widely explored for remote sensing retrieval tasks. Newsam and Kamath [15] explored different ways to represent textures in multispectral satellite images from the IKONOS dataset. They found that using all available spectral bands leads to better image retrieval accuracy. Co-occurrence, wavelet, and Gabor feature methods all did equally as well, according to some studies. For example, in another study by Joshi and Mukherjee [16], SIFT and Gabor descriptors were used by the authors, along with their suggestion of combining these descriptors for improved retrieval results. However, all these base methods, although effective for obtaining basic images, still cannot bridge the gap between raw pixel values and what the picture truly shows to the user.

2.2. Semantic and Spectral Feature Learning Approaches

Researchers utilized a method involving creating a multi-spectral image space, using spectral features to compute similarity among these images and keywords related to users' queries. Wijitdechakul et al. [17] demonstrated that creating this multispectral semantic image improved the accuracy of semantic relatedness for image retrieval based on semantics, as they included more than just raw spectral values when calculating the degree of semantic similarity between various images and spectra.

2.3. Deep Learning and GAN-Based Methods

Deep learning approaches have significantly advanced multispectral retrieval by learning hierarchical spectral–spatial representations. Chen and Lu [18] proposed an unsupervised method for multispectral retrieval using hashing and Generative Adversarial Networks (GANs), introducing a reconstruction loss to better utilize latent codes. Sathiyaprasad and Kumar [20] developed a Fast Recurrent Convolutional Neural Network (FRCNN) architecture for satellite image feature extraction and object localization. Although these models improved representation learning, they required substantial computational resources and large training datasets.

2.4. Hashing-Based Retrieval Methods

To enhance the retrieval efficiency of large-scale databases, hashing methods were created. Multiview Graph Convolutional Hashing (MGCH) is the method developed by Gao et al. [19] that combines multisource image fusion with Graph Convolutional Networks (GCNs) to enhance the capture of structural similarities found in multiview data. Some traditional unsupervised hashing methods [21-24] mapped high-dimensional features to compact hash spaces, and they also retain the underlying similarity structure when achieving this ascertainment. Various deep hashing methods [25, 26] produce binary codes in an end-to-end fashion directly from images, often utilizing rotation invariance or pre-trained CNNs. While hashing methods are efficient, they generally compromise on either discriminative detail or compact representation.

2.5. Foundation Models and Transformer-Based Approaches

The development of recent technologies is directing attention toward foundation models, self-supervised learning methods, and transformer architectures for multispectral image retrieval.

Blumenstiel et al. [27] presented a framework for multispectral image retrieval based on a geospatial foundation model. This approach relies on pre-trained large foundation models to encode multispectral satellite images without requiring extensive fine-tuning. They demonstrated competitive performance in terms of mean average precision (MAP) with a small embedding size, making them well-suited for large-scale retrieval. Choudhury et al. [28] proposed an architectural framework based on joint-embedding predictive modelling that uses feature space instead of pixel space. This self-supervised architecture uses variance-invariance-covariance regularization as a way to obtain more discriminative representations, with lower overhead costs and improved performance on multispectral datasets than traditional methods. Marimo et al. [29] developed Llama3-MS-CLIP, a vision-language foundation model using contrastive learning to learn from very large multispectral datasets. By extending the spectral range beyond that of RGB data and including supervised learning from language inputs, the authors demonstrated significant improvements in zero-shot retrieval of multispectral images. However, transformer and foundation model-based approaches require high hardware and memory resources, which may limit their use in resource-constrained environments. In conclusion, there has been a shift from the use of traditional texture-based methods such as co-occurrence, wavelet, and Gabor filter-based methods for retrieving multispectral images to more advanced methods being developed to reduce the semantic gap. Research has demonstrated that the performance of multispectral retrieval is improved by combining several spectral bands, and current work has focused on developing more advanced methodologies such as high-level feature fusion, semantic computing, unsupervised deep learning techniques with GANs, and hashing techniques to develop more intelligent methods of searching and retrieving multispectral images. Neural networks such as FRCNN have also been employed, providing evidence of the shift toward more intelligent retrieval methods. While these methods have provided improved retrieval capabilities, they do require significant computational power and memory resources.

3. PROPOSED WORK

3.1. Experimental Setup

The EuroSAT dataset contains 27,000 images with a pixel size of 64x64, utilizing all 13 bands for experimentation. All images are preprocessed through normalization, resizing, and Gabor filtering. Features are extracted using DenseNet, built and tested with the Keras open-source framework. The platform includes an Intel Core i9-11900K, 16GB RAM (x2), 500GB hard disk space, and an NVIDIA GeForce RTX 3070 graphics card for experimentation.

3.2. Dataset

The experimentation is conducted on the EuroSAT [27] multispectral image dataset, which comprises 13 spectral bands in total. This dataset is structured into 10 different classes, each including between 2,000 and 3,000 images. The classes include Annual Crop, Forest, Herbaceous Vegetation, Highway, Industrial, Pasture, Permanent, Residential, Rivers, and Sea Lake, each captured at various wavelengths as outlined in Table 1. The entire EuroSAT dataset contains 27,000 images, with two types of images: one database with 27,000 RGB images and another with 27,000 multispectral images.

These images serve as valuable resources for exploring and analyzing the impact of spectral information on image classification, retrieval, and related tasks. They offer a diverse representation of terrestrial environments and land cover types, making them suitable for extensive remote sensing applications and research endeavors.

Table 1. Spectral bands of EuroSAT [27].

Sr. No.	Band	Band Name	Wavelength (µm)	Typical Use
1	B01	Coastal Aerosol	0.433 –0.453	Atmospheric correction, coastal monitoring
2	B02	Blue	0.458 – 0.523	True color (blue), water analysis
3	B03	Green	0.543 – 0.578	True color (green), vegetation monitoring
4	B04	Red	0.650 – 0.680	True color (red), vegetation, and soil detection
5	B05	Vegetation Red Edge 1	0.698 – 0.713	Vegetation health detection
6	B06	Vegetation Red Edge 2	0.733 – 0.748	Monitoring vegetation stress
7	B07	Vegetation Red Edge 3	0.773 – 0.793	Vegetation classification
8	B08	Near-Infrared (NIR)	0.785 – 0.899	Biomass, vegetation vigor
9	B08A	Narrow NIR (Red Edge 4)	0.855 – 0.875	Detailed vegetation analysis
10	B09	Water Vapor	0.935 – 0.955	Atmospheric correction (humidity)
11	B10	Shortwave Infrared – Cirrus	1.365 – 1.385	Cirrus cloud detection
12	B11	Shortwave Infrared 1	1.565 – 1.655	Soil, vegetation moisture content
13	B12	Shortwave Infrared 2	2.100 – 2.280	Snow/ice detection, soil study

3.3. Methodology

The proposed multispectral image retrieval technique consists of two main stages: offline feature extraction and online query retrieval. The system design is shown in Figure 5. During the offline stage, multispectral images are preprocessed using grayscale conversion and Gabor-based texture enhancement before passing through a DenseNet model for deep feature extraction. The resulting feature embeddings are stored in a feature database. During the online stage, a query image undergoes the same preprocessing and feature extraction pipeline. The extracted query embedding is then compared with stored database features using similarity measures, such as Euclidean distance and Cosine similarity, to retrieve the Top-N most similar images.

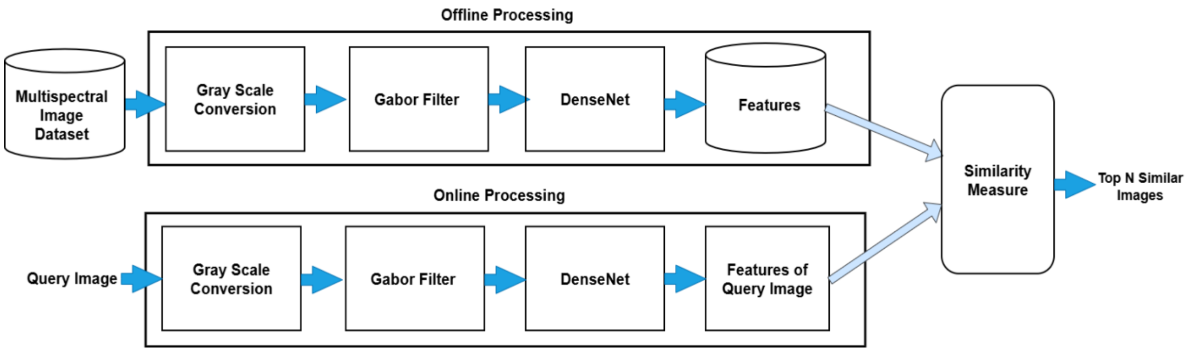

Figure 5. Proposed multispectral image retrieval technique.

