

Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis

Girish D. Chate ¹, , S. S. Bhamare ¹

1. School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, IND

Received: April 24, 2026 | Review began: May 05, 2026 | Review ended: May 18, 2026 | Published: May 18, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

This paper presents a comprehensive framework for soil image feature extraction and classification using RGB color, luminance, and gray-level co-occurrence matrix (GLCM)-based texture descriptors. A dataset of 1,300 soil images representing black, laterite, red, and white soil categories was used for evaluation. The proposed framework integrates color, brightness, and texture information to improve soil representation and classification performance. K-Nearest Neighbors classification was evaluated using Euclidean, Manhattan, Chebyshev, and Minkowski distance metrics. Experimental results demonstrate that feature fusion improves classification performance compared to individual feature domains. Among the evaluated metrics, Euclidean and Chebyshev distances achieved strong classification performance across different feature sets. The proposed framework provides a scalable and efficient solution for image-based soil analysis and precision agriculture applications. Future work will focus on larger and more diverse datasets for real-world validation.

Categories: Image Processing and Analysis, Computer Vision, Machine Learning (ML)

Keywords: glcm, analysis of soil, image processing, machine learning, indian agriculture

Introduction

Soil plays a fundamental role in agriculture, influencing crop productivity, nutrient availability, and overall ecosystem sustainability. Accurate soil analysis is therefore essential for effective land management and precision agriculture. Traditional soil assessment methods rely heavily on laboratory testing [1], which is often time-consuming, labor-intensive, and not suitable for real-time or large-scale applications. With the advancement of computer vision and image processing techniques, soil image analysis has emerged as a promising alternative for rapid and non-destructive soil characterization.

Soil images contain rich visual information such as color, texture, and brightness, which are closely related to soil composition, moisture content, and organic matter. Among these, color features provide useful insights into mineral composition and fertility levels, while luminance captures variations in brightness that may indicate moisture or surface conditions [2]. In addition, texture features play a critical role in distinguishing soil types that may appear visually similar in terms of color but differ in structural patterns.

Recent research has explored the use of image-based feature extraction techniques for soil classification; however, many existing approaches focus on a limited set of features, which may reduce classification accuracy and robustness. In particular, methods that rely solely on color information often fail to capture spatial relationships between pixels, while texture-only approaches may overlook important spectral characteristics [3]. Therefore, there is a need for a comprehensive framework that integrates multiple feature types to improve the reliability of soil analysis.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

In this paper, a unified framework is proposed for extracting and analyzing soil image features using red, green, and blue (RGB) color descriptors, luminance information, and texture properties derived from the Gray-Level Co-occurrence Matrix (GLCM) [4]. RGB features are used to represent the color distribution of soil samples, while luminance is computed to capture brightness variations. GLCM-based texture features, including contrast, dissimilarity, homogeneity, and energy, are extracted to model spatial relationships between pixel intensities.

To validate the effectiveness of the proposed approach, a dataset of 1,300 soil images is used, representing a variety of soil types [5]. The extracted features are further evaluated using multiple distance-based similarity measures, including Euclidean, Manhattan, Chebyshev, and Minkowski distances. This allows a comparative analysis of feature effectiveness across different metrics and feature combinations.

The main contribution of this work lies in the integration of color, luminance, and texture features into a single framework, along with a detailed evaluation of distance metrics for soil classification. The proposed approach aims to improve classification accuracy, enhance feature representation, and provide a scalable solution for real-world agricultural applications.

Unlike existing studies that rely only on color or texture information, the proposed framework integrates RGB, luminance, and GLCM-based texture descriptors into a unified feature extraction pipeline. In addition, this work provides a comparative analysis of multiple distance metrics within the k-nearest neighbors (KNN) framework for soil image classification. The research aims to support rapid, low-cost, and non-destructive soil analysis for precision agriculture applications.

Literature review

Soil analysis using traditional laboratory methods has long been considered the standard approach for determining soil properties such as texture, composition, and fertility. However, these methods are often time-consuming, expensive, and require expert intervention. To overcome these limitations, researchers have increasingly explored image processing and computer vision techniques for rapid and non-destructive soil analysis.

Early studies in soil image analysis primarily focused on color-based feature extraction, as soil color is closely related to its mineral composition and organic content. RGB color models have been widely used to represent soil images due to their simplicity and effectiveness. For instance, several researchers have demonstrated that mean and standard deviation values of RGB channels can be used for basic soil classification tasks [6]. However, relying solely on color features often leads to misclassification, especially when different soil types exhibit similar color distributions under varying lighting conditions.

To address these limitations, researchers have incorporated luminance and color space transformations such as grayscale; hue, saturation, and value; and LAB models. Luminance-based features help capture brightness variations and improve robustness against illumination changes. Studies have shown that separating luminance from chromatic components enhances feature discrimination and improves classification accuracy in soil image analysis [7,8].

In addition to color and luminance, texture analysis has gained significant attention for its ability to describe spatial patterns within soil images. Among various texture extraction techniques, the GLCM is one of the most widely used methods. GLCM captures second-order statistical information by analyzing the spatial relationship between pixel intensities. Features such as contrast, dissimilarity, homogeneity, and energy derived from GLCM have been successfully applied in soil classification and segmentation tasks [9]. These features are particularly useful for distinguishing soils with similar color but different structural characteristics.

Hybrid approaches have demonstrated better accuracy compared to single-feature methods by capturing complementary information from different feature domains [10]. Recent advancements have emphasized the importance of multi-feature integration, combining color, luminance, and texture descriptors to improve classification performance [11]. Such integration is especially beneficial in complex soil datasets where variability in lighting, texture, and composition can affect model performance.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. *Cureus J Comput Sci* 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

Furthermore, various distance-based and machine learning methods have been employed to evaluate and classify soil features. Distance metrics such as Euclidean, Manhattan, Chebyshev, and Minkowski distances are commonly used for similarity measurement and classification tasks due to their computational simplicity and effectiveness [12]. Comparative studies indicate that the performance of these metrics varies depending on the feature space, highlighting the need for systematic evaluation across different feature combinations.

The proposed a smartphone-based approach for soil pH prediction using digital image processing techniques. Their method extracts color features such as hue, saturation, and value from soil images and applies machine learning models including Linear Regression, Artificial Neural Networks, and KNN for prediction [13].