3.3.1. Dataset and Data Splitting Strategy

The experiments were conducted on multispectral satellite images of size 64 × 64 pixels with 13 spectral bands. To ensure unbiased evaluation and eliminate data leakage, a strict 70%–30% train–test split was adopted.

- 70% of the dataset was used exclusively for model training.
- 30% of the dataset was reserved strictly for testing and retrieval evaluation.

No overlap between training and testing samples was allowed. All retrieval experiments and Mean Average Precision (MAP) calculations were performed exclusively on the unseen 30% test set.

3.3.2. Multispectral Preprocessing and Texture Enhancement

To enhance spatial texture characteristics prior to deep feature learning, a Gabor-based preprocessing strategy was applied. Since multispectral images contain rich spectral information, band processing and texture extraction were carefully designed. The 13 spectral bands were first combined to generate a grayscale representation, followed by normalization and multi-orientation Gabor filtering. The details are summarized in Table 2.

Table 2. Multispectral preprocessing and Gabor configuration.

Category	Parameter	Value / Setting	Explanation
Original Input	Image size	64 × 64 × 13	Multispectral satellite image with 13 spectral bands
Band Processing	Band combination	Summation across 13 bands	Produces a single grayscale representation for texture extraction
Normalization	Intensity range	0–255	Ensures uniform scale before filtering
Gabor Kernel	Kernel size	31 × 31	Captures spatial texture patterns
	Sigma (σ)	4.0	Controls the Gaussian envelope spread
	Lambda (λ)	10.0	Controls sinusoidal wavelength
	Gamma (γ)	0.5	Controls spatial anisotropy
	Psi (ψ)	0	Zero phase (even-symmetric filter)
Orientations	Theta (θ)	16 orientations (0 → π)	Captures directional textures
Gabor Output	Orientation responses	64 × 64 × 16	Each orientation produces one feature map
Fusion Method	Orientation fusion	Stacking	Orientation responses stacked as channels
Final Output	CNN input tensor	64 × 64 × 16	Texture-enhanced input to DenseNet
Purpose	Role	Texture enhancement before CNN	Improves land-cover discrimination

The stacked 16-channel tensor was used as input to the DenseNet architecture.

3.3.3. DenseNet-Based Feature Extraction

A modified DenseNet architecture was used to develop a discriminative representation for retrieving images using multiple spectra. This approach takes advantage of the dense connection between each layer and allows for efficient reuse of features, improved gradient propagation, and better learned representations. Each of the three dense blocks has been separated from one another by a transition layer, which has been followed by global average pooling and an embedding layer made up of fully connected methods. The detailed configuration of DenseNet is shown in Table 3.

Table 3. DenseNet architecture configuration.

Layer	Operation	Dense Connectivity	Output Size
Input	Gabor Feature Maps	—	(64, 64, 16)
Conv Layer	3×3 Conv (64 filters)	Standard convolution	(64, 64, 64)
Dense Block 1 (4 layers, growth=32)	BN → ReLU → 1×1 Conv (128) → BN → ReLU → 3×3 Conv (32)	Concatenated with previous maps	(64, 64, 192)
Transition Layer 1	BN → ReLU → 1×1 Conv (compression=0.5) → AvgPool(2×2)	Channel & spatial reduction	(32, 32, 96)
Dense Block 2 (6 layers, growth=32)	Same structure	Dense concatenation	(32, 32, 288)
Transition Layer 2	BN → ReLU → 1×1 Conv (compression=0.5) → AvgPool(2×2)	Downsampling	(16, 16, 144)
Dense Block 3 (8 layers, growth=32)	Same structure	Dense concatenation	(16, 16, 400)
BN + ReLU	Feature normalization	—	(16, 16, 400)
Global Average Pooling	Spatial aggregation	—	(400)
Dense Layer	Fully Connected (ReLU)	—	(512)
BatchNorm + Dropout	Regularization	—	(512)
Embedding Layer	Dense (128)	Feature embedding	(128)
L2 Normalization	Unit-length normalization	—	(128)

The final output is a 128-dimensional normalized embedding used for similarity-based retrieval.

3.3.4. Similarity Measurement

To evaluate retrieval effectiveness, two similarity measures were investigated.

- Euclidean Distance.
- Cosine Similarity.

Both metrics were applied to the 128-dimensional L2-normalized embeddings. Retrieval ranking was performed by computing similarity scores between the query embedding and database embeddings within the 30% test set.

3.3.5. Evaluation Protocol

Retrieval performance was evaluated using Mean Average Precision (MAP).

- All MAP values were computed exclusively on the unseen 30% test set.
- The 70% training set was used only for model learning.

This strictly disjoint protocol ensures unbiased performance estimation and eliminates data leakage.

3.4. Pseudocode

```
Input:
   D = Multispectral image dataset (64×64×13)
   N = Number of top similar images to retrieve

Output:
   Top-N retrieved images
   MAP score

--------------------------------------------------------
Step 1: Dataset Splitting
--------------------------------------------------------
1. Randomly split D into:
      Train_Set (70%)
      Test_Set  (30%)
2. Ensure no overlap between Train_Set and Test_Set

--------------------------------------------------------
Step 2: Offline Processing (Feature Database Creation)
--------------------------------------------------------
3. Initialize empty Feature_Database

4. For each image I in Train_Set:
     a. Convert multispectral image (13 bands)
        to grayscale by summation across bands
     b. Normalize intensity to range [0, 255]
     c. Apply Gabor filter with:
          - Kernel size = 31×31
          - 16 orientations (0 → π)
     d. Stack 16 orientation responses → Tensor (64×64×16)
     e. Input tensor to DenseNet
     f. Extract 128-dimensional embedding
     g. Apply L2 normalization
     h. Store embedding in Feature_Database

--------------------------------------------------------
Step 3: Online Processing (Query Retrieval)
--------------------------------------------------------
5. For each query image Q in Test_Set:
     a. Convert Q to grayscale (band summation)
     b. Normalize to [0, 255]
     c. Apply Gabor filter (same parameters)
     d. Stack 16 orientation responses
     e. Pass through DenseNet
     f. Extract 128-D embedding
     g. Apply L2 normalization

--------------------------------------------------------
Step 4: Similarity Computation
--------------------------------------------------------
6. For each embedding F in Feature_Database:
     a. Compute Euclidean_Distance(Q, F)
     b. Compute Cosine_Similarity(Q, F)

7. Rank database images based on similarity score
     - Ascending order for Euclidean
     - Descending order for Cosine

8. Retrieve the top-N most similar images
```

Step 5: Performance Evaluation
----
9. Compute Precision and Average Precision (AP)
10. Compute Mean Average Precision (MAP)
11. Report MAP for:
 - Euclidean similarity
 - Cosine similarity

3.5. Mathematical Model

3.5.1. Dataset Representation

Let the multispectral dataset be defined as shown in Equation 1.

$$\mathcal{D} = \{I_i\}_{i=1}^{M} \tag{1}$$

Where each image is represented using Equation 2.

$$I_i \in \mathbb{R}^{64\times64\times13} \tag{2}$$

The dataset is partitioned as shown in Equation 3.

$$\mathcal{D}_{train} \cup \mathcal{D}_{test}, |\ \mathcal{D}_{train}\ |= 0.7M, |\ \mathcal{D}_{test}\ |= 0.3M \tag{3}$$

With:

$$\mathcal{D}_{train} \cap \mathcal{D}_{test} = \emptyset$$

3.5.2. Grayscale Conversion

Each multispectral image consists of 13 spectral bands as shown in Equation 4.

$$I_i = \{B_1, B_2, \dots, B_{13}\} \tag{4}$$

The grayscale representation is obtained via band summation as shown in Equation 5.

$$G(x,y) = \sum_{k=1}^{13} B_k\,(x,y) \tag{5}$$

3.5.3. Normalization

Pixel intensities are normalized to [0,255] as shown in Equation 6:

$$G_{norm}(x,y) = 255 \cdot \frac{G(x,y)-G_{min}}{G_{max}-G_{min}} \tag{6}$$

3.5.4. Gabor Filtering

The 2D Gabor kernel is defined as shown in Equation 7.

$$g(x,y;\theta,\lambda,\sigma,\gamma,\psi) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) cos\left(\frac{2\pi x'}{\lambda}+\psi\right) \tag{7}$$

Where:

$$x' = x\cos\ \theta + y\sin\ \theta$$
$$y' = -x\sin\ \theta + y\cos\ \theta$$

For 16 orientations:

$$\theta_j = \frac{j\pi}{16}, j = 0,1, \dots ,15$$

Each orientation response:

$$R_j = G_{norm} * g(\theta_j)$$

Stacking all responses:

$$T(x,y) = \{R_0, R_1, \dots, R_{15}\}$$
$$T \in \mathbb{R}^{64\times64\times16}$$

3.5.5. DenseNet Feature Extraction

Let the DenseNet mapping function be as shown in Equation 8.

$$f_\theta: \mathbb{R}^{64\times64\times16} \to \mathbb{R}^{128} \tag{8}$$

The embedding vector is as shown in Equation 9.

$$z_i = f_\theta(T_i) \tag{9}$$

L2 Normalization:

The embedding is normalized to unit length using Equation 10.

$$\hat{z}_i = \frac{z_i}{\|z_i\|_2} = \frac{z_i}{\sqrt{\sum_{k=1}^{128} z_{ik}^2}} \tag{10}$$

3.5.6. Similarity Measures

Let:

$$\hat{z}_q$$

Be the query embedding, and

$$\hat{z}_d$$

Be a database embedding.

Euclidean Distance is shown in Equation 11.

$$d_E(\hat{z}_q, \hat{z}_d) = \| \hat{z}_q - \hat{z}_d \|_2 = \sqrt{\sum_{k=1}^{128} (\hat{z}_{qk} - \hat{z}_{dk})^2} \tag{11}$$

Images are ranked in ascending order.

Cosine Similarity is shown in Equation 12.

$$S_C(\hat{z}_q, \hat{z}_d) = \frac{\hat{z}_q \cdot \hat{z}_d}{\|\hat{z}_q\|_2 \|\hat{z}_d\|_2} \tag{12}$$

Since embeddings are L2-normalized as shown in Equation 13.

$$S_C(\hat{z}_q, \hat{z}_d) = \hat{z}_q^T \hat{z}_d \tag{13}$$

Images are ranked in descending order.

3.5.7. Retrieval Ranking

Let:

$$\mathcal{R}_q = \text{ranked list of database images}$$

Top-N retrieval is given as Equation 14.

$$\mathcal{R}_q^{(N)} = \{I_1, I_2, \ldots, I_N\} \tag{14}$$

3.5.8. Evaluation Metric (MAP)

For each query, Average Precision (AP) is calculated as Equation 15.

$$AP(q) = \frac{1}{R_q} \sum_{k=1}^{n} P(k) \cdot rel(k) \tag{15}$$

Where:

- R_q= Number of relevant images.
- $P(k)$= Precision at rank k.
- $rel(k) \in (0,1)$.

Mean Average Precision is calculated using Equation 16.

$$MAP = \frac{1}{|\mathcal{D}_{test}|} \sum_{q \in \mathcal{D}_{test}} AP(q) \tag{16}$$

All MAP values are computed exclusively on the 30% test set.

4. RESULTS AND DISCUSSION

For experimentation, images from each class are considered with all spectral bands. Figure 6 shows an example of representative spectral bands for the Annual Crop class. Figure 6 illustrates that each spectral band in a multispectral image captures distinct details, highlighting the importance of including all bands for effective image representation. Given this, retrieval experiments were conducted using all 13 spectral bands from the EuroSAT dataset.

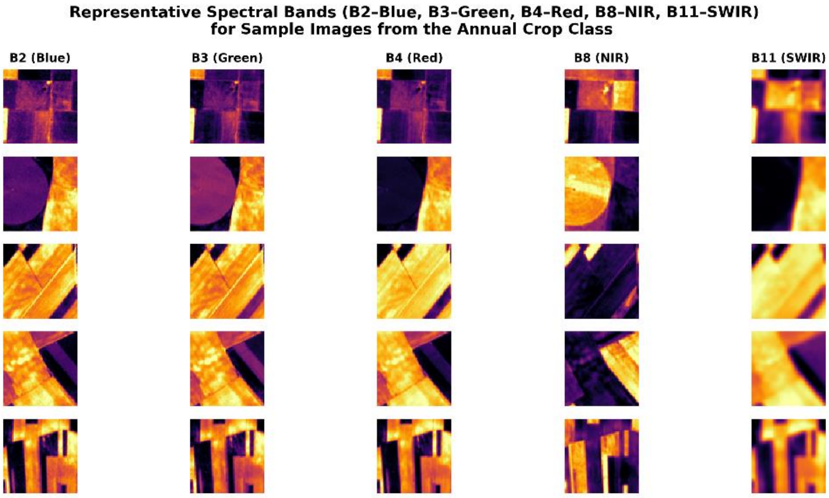

Figure 6. Representative bands of multispectral images.

For evaluation, Euclidean distance was first applied as the similarity measure. Retrieval performance was assessed at different cut-off points: Top 10, Top 20, Top 30, Top 50, and Top 100 retrieved images. A set of 50 randomly selected query images was used for comparative analysis. The selection of 50 query images was motivated by the need to ensure a fair and consistent comparison with existing state-of-the-art methods. For each query image, precision was computed at each cut-off, and Figures 7–11 present precision values for Top 10 through Top 100 retrievals, respectively.

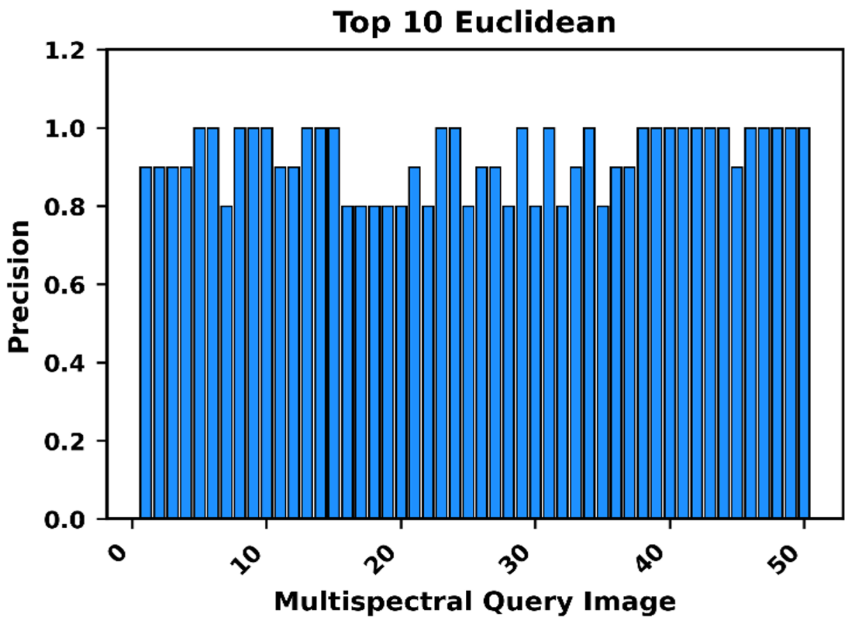

Figure 7. Precision values for the retrieval results of the Top 10 similar images using Euclidean Distance.

In Figure 8, for each of the query image Precision is calculated for the similar Top 20 images.