Several studies have also emphasized the importance of texture analysis in image classification. GLCM-based texture features have been widely used to capture spatial relationships between pixels, enabling effective differentiation between visually similar classes. These methods are particularly useful in soil analysis, where texture variations play a key role in classification [14].

Kiran Pandiri et al. [15] proposed a material classification approach using combined color and texture features. Their work demonstrated that integrating multiple feature types improves classification performance compared to using a single feature domain. This finding aligns with the concept of feature fusion, where complementary information from color and texture enhances discriminative capability [15].

Despite these advancements, existing studies often focus on limited feature sets or lack a comprehensive evaluation of multiple distance metrics across diverse feature domains. There is still a gap in developing a unified framework that integrates RGB, luminance, and GLCM-based texture features while systematically analyzing their performance using different similarity measures.

This study proposes a comprehensive feature extraction and evaluation framework that combines color, luminance, and texture descriptors with distance-based analysis. This approach aims to enhance soil classification accuracy and provide a more robust and scalable solution for real-world agricultural applications.

Materials And Methods

Dataset of soil images

A total of 1,300 soil images are collected and prepared to construct the dataset used in this study. The images are captured directly from agricultural fields using a digital camera under natural environmental conditions, ensuring that the dataset reflects real-world variability in soil appearance. The dataset includes multiple soil categories, namely black, laterite, red, and white soils, covering a diverse range of soil textures and color characteristics.

All captured images have an original resolution of 800×800 pixels, providing sufficient detail for feature extraction. To enhance the robustness and generalization capability of the proposed framework, several preprocessing and data augmentation techniques were applied. These include image scaling into patches, rotation (twisting), cropping, and inversion, which help simulate variations in orientation, perspective, and lighting conditions commonly encountered in practical scenarios.

After preprocessing, all images were resized to a uniform dimension of 224×224 pixels to ensure compatibility with standard image processing and machine learning workflows. This normalization step also reduces computational complexity while preserving essential visual features required for analysis.

The prepared dataset provides a balanced representation of soil types and supports effective extraction of color, luminance, and texture features. These processed images serve as the input for subsequent feature extraction and classification stages in the proposed framework. Representative samples of the soil categories used in this paper are illustrated in Figure 7, showing visual differences among black, laterite, red, and white soils.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

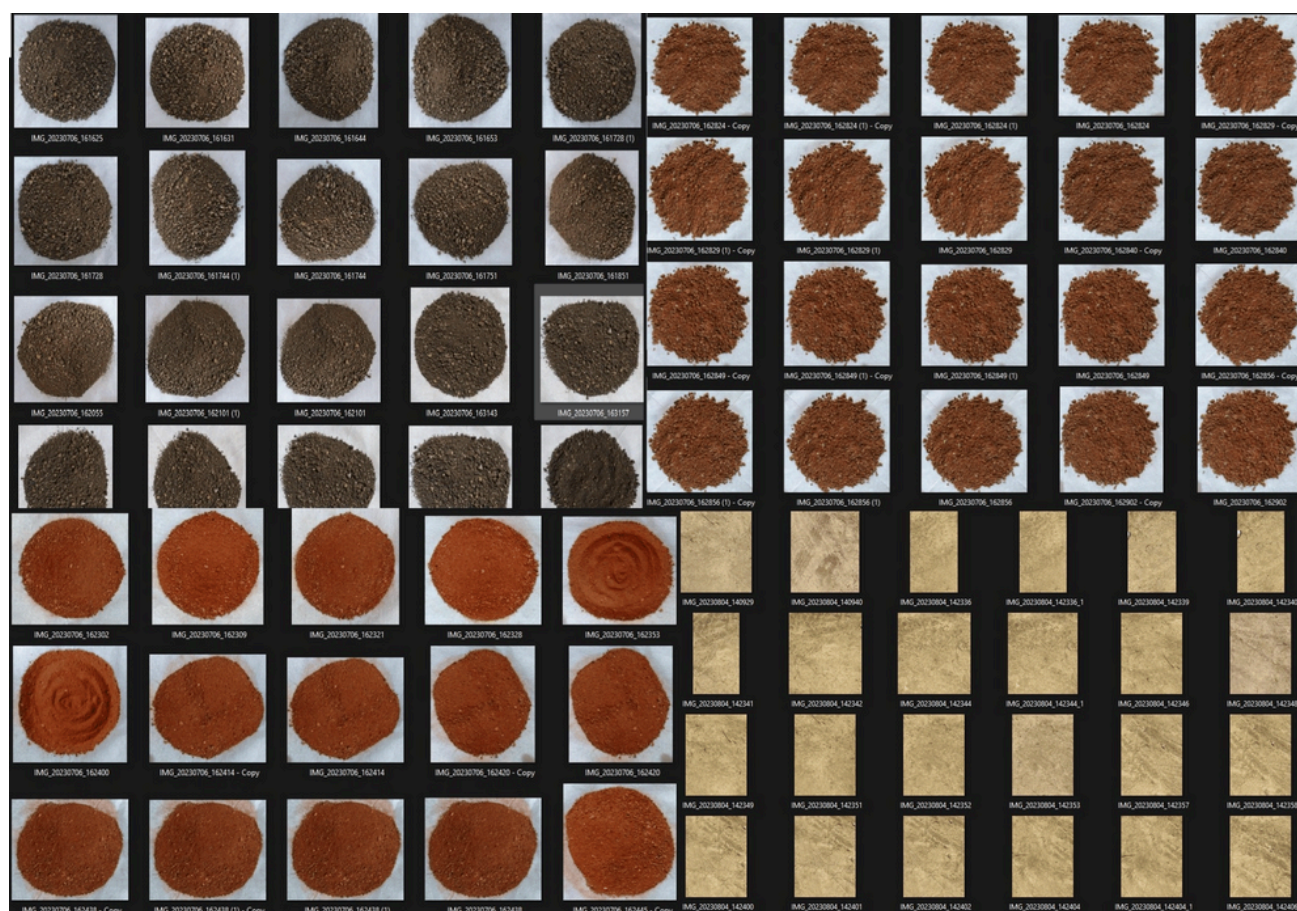


FIGURE 1: Soil images from the proposed dataset

Source: Author's own dataset developed during field survey

Dataset splitting and validation

The dataset was divided into training and testing sets using an 80:20 ratio. To avoid data leakage, augmented versions of the same image are maintained within the same subset during splitting. Training images are used for feature extraction and classifier learning, while testing images are used only for performance evaluation. To improve the reliability of the experimental evaluation, k-fold cross-validation was considered during model analysis. This helps reduce bias and improves the robustness of the reported classification results.