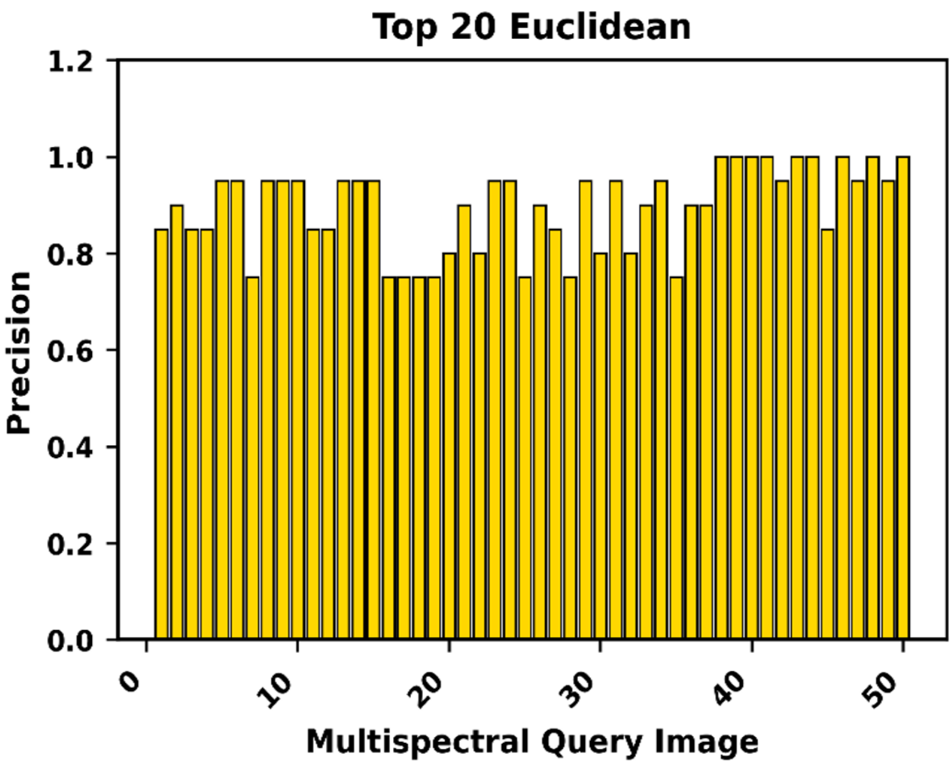

Figure 8. Precision values for retrieval results of the Top 20 similar images using Euclidean Distance.

In Figure 9, for each of the query image Precision is calculated for the similar Top 30 images.

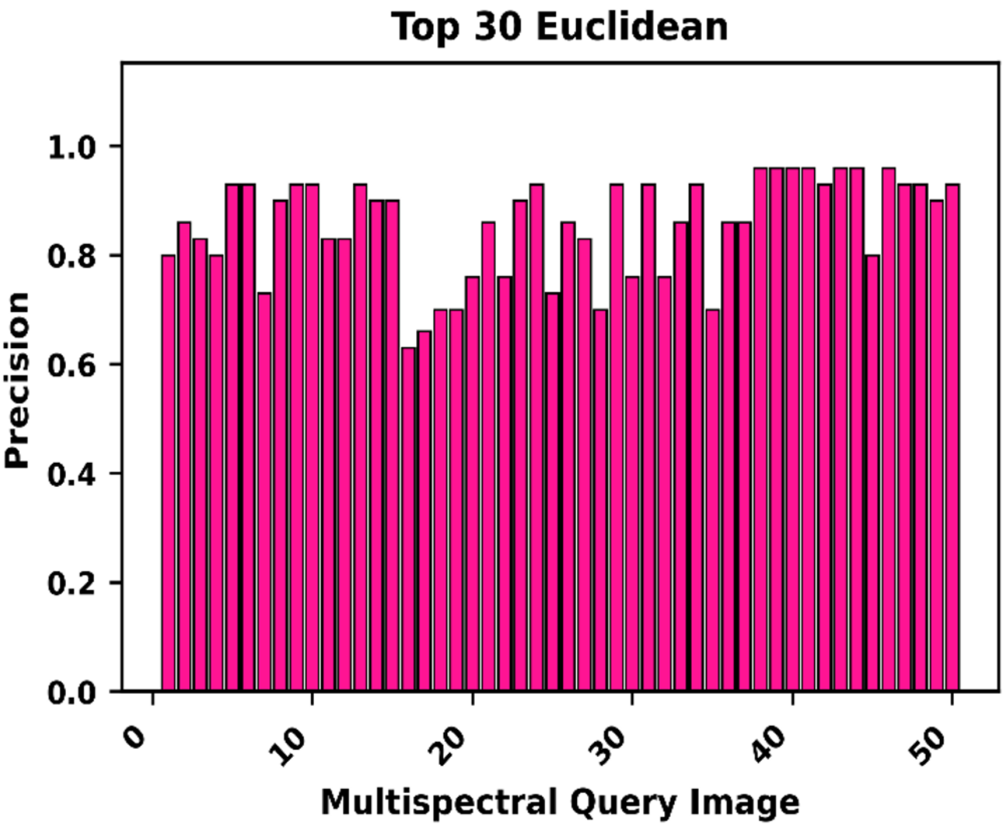

Figure 9. Precision values for the retrieval results of the Top 30 similar images using Euclidean Distance.

In Figure 10, for each of the query image Precision is calculated for similar Top 50 images.

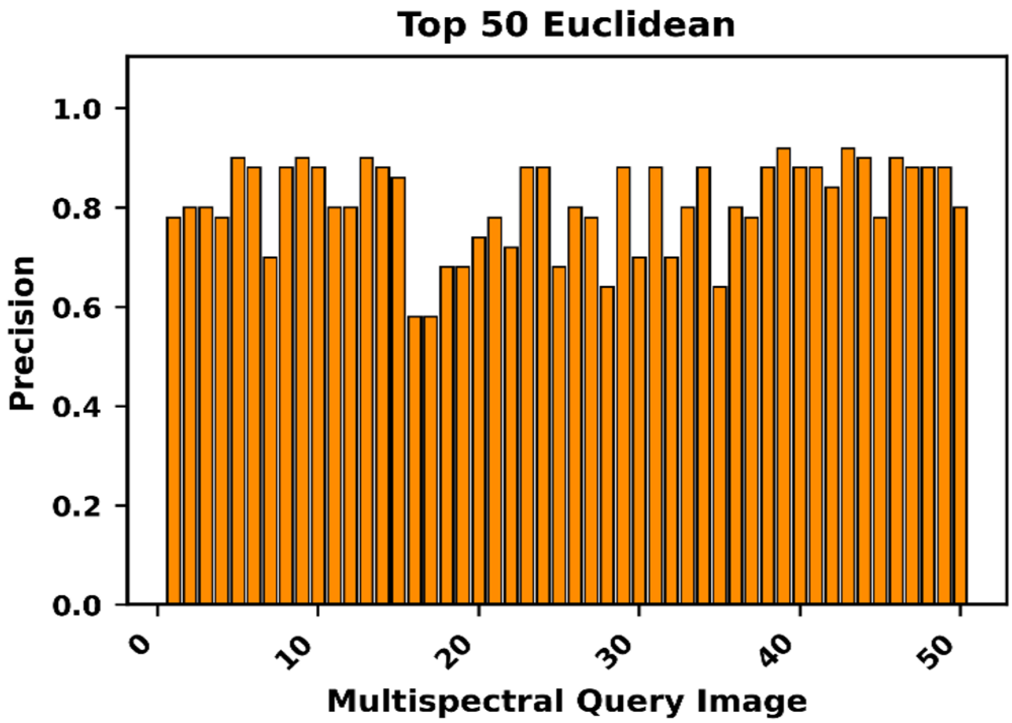

Figure 10. Precision values for retrieval results of the Top 50 similar images using Euclidean Distance.

In Figure 11, for each of the query image Precision is calculated for similar Top 100 images.

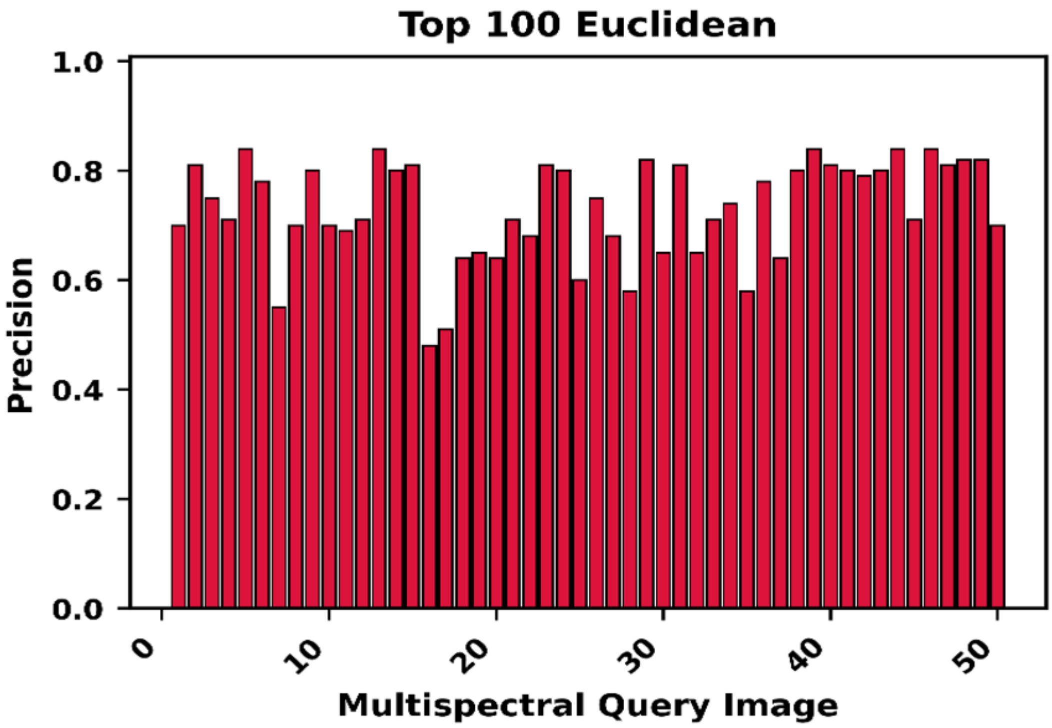

Figure 11. Precision values for the retrieval results of the Top 100 similar images using Euclidean Distance.

For evaluation, the cosine distance was also applied as the similarity measure. Figures 12–15 present precision values for Top 10 through Top 100 retrievals, respectively.

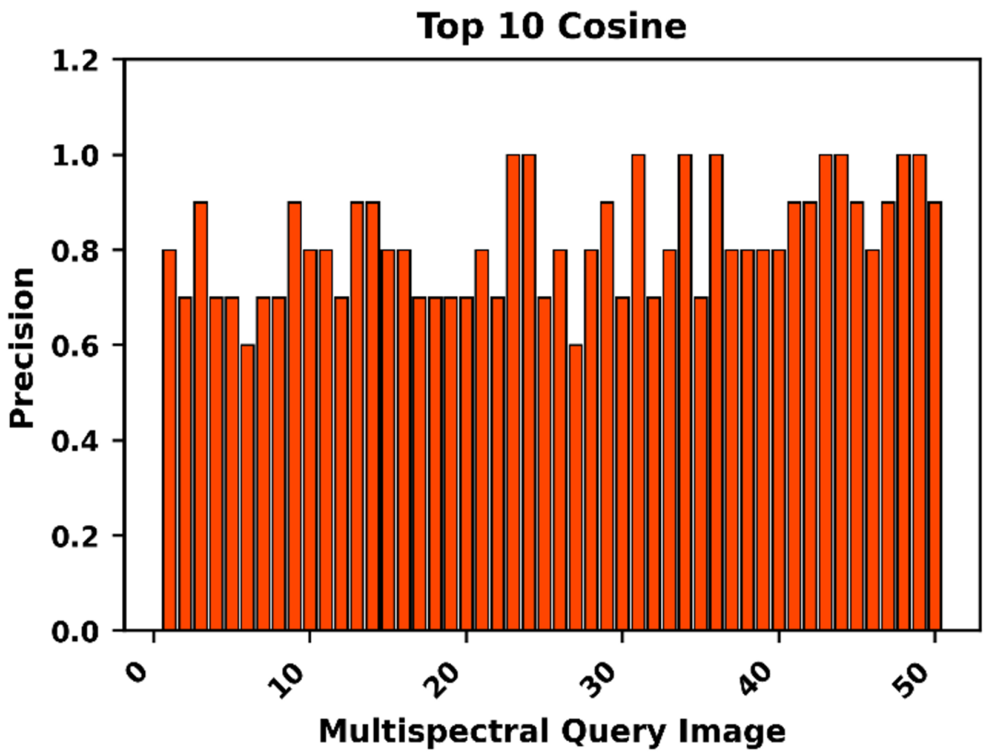

Figure 12. Precision values for retrieval results of the Top 10 similar images using Cosine Distance.

In Figure 13, for each of the query image Precision is calculated for the similar Top 20 images.

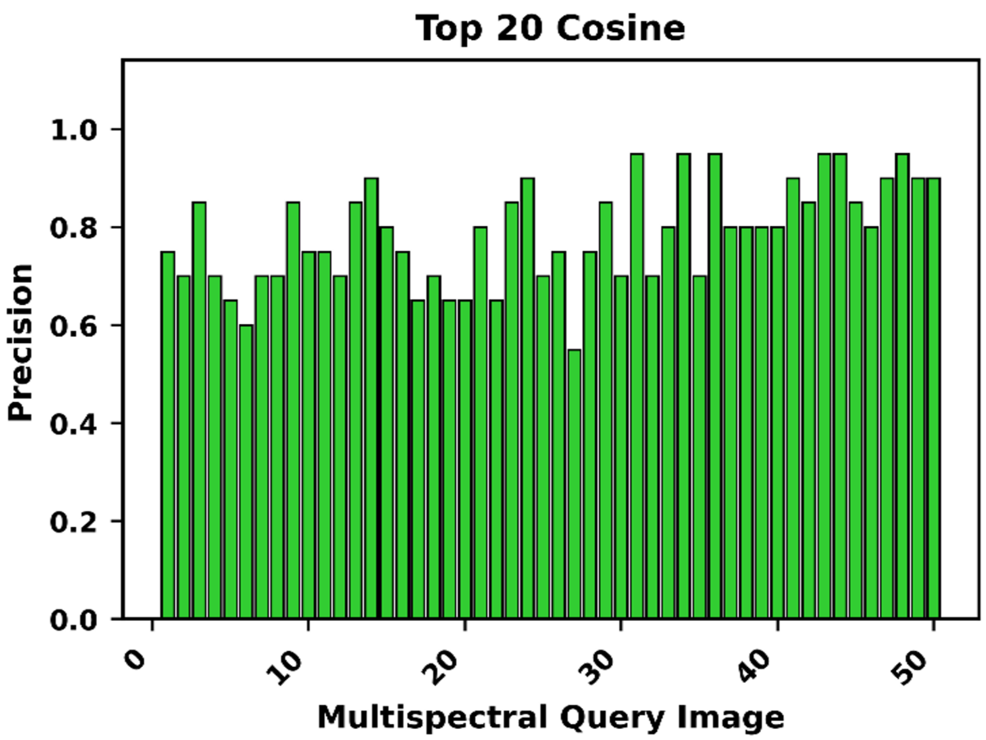

Figure 13. Precision values for the retrieval results of the Top 20 similar images using Cosine Distance.

In Figure 14, for each of the query image Precision is calculated for similar Top 30 images.