Methodology

Figure 2 illustrates the overall framework proposed for soil image feature extraction and classification. The system is designed to process a structured soil image dataset, where images are organized into directories corresponding to different soil classes. The methodology consists of four major stages: preprocessing, feature extraction (RGB, luminance, and texture), feature integration, and classification using the KNN algorithm with multiple distance metrics.

Image acquisition and preprocessing

The input to the system is a soil image dataset consisting of labeled images stored in class-specific folders. Each image is first subjected to preprocessing steps to ensure uniformity and improve feature quality. These steps include resizing all images to 224×224 pixels, normalization, and optional augmentation techniques such as rotation, cropping, and inversion. Preprocessing helps reduce noise, standardize input dimensions, and enhance the robustness of the feature extraction process.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. *Cureus J Comput Sci* 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

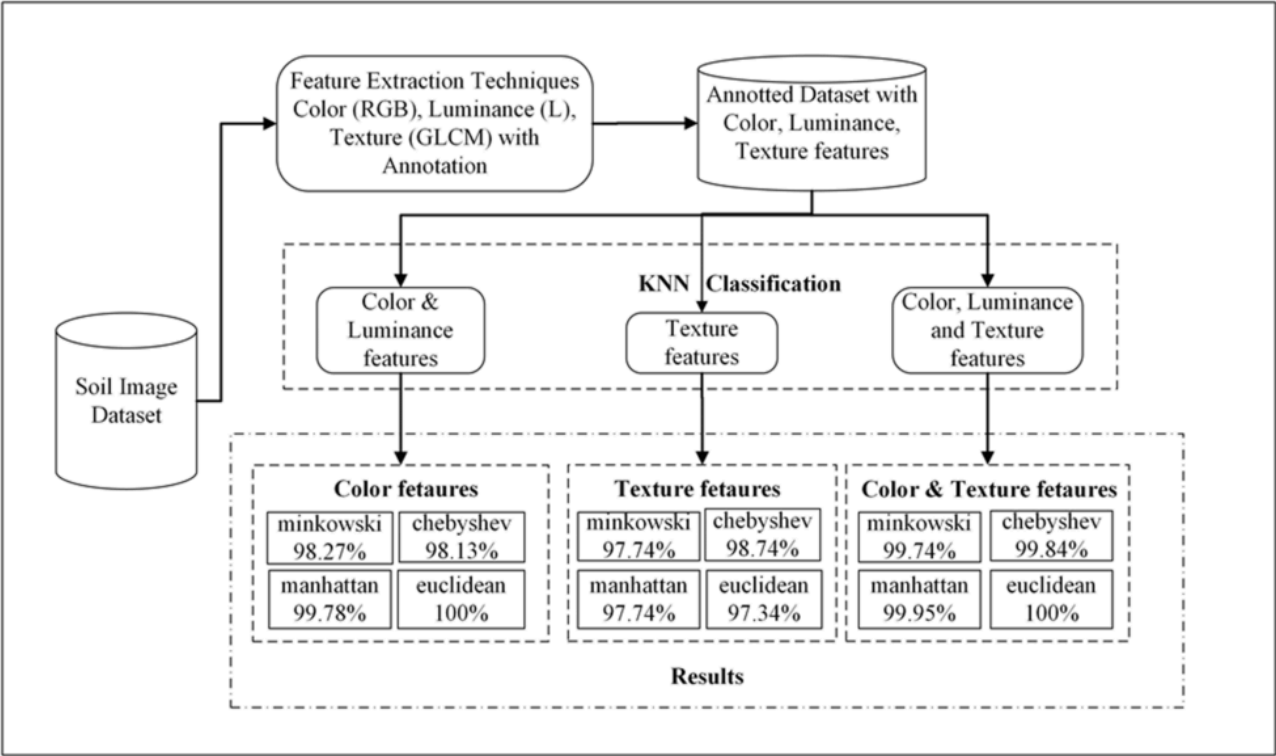


FIGURE 2: Proposed framework
GLCM, Gray Level Co-occurrence Matrix; KNN, K-Nearest Neighbors

RGB color feature extraction

Color features are extracted from each soil image using the RGB color model. The RGB channels are separated, and statistical measures such as mean values are computed for each channel. These features represent the overall color distribution of the soil sample and provide important information related to mineral composition and soil type. RGB features are simple yet effective for distinguishing soils with distinct color characteristics. The algorithm in Figure 3 is used to compute RGB descriptor for an image.

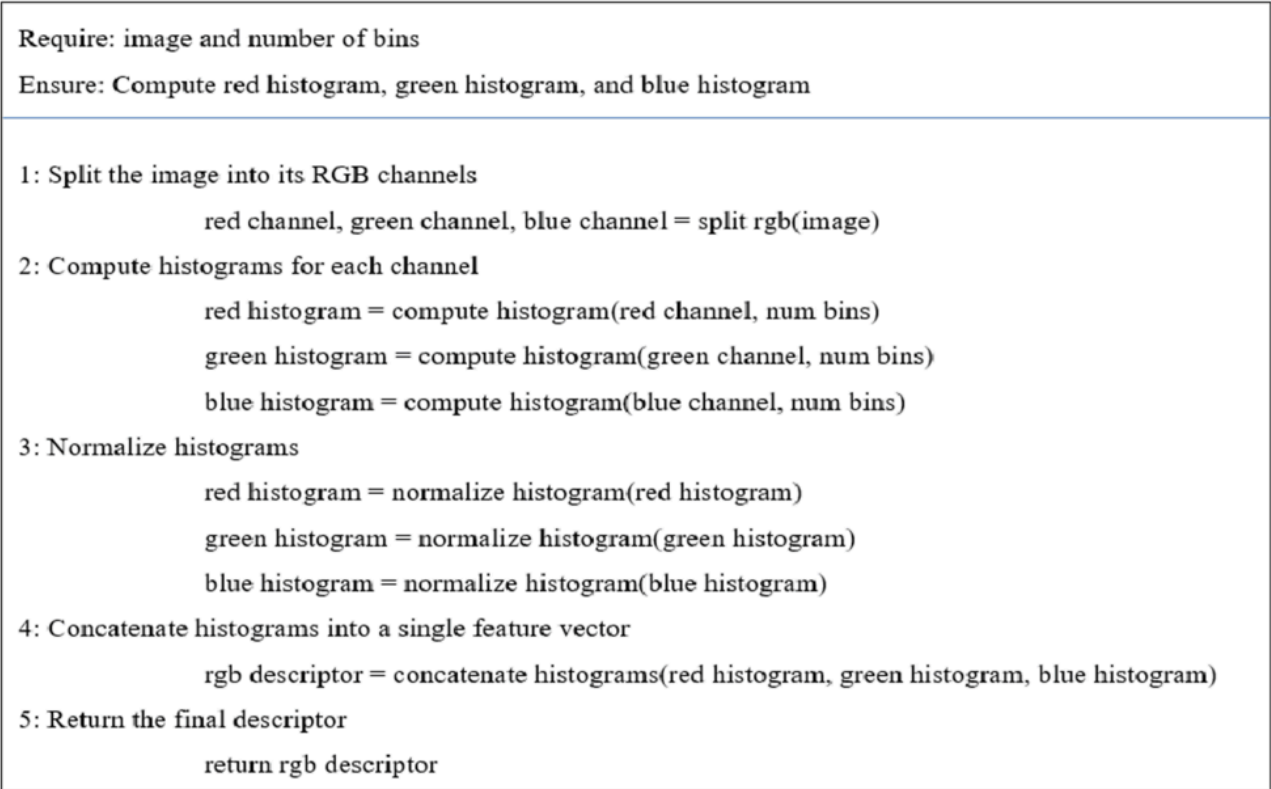


FIGURE 3: Algorithm for RGB color feature extraction for an image

RGB, Red Green Blue

Luminance feature extraction

In addition to RGB features, luminance (brightness) is computed to capture intensity variations within the image. Luminance is derived from the RGB channels using standard transformation techniques and represents the grayscale intensity of the image. This feature is particularly useful for handling variations in lighting conditions and enhances the reliability of the system under different environmental settings. Luminance provides brilliant statistics for a colorless image. Hues become brighter or darker as luminosity values are increased or decreased, accordingly. Luminance is characterized by,

$$YLuminance = 0.216R + 0.7152G + 0.0722B$$

Table 1 presents the RGBL descriptors, which provide a quantitative representation of the color composition of black, red, laterite, and white soil images.

How to cite this article:

Black Soil	Red Soil	Laterite Soil	White Soil
Red = 102.2643	Red = 95.8775	Red = 102.0306	Red = 167.6828
Green = 105.1301	Green = 107.7677	Green = 144.1730	Green = 168.3074
Blue = 110.1706	Blue = 134.4262	Blue = 117.5009	Blue = 167.3117
Luminance = 142.4649	Luminance = 151.6730	Luminance = 123.7159	Luminance = 167.6617

TABLE 1: RGBL features of black soil, red soil, laterite soil, and white soil

RGBL, Red Green Blue Light

Texture feature extraction using GLCM

To capture spatial and structural information, texture features are extracted using the GLCM. The GLCM analyzes the frequency of pixel intensity pairs at specific orientations and distances within the image. From the GLCM, several statistical texture features are computed, including:

Contrast: Contrast measures the local variations in the image. Significant variations in the brightness of nearby pixels are indicated by high contrast values. The contrast formula in the context of GLCM is defined as follows:

$$\text{Contrast} = \sum_{(i,j=0)} (N-1)p_{ij}(i-j)^2$$

here,

- N is the number of different gray levels in the image of 256 for an 8-bit grayscale image.
- i and j are the gray levels pixel intensities in the image.
- $p(i,j)$ is the probability that a pixel in the GLCM with intensity i appears adjacent to a pixel with intensity j .

Dissimilarity: Dissimilarity calculates the average intensity difference between adjacent pixels. Greater textural heterogeneity is indicated by high dissimilarity scores. The dissimilarity formula is as follows:

$$\text{Dissimilarity} = \sum_{(i,j=0)} (M-1)p_{ij}|i-j|$$

here,

- M is the number of distinct gray levels in the image.
- i and j are the gray levels (pixel intensities) in the image.
- $p(i,j)$ is the probability of a pixel with intensity i occurring next to a pixel with intensity j in the GLCM.
- $|i-j|$ represents the absolute difference between the intensity values of the pixel pairs.

Homogeneity: Homogeneity shows how closely the GLCM diagonal and the element distribution are related. A more uniform texture is suggested by high homogeneity values, which show that the components are concentrated along the diagonal.

$$\text{Homogeneity} = \sum_{(i,j)} 1/(1+(i-j)^2)p(i,j)$$

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

here,

- i and j are the gray levels (pixel intensities) in the image.
- $P(i,j)$ is the probability of a pixel with intensity i occurring adjacent to a pixel with intensity j in the GLCM.
- $(i-j)^2$ is the squared difference between the gray levels.
- The term $1/1+(i-j)^2$ gives higher weight to lower differences in intensity, contributing more to homogeneity.

Energy: Energy represents textural uniformity.

$$Energy = \sum_{i,j=0}^{N-1} P(i,j)^2$$

here,

- N is the number of distinct gray levels in the image.
- i and j are the gray levels (pixel intensities) in the image.
- $P(i,j)$ is the probability of a pixel with intensity i occurring adjacent to a pixel with intensity j in the GLCM.
- $P(i,j)^2$ emphasizes repeated or dominant texture patterns, contributing to the uniformity of the image.

These features are essential for distinguishing soil types that may appear similar in color but differ in texture. Table 2 shows the GLCM features of the black, red, laterite, and white soil images.

Black Soil	Red Soil	Laterite Soil	White Soil
Contrast = 351.4795	Contrast = 101.4394	Contrast = 174.2718	Contrast = 200.2302
Energy = 599.7709	Energy = 149.1404	Energy = 241.2088	Energy = 416.2798
Dissimilarity = 899.3403	Dissimilarity = 143.0209	Dissimilarity = 241.8993	Dissimilarity = 373.4308
Homogeneity = 559.2356	Homogeneity = 105.1345	Homogeneity = 175.4430	Homogeneity = 450.2612

TABLE 2: GLCM features of black, red, laterite, and white soil images

GLCM, Gray Level Co-occurrence Matrix

Feature integration and dataset annotation

The extracted RGB, luminance, and GLCM texture features are combined to form a comprehensive feature vector for each image. The dataset is then annotated with these features and corresponding class labels. This structured dataset is stored in CSV format, enabling easy integration with machine learning algorithms and further analysis.