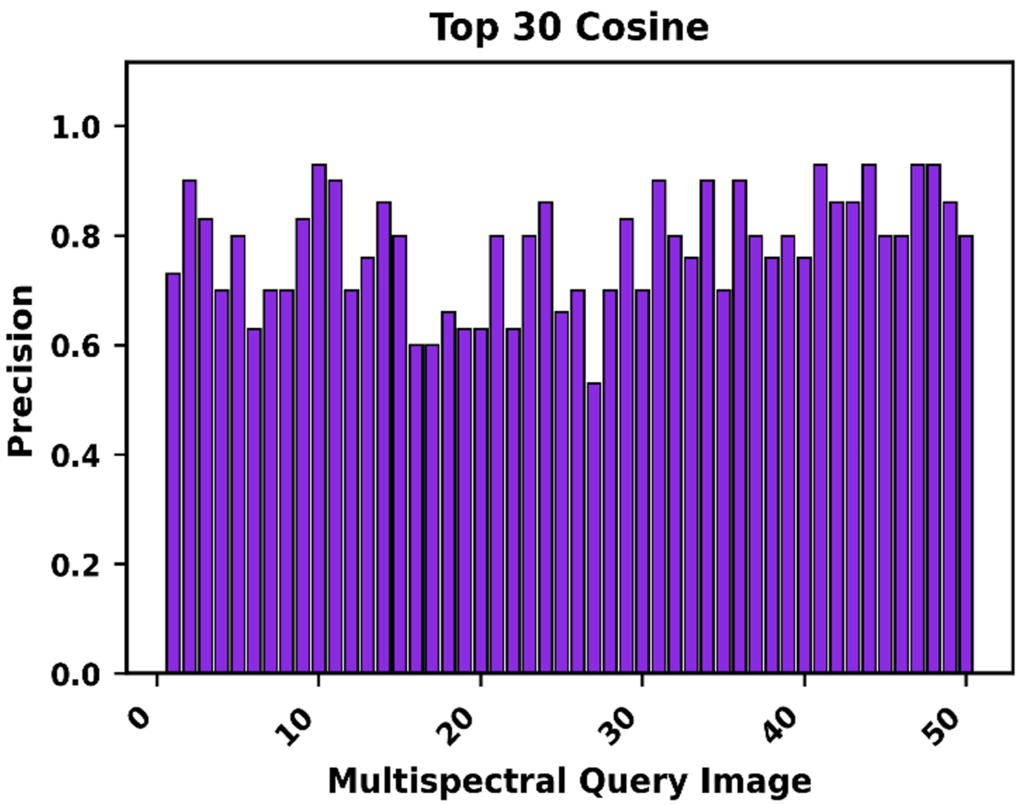

Figure 14. Precision values for the retrieval results of the Top 30 similar images using Cosine Distance.

In Figure 15, for each of the query image Precision is calculated for similar Top 50 images.

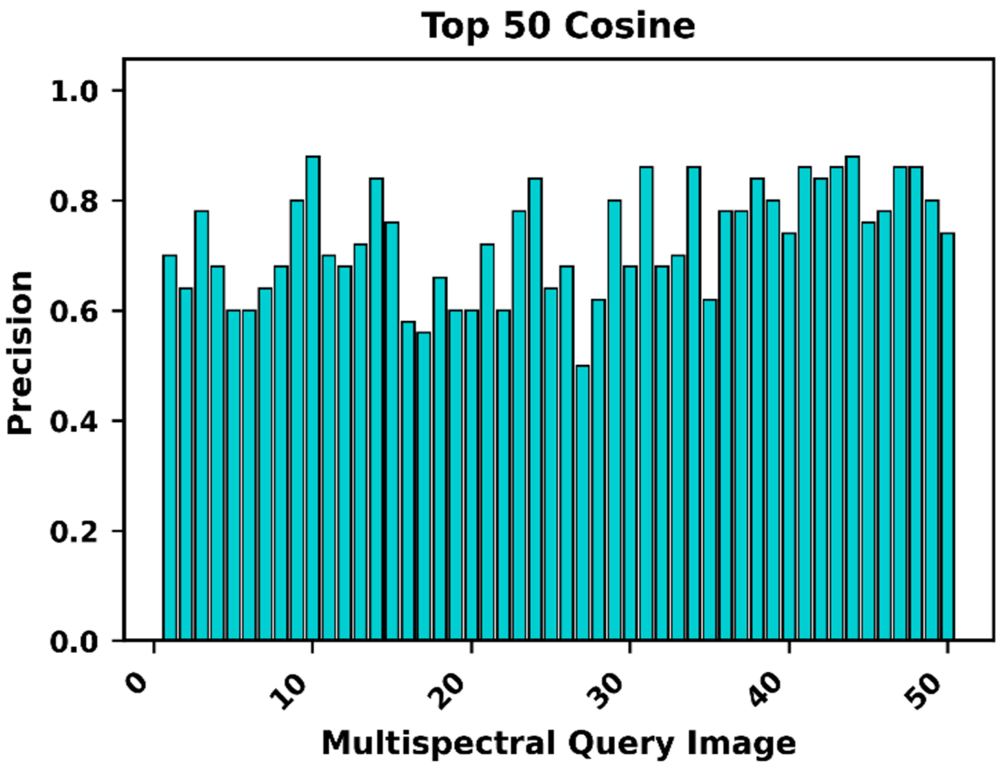

Figure 15. Precision values for the retrieval results of the Top 50 similar images using Cosine Distance.

In Figure 16, for each of the query image Precision is calculated for similar Top 100 images.

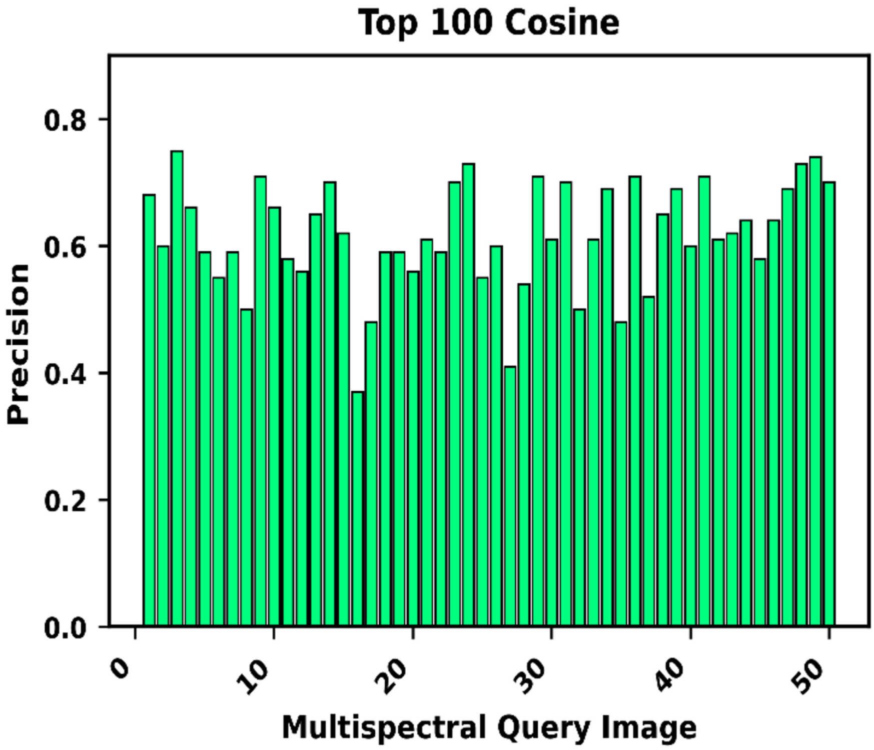

Figure 16. Precision values for retrieval results of the Top 100 similar images using Cosine Distance.

The following observations are made from these comparisons.

1. Nature of Data in EuroSAT – The dataset contains 13 spectral bands capturing distinct physical and spectral properties (e.g., visible, near-infrared), which can be more effectively leveraged by Euclidean distance, as it measures direct magnitude differences across all bands.
2. Magnitude Sensitivity – Euclidean distance preserves differences in absolute reflectance values across bands, which can be important for discriminating land cover types.
3. High-Dimensional Representation – While high-dimensional feature vectors may increase computational cost, Euclidean distance can still perform effectively when the dimensions represent meaningful, independent measurements such as spectral reflectance.

In the EuroSAT dataset, each band captures a specific aspect of the terrain, and differences in values across these bands can indicate different land cover types. Euclidean distance can leverage these differences effectively for analysis and classification purposes.

Multispectral images generate high-dimensional feature vectors, with each band adding a new dimension. Euclidean distance more effectively manages high-dimensional vectors where magnitude variations across bands are essential for distinguishing images, emphasizing the importance of accurate analysis.

After finding precision for each query image at various recall values, such as Top 10, Top 20, Top 30, Top 50, and Top 100, MAP is calculated by averaging these precision values across different query images.

This MAP is summarized in Table 4.

Table 4. MAP for Top N values.

Dataset	Top 10	Top 20	Top 30	Top 50	Top 100
EuroSAT	0.93	0.90	0.86	0.81	0.73

For further comparison, retrieval performance was also measured using Cosine similarity. Figure 17 shows that Euclidean similarity consistently outperforms Cosine similarity for the EuroSAT dataset across the evaluated cut-off points.