Classification using KNN

The classification stage uses the KNN algorithm, a simple yet effective supervised learning method. KNN classifies a given test sample based on the majority class of its nearest neighbors in the feature space. To evaluate the effectiveness of different feature sets, classification is performed using three configurations such as 1. Color and Luminance features, 2. Texture features, and 3. Combined Color, Luminance, and Texture features. In this research, the value of k was

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

experimentally selected based on classification performance. Several k values are evaluated, and $k = 3$ provided the most stable and accurate results across different feature sets. The classifier was implemented using normalized feature vectors to ensure fair distance computation across all attributes.

Distance metrics for performance evaluation

The KNN looked at each other and used four different distance measurements to determine how far apart they were. All distance measurements are combined with the classifier to decide which one directly identifies the class of the picture and may use the distance function to compute, combining the differences between two objects.

$$Distance(p, q)$$

where p and q are objects with features, and

$$(p_1, p_2, \dots, p_N), q = (q_1, q_2, \dots, q_N)$$

The performance of the KNN classifier is evaluated using multiple distance metrics, including:

Euclidean Distance

$$Distance_{Euclidean}(q, p) = \sqrt{\sum_{i=0}^n (q_i - p_i)^2}$$

here,

- $p = (p_0, p_1, \dots, p_n)$ and $q = (q_0, q_1, \dots, q_n)$ are two objects or feature vectors in an $(n+1)$ dimensional space.
- p_i and q_i are the features of objects p and q , respectively.
- $(q_i - p_i)^2$ is the squared difference between the corresponding features.
- The sum $\sum_{i=0}^n (q_i - p_i)^2$ gives the total squared distance across all features.
- The square root computes the actual Euclidean distance.

Manhattan Distance

$$Distance_{Manhattan}(q, p) = \sum_{i=0}^n |q_i - p_i|$$

here,

- $p = (p_0, p_1, \dots, p_n)$ and $q = (q_0, q_1, \dots, q_n)$ are two objects or feature vectors.
- p_i and q_i are the i^{th} features of objects p and q , respectively.
- $|q_i - p_i|$ is the absolute difference between the corresponding features.
- The sum $\sum_{i=0}^n |q_i - p_i|$ gives the total distance between the two vectors.

Chebyshev Distance

$$Distance_{Chebyshev}(q, p) = (\max_i) (|q_i - p_i|)$$

here,

- $p = (p_0, p_1, \dots, p_n)$ and $q = (q_0, q_1, \dots, q_n)$ are two objects (feature vectors).

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

- p_i and q_i are the i^{th} features of objects p and q , respectively.
- $|q_i - p_i|$ is the absolute difference between the corresponding features.
- The sum $\sum_{i=0}^n |q_i - p_i|$ gives the total distance between the two vectors.

Minkowski Distance

$$\llbracket \text{Distance} \rrbracket_{\text{Minkowski}}(q, p) = \left(\sum_{i=0}^n |q_i - p_i|^F \right)^{1/F}$$

here,

- $p = (p_0, p_1, \dots, p_n)$ and $q = (q_0, q_1, \dots, q_n)$ are two objects (feature vectors).
- q_i and p_i represent the i^{th} features of q and p , respectively.
- F is the order of the norm (e.g., $F = 1$ for Manhattan, $F = 2$ for Euclidean).
- $|q_i - p_i|^F$ is the absolute difference raised to the power of F .
- The entire expression is raised to the power $1/F$ to compute the Minkowski distance.

Each metric measures similarity between feature vectors in a different way, influencing classification accuracy. The comparative analysis of these metrics helps identify the most suitable distance measure for different feature combinations.

KNN-based classification for soil image label retrieval

To classify soil images based on the extracted features, the KNN algorithm is employed as a supervised learning approach. KNN is chosen due to its simplicity, effectiveness, and suitability for distance-based similarity analysis. The algorithm assigns a class label to each test sample by identifying the k closest training samples in the feature space and determining the majority class among them. The performance of the algorithm depends on the choice of distance metric and the value of k , which controls the number of neighbors considered during classification.

The algorithm in Figure 4 begins by loading both the training dataset and the testing dataset, which contain feature vectors derived from soil images along with their corresponding class labels.

Initially, an empty set is created to store the predicted class labels for the test samples. The algorithm then iterates through each test sample and computes its distance from every training sample using a selected distance metric (such as Euclidean, Manhattan, Chebyshev, or Minkowski).

For each test sample, a list of nearest neighbors is maintained. As distances are computed, they are compared and stored in the nearest neighbor set. If the number of neighbors is less than k , new distances are added to the set. Once the number of stored neighbors reaches k , the algorithm ensures that only the closest k distances are retained by discarding larger distances if necessary.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

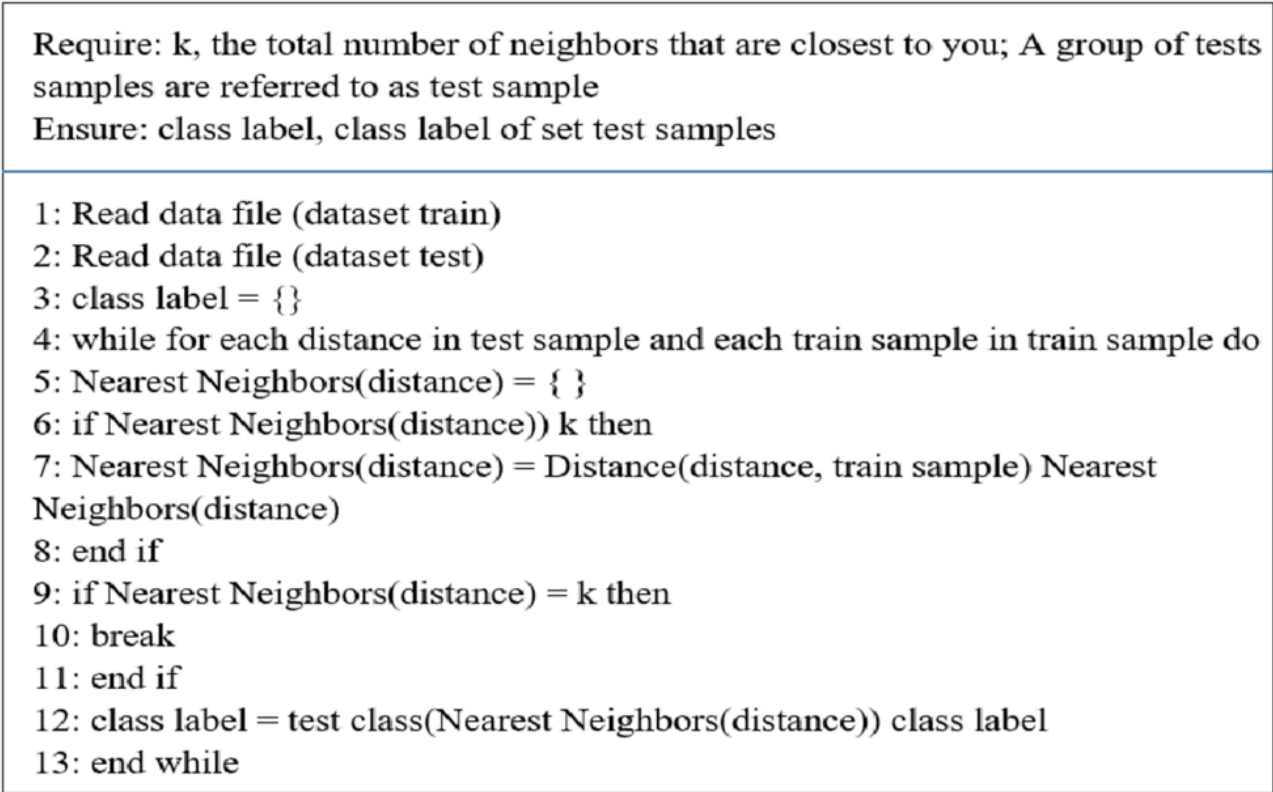


FIGURE 4: Proposed KNN algorithm for retrieve label of a class

KNN, K-Nearest Neighbors

When exactly KNNs are identified, the search process for that test sample is terminated. The class labels of these nearest neighbors are then examined, and the final class label is assigned to the test sample based on the majority voting principle.

This process is repeated for all test samples, and the predicted class labels are stored as the final output. The algorithm effectively classifies soil images by leveraging similarity measures in the feature space and demonstrates strong performance when combined with well-extracted color, luminance, and texture features.

Results And Discussion

The performance of the proposed soil image feature extraction and classification framework was evaluated using a dataset of 1,300 soil images. The evaluation was carried out using the KNN classifier with different distance metrics, namely Euclidean, Manhattan, Chebyshev, and Minkowski distances. The analysis was conducted across three feature configurations: color and luminance features, texture features, and the combined feature set.

For the color and luminance features, the results demonstrate that the Euclidean distance metric achieves the highest classification accuracy of 100%, followed by Manhattan distance with 99.78%, Minkowski distance with 98.27%, and Chebyshev distance with 98.13%. This indicates that color and brightness information are highly discriminative for soil classification, especially when measured using Euclidean distance, which effectively captures overall feature similarity.

The training and testing accuracy curves, as shown in Figure 5, demonstrate that the Euclidean distance metric achieves high classification performance and no observable gap between training and testing results, indicating strong generalization capability. Manhattan distance shows slightly lower performance with minimal variation between training

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

and testing accuracy. In contrast, Minkowski and Chebyshev distances exhibit a noticeable decrease in accuracy.