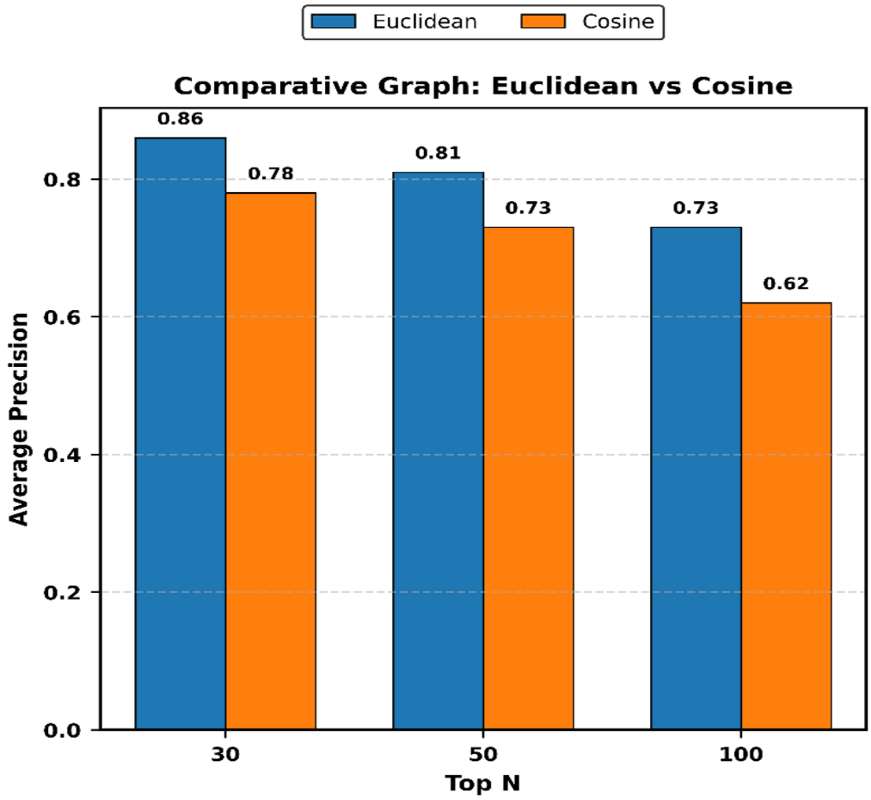

Figure 17. Comparison of multispectral image retrieval for Euclidean and Cosine similarity.

5. COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART MULTISPECTRAL IMAGE RETRIEVAL METHODS

Spectral bands play a vital role in representing multispectral images. It is observed from the above results that considering all 13 spectral bands for feature representation improves the representation of multispectral images, leading to better retrieval results. The results are obtained for various recall values for retrieving the Top 10, Top 20, Top 30, Top 50, and Top 100 images for a given query image, summarized in Table 4 and in Figure 18. Among these results, for the top 100, the MAP value is 0.73, showing improvement compared to results using end-to-end hashing, which has a MAP@100 of 0.68 [18].

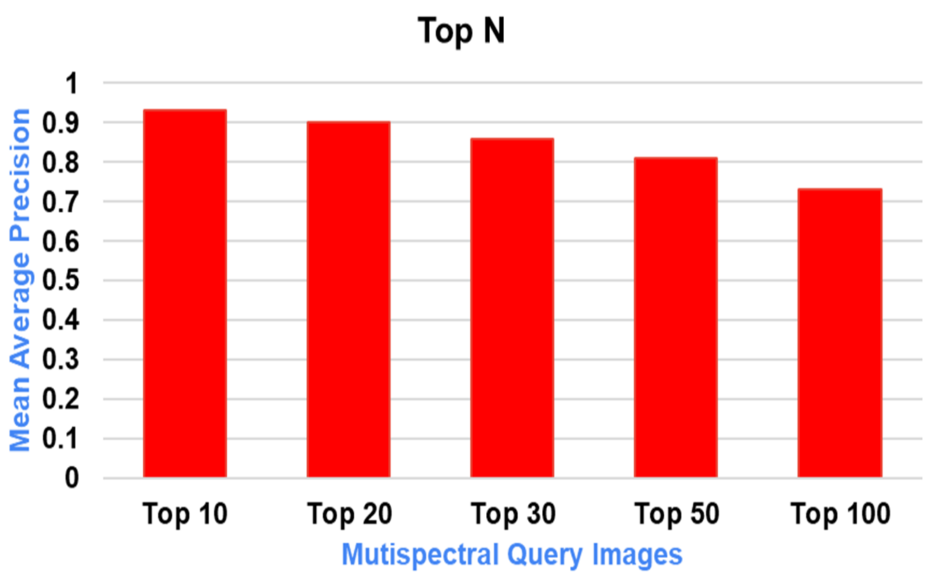

Figure 18. Retrieval results of mean average precision for top 10, Top 20, Top 30, Top 50, and top 100 images using Euclidean similarity.

Figure 18 summarizes the MAP values for different retrieval depths using Euclidean similarity. The MAP for Top 100 retrievals was 0.73, which is higher than the MAP@100 value of 0.68 reported by Chen and Lu [18] using an unsupervised GAN-based hashing method on the same dataset. The difference indicates an improvement in retrieval accuracy for deeper-ranked results under the experimental conditions used in this study.

The proposed method was evaluated against several established approaches for multispectral image retrieval, including SKLSH [21], ITQ [22], SpH [23], SP [24], and End-to-End Hashing [18]. These baseline methods transform high-dimensional feature vectors into binary codes while attempting to preserve similarity or distance relationships. As summarised in Table 5, the proposed method achieved a MAP score of 0.73, the highest among the compared approaches across all three evaluated code lengths.

Table 5. MAP@100 of image retrieval using different methods.

Method	MAP
SKLSH [21]	0.14
ITQ [22]	0.50
SpH [23]	0.51
SP [24]	0.47
End-to-End Hashing [18]	0.68
Proposed Method	0.73

For direct comparison, when using Densely Connected Convolutional Networks (DenseNet) for feature representation with Euclidean similarity, the MAP score also reached 0.73. While these results are promising, they are based on the EuroSAT dataset under current experimental settings, and further evaluation on additional datasets is necessary to confirm the method's robustness.

5.1. Strengths of the Proposed Model

The new DenseNet-based Multispectral Image Retrieval Framework utilizes the full spectral diversity available from the EuroSAT dataset for feature learning by leveraging all 13 spectral bands. This framework leverages the advantages of dense connectivity over traditional handcrafted descriptors and compact hashing-based techniques; specifically, the dense connectivity design encourages reused features and helps avoid gradient vanishing, leading to a greater ability to learn features representing spectral and spatial information. Additionally, Gabor enhancement used in conjunction with deep feature extraction contributes significantly to the framework's ability to provide spatial patterns, especially over complex land cover types. The experimental results provide evidence that this framework achieved an average precision (MAP@100) of 0.73, outperforming other hashing-based retrieval systems evaluated under the same testing conditions. Thus, the framework can retain critical spectral information that determines the discrimination of imagery for use in remote sensing retrieval tasks.

The systematic evaluation of both Euclidean and Cosine similarity measures as part of a multispectral retrieval framework demonstrates an important aspect of this framework's performance. The evaluation was performed using strict data depot separation (70% – 30% for training – testing), which prevents data leakage and allows for a true assessment of retrieval performance.

The use of L2 normalization of 128D embeddings results in compact yet highly discriminative representations for retrieval at scale. The computational cost of the framework is also lower than that of the most recent transformer-based or foundation model-based retrieval frameworks while providing competitiveness in retrieval accuracy. Additionally, this framework's architecture is modular and extensible, allowing integration with more advanced metric learning techniques or deployment in real-world remote sensing applications such as land-use monitoring, environmental assessment, and managing geospatial archives.

5.2. Limitations of the Proposed Model

Several limitations of this work must be addressed, and while the results are encouraging in terms of promising little drawback to a final product, improving texture representation in multiband remotely sensed data using Gabor filters, further validation needs to be conducted before claiming it can be used on other datasets. First, validation was conducted only on images within the EuroSAT dataset, and images were all captured at an identical spatial resolution (64×64 pixels) and shared the same physical attributes and spectral values. Therefore, it is currently unknown whether these results will generalize to other multispectral datasets capturing images with different resolutions, noise levels, seasons, or sensors. Performance variances may also be very large, depending upon which satellite platform captured the dataset or what type of atmospheric conditions existed at the time of capture.