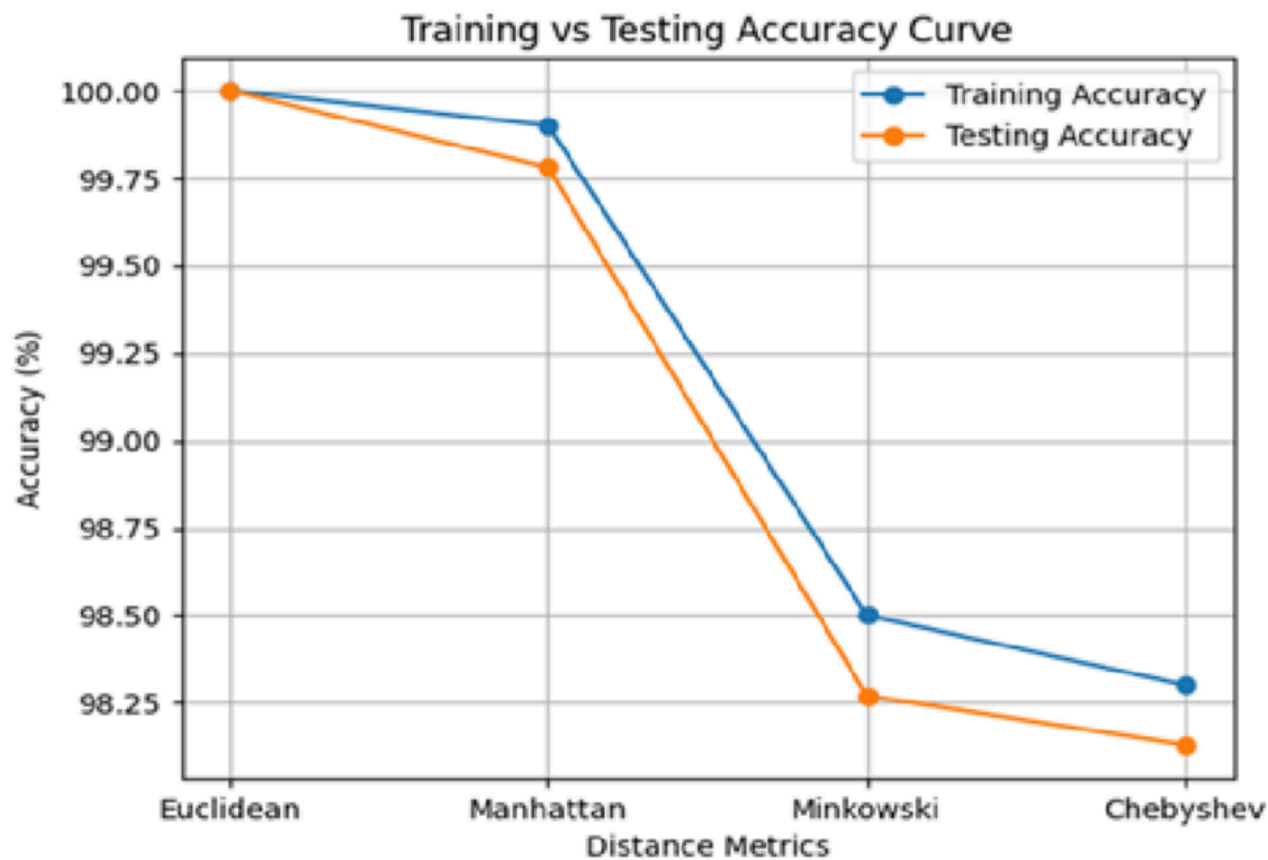


FIGURE 5: Training and testing accuracy curve on color and luminance features

The corresponding loss curves further support these observations. Euclidean distance results in near-zero training and testing loss, as shown in Figure 6, while the other metrics show a gradual increase in loss values. The small gap between training and testing performance across all metrics indicates relatively stable classification behavior under the evaluated experimental conditions. Overall, the results confirm that Euclidean distance is the most effective metric for classification using color and luminance features.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>



FIGURE 6: Training and testing loss curve on color and luminance features

Figure 7 shows a confusion matrix with labeled classes such as black, laterite, red, and white soils. It clearly demonstrates that most samples are correctly classified, as indicated by the dominant diagonal values. The black soil class achieves perfect classification with no misclassifications. Minor confusion is observed between red and laterite soils, and between laterite and white soils, which may be due to similarities in their visual texture and color properties. However, these misclassifications are very limited, confirming the robustness of the proposed feature extraction framework. Overall, the confusion matrix validates the high classification accuracy achieved using Euclidean distance with color and luminance features.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

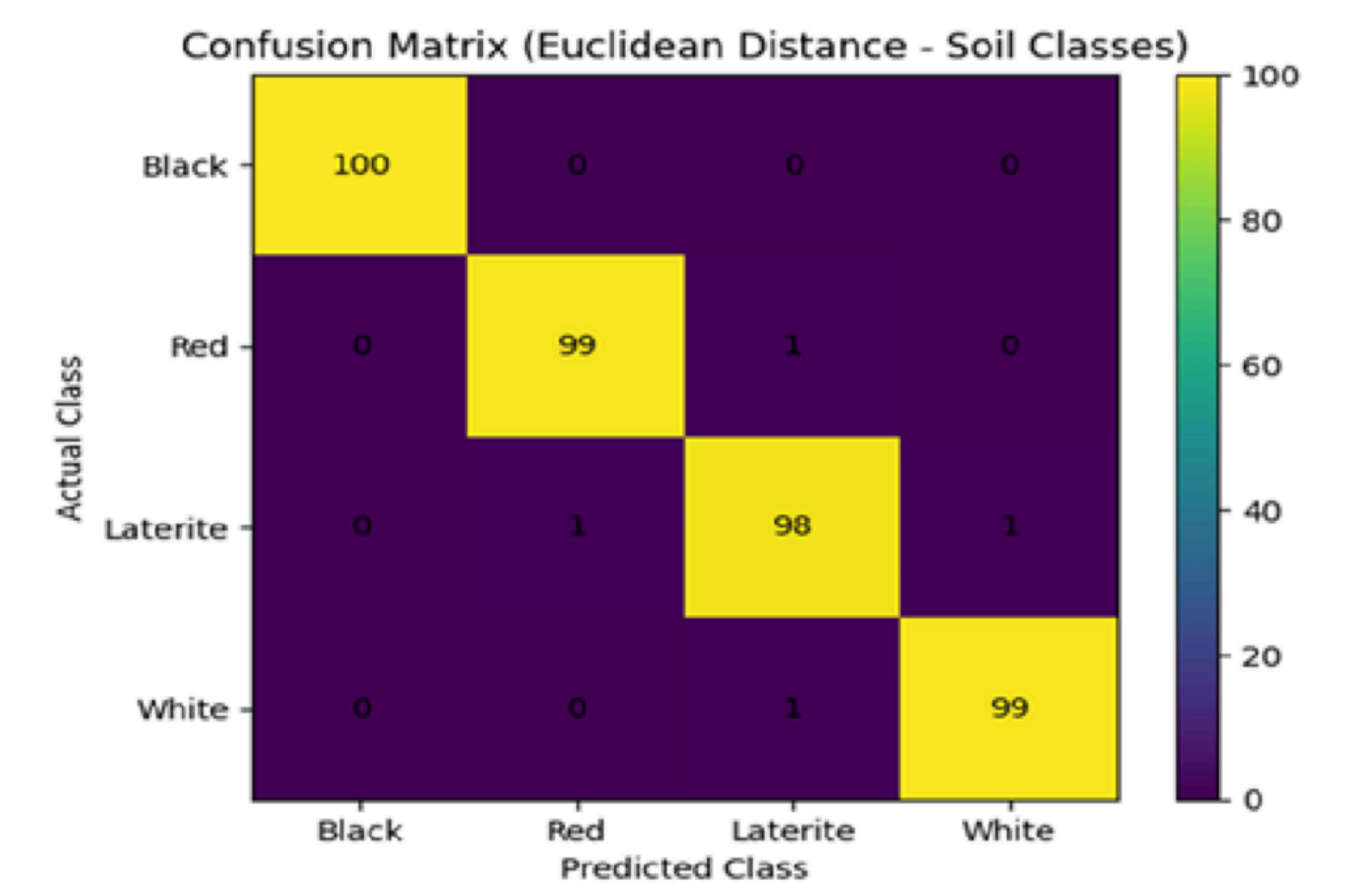


FIGURE 7: Confusion matrix for Euclidean distance on color and luminance features

Figure 8 shows the performance of different distance metrics used in the KNN algorithm. The metrics compared are Euclidean, Manhattan, Chebyshev, and Minkowski. The metrics are evaluated based on features such as color and luminance. The Euclidean metric is the most effective for this dataset, while Chebyshev has the lowest performance.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