Second, the authors converted all 13 spectral bands into a single-band image (grayscale) by summing the bands before filtering with Gabor filters for texture-enhancing profiles. This process loses some fine detail in each band that would otherwise be preserved by using 3D spectral–spatial convolutions directly. Furthermore, including all bands results in greater computational complexity, increased memory usage, and higher feature dimensionality, potentially making large archives of remote sensing data impractical. The current analysis also examined only two similarity distance metrics, Euclidean and Cosine similarity, and did not explore advanced metric learning techniques such as Mahalanobis distance, learned similarity functions, or triplet-loss optimization, especially where millennial-scale data were present. Lastly, retrieval analysis was conducted using a predefined number of query images and did not incorporate index structures, approximate nearest neighbor search, or retrieval latency related to real-time performance. Addressing these limitations could improve both the robustness and efficiency of the proposed framework, enhancing its practical application in operational deployments.

6. CONCLUSION

The objective of this research was to examine how utilizing the entire range of available spectral bands for multispectral image representation impacts content-based image retrieval effectiveness. The analysis determined that using all available spectral bands for DenseNet-based feature extraction significantly enhances the overall performance of top-N image retrieval tasks when assessed with both Euclidean and Cosine distance criteria. It is hypothesized that the improvements in retrieval performance result from the use of richer, more complete spectral data, which provides a greater ability to create distinct features for determining image similarity. Conversely, because employing all available spectral bands increases the total number of dimensions of the feature vectors produced, it requires additional computational resources and takes longer to retrieve images. The experiments were conducted on a single multispectral image dataset, limiting the extent to which the retrieved image results may directly apply to other multispectral datasets with different acquisition conditions, resolution, and noise characteristics. The study also did not consider alternative dimensionality reduction or spectral band selection techniques to reduce computational burden while maintaining retrieval performance. Future research could focus on identifying optimal spectral band selection strategies, integrating feature compression techniques, exploring alternative distance metrics such as Mahalanobis and Canberra distance, and validating the approach using larger, more diverse multispectral datasets to demonstrate both the reliability and scalability of the method.

Funding: The authors declare that no funding was received for this research.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.
Disclosure of AI Use: The author used OpenAI's ChatGPT (GPT-4) to edit and refine the wording of the Introduction and Literature Review. All outputs were thoroughly reviewed and verified by the author.

REFERENCES

[1] M. P. Mahajan and S. M. Kamalapur, "Spectral imaging," *International Journal of Modern Electronics and Communication Engineering*, vol. 7, no. 1, pp. 144-148, 2019.

[2] P. Kot, A. Shaw, K. O. Jones, J. D. Cullen, A. Mason, and A. I. Al-Shamma'a, "The feasibility of using electromagnetic waves in determining the moisture content of building fabrics and the cause of the water ingress," *International Journal on Smart Sensing and Intelligent Systems*, vol. 7, no. 5, pp. 1-5, 2014. https://doi.org/10.21307/ijssis-2019-136

[3] M. Aboras, H. Amasha, and I. Ibraheem, "Early detection of melanoma using multispectral imaging and artificial intelligence techniques," *American Journal of Biomedical and Life Sciences*, vol. 3, no. 2-3, pp. 29-33, 2015.

[4] V. B. Raju and E. Sazonov, "Detection of oil-containing dressing on salad leaves using multispectral imaging," *IEEE Access*, vol. 8, pp. 86196-86206, 2020. https://doi.org/10.1109/ACCESS.2020.2992326

[5] S. Chand, A. Raj, and M. Sivaprakasam, "Identifying oral cancer using multispectral snapshot camera," presented at the 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2022.

[6] L. L. Chambino, J. S. Silva, and A. Bernardino, "Multispectral facial recognition: A review," *IEEE Access*, vol. 8, pp. 207871-207883, 2020. https://doi.org/10.1109/ACCESS.2020.3037451

[7] F. Hu, M. Zhou, P. Yan, K. Bian, and R. Dai, "Multispectral imaging: A new solution for identification of coal and gangue," *IEEE Access*, vol. 7, pp. 169697-169704, 2019. https://doi.org/10.1109/ACCESS.2019.2955725

[8] A. Spoladore, T. Cavaleri, S. Legnaioli, and V. Palleschi, "Multispectral imaging to reveal ancient hieroglyphic text in an Egyptian Stele," presented at the 2020 6th IEEE Congress on Information Science and Technology (CiSt), IEEE, 2021.

[9] Z. Cui and H. Ji, "Target tracking based on staring multispectral imager," presented at the 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), IEEE, 2022.

[10] M. A. Castillo-Guevara, F. Palomino-Quispe, A. B. Alvarez, and R. J. Coaquira-Castillo, "Water stress analysis using aerial multispectral images of an avocado crop," presented at the 2020 IEEE Engineering International Research Conference (EIRCON), IEEE, 2020.

[11] G. ElMasry, N. Mandour, S. Al-Rejaie, E. Belin, and D. Rousseau, "Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview," *Sensors*, vol. 19, no. 5, p. 1090, 2019. https://doi.org/10.3390/s19051090

[12] J. R. Mansfield, "Multispectral imaging: A review of its technical aspects and applications in anatomic pathology," *Veterinary Pathology*, vol. 51, no. 1, pp. 185-210, 2014. https://doi.org/10.1177/0300985813506918

[13] O. Kleynen, V. Leemans, and M.-F. Destain, "Development of a multi-spectral vision system for the detection of defects on apples," *Journal of Food Engineering*, vol. 69, no. 1, pp. 41-49, 2005. https://doi.org/10.1016/j.jfoodeng.2004.07.008

[14] M. P. Mahajan and S. M. Kamalapur, "Impact of feature representation on remote sensing image retrieval," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 7, pp. 68–77, 2023.

[15] S. D. Newsam and C. Kamath, "Retrieval using texture features in high-resolution multispectral satellite imagery," in *UCRL Conference Proceedings (UCRL-CONF-201981)*, 2004.

[16] C. Joshi and S. Mukherjee, "Empirical analysis of SIFT, Gabor and fused feature classification using SVM for multispectral satellite image retrieval," presented at the 2017 Fourth International Conference on Image Information Processing (ICIIP), IEEE, 2017.

[17] J. Wijitdechakul, S. Sasaki, Y. Kiyoki, and C. Koopipat, "UAV-based multispectral aerial image retrieval using spectral feature and semantic computing," presented at the 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), IEEE, 2017.

[18] X. Chen and C. Lu, "An end-to-end adversarial hashing method for unsupervised multispectral remote sensing image retrieval," presented at the 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020.

[19] J. Gao, X. Shen, P. Fu, Z. Ji, and T. Wang, "Multiview graph convolutional hashing for multisource remote sensing image retrieval," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022. https://doi.org/10.1109/LGRS.2021.3093884

[20] B. Sathiyaprasad and B. S. Kumar, "Retracted: Multi Spectral Image retrieval in remote sensing big data using Fast recurrent convolutional neural network," presented at the 2022 International Conference for Advancement in Technology (ICONAT), IEEE, 2022.

[21] M. Raginsky and S. Lazebnik, "Locality-sensitive binary codes from shift-invariant kernels. In Advances in Neural Information Processing Systems," presented at the Neural Information Processing Systems (NeurIPS) Foundation, United States, 2009.

[22] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2916-2929, 2013. https://doi.org/10.1109/TPAMI.2012.193

[23] J.-P. Heo, Y. Lee, J. He, S.-F. Chang, and S.-E. Yoon, "Spherical hashing Computer Vision and Pattern Recognition (CVPR)," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[24] Y. Xia, K. He, P. Kohli, and J. Sun, "Sparse projections for high-dimensional binary codes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3332-3339.

[25] V. Erin Liong, J. Lu, G. Wang, P. Moulin, and J. Zhou, "Deep hashing for compact binary codes learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2475-2483.

[26] E. Yang, C. Deng, T. Liu, W. Liu, and D. Tao, "Semantic structure-based unsupervised deep hashing," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 1064-1070.

[27] B. Blumenstiel, V. Moor, R. Kienzler, and T. Brunschwiler, "Multi-spectral remote sensing image retrieval using geospatial foundation models," presented at the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2024.

[28] S. Choudhury, Y. Salunkhe, S. Mehrotra, and B. Banerjee, "REJEPA: A novel joint-embedding predictive architecture for efficient remote sensing image retrieval," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 2373-2382.

[29] C. T. Marimo, B. Blumenstiel, M. Nitsche, J. Jakubik, and T. Brunschwiler, "Beyond the visible: Multispectral vision-language learning for earth observation," presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2025.