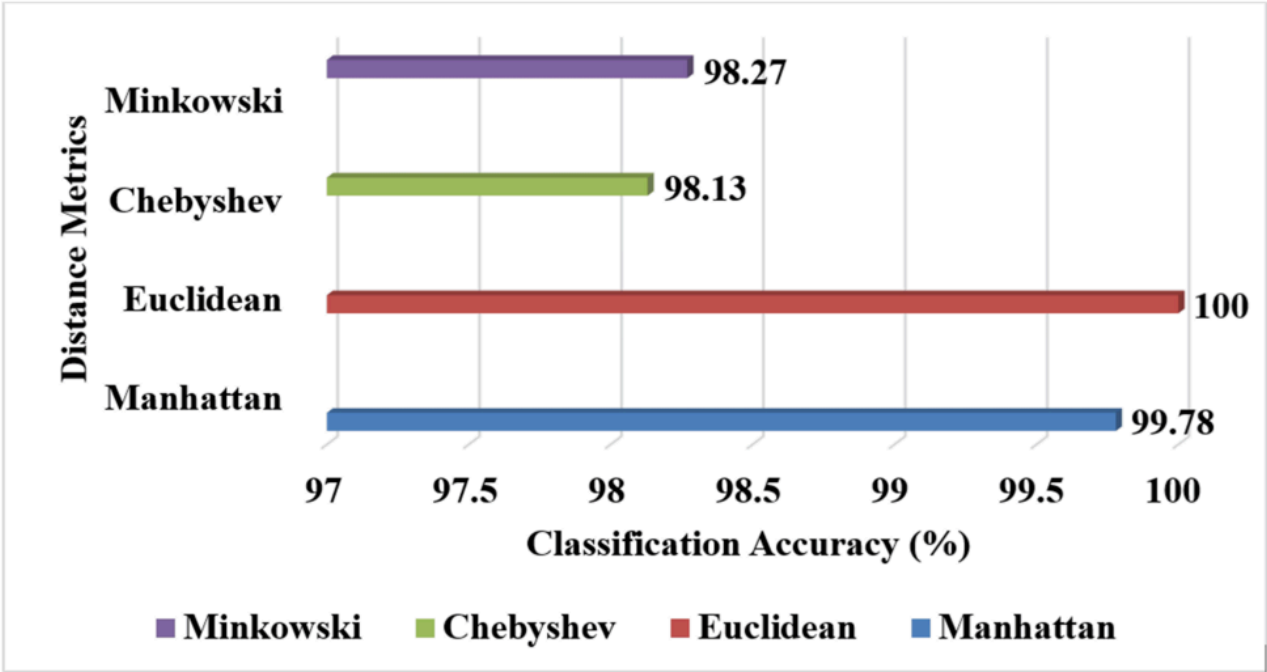


FIGURE 8: Performance evaluation of KNN distance metrics on color and luminance features

KNN, K-Nearest Neighbors

In the case of texture-based features, the performance shows a different trend (Figure 9). The Chebyshev distance achieves the highest accuracy of 98.74%, while Manhattan and Minkowski distances both achieve 97.74%, and Euclidean distance records 97.34%. This suggests that texture features derived from GLCM are sensitive to the choice of distance metric, and Chebyshev distance is more effective in capturing maximum variation in texture patterns. The small gap between training and testing accuracy across all metrics suggests stable model performance with good generalization.

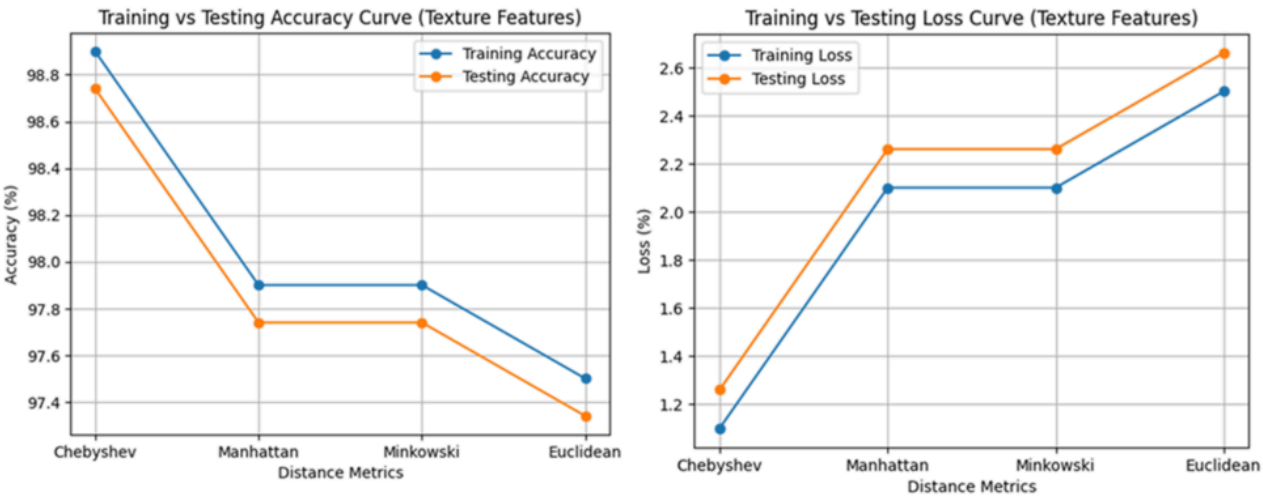


FIGURE 9: Training vs testing accuracy and loss curve

For the combined feature set (color, luminance, and texture), the results indicate an overall improvement in classification performance. The Euclidean distance again achieves high accuracy, followed closely by Manhattan distance at 99.95%, Chebyshev distance at 99.84%, and Minkowski distance at 99.74%. This demonstrates that integrating multiple feature types enhances the discriminative power of the model and leads to more robust classification outcomes.

Overall, the results confirm that feature fusion significantly improves soil classification accuracy compared to individual feature sets. While Euclidean distance performs best for color-dominated features, Chebyshev distance shows superior performance for texture-based analysis. The combined feature approach provides the most consistent and reliable results across all metrics, highlighting the effectiveness of the proposed framework for real-world soil analysis applications.

The experimental results indicate that color and luminance descriptors contribute more significantly than texture features for the selected soil categories. This may be due to the visually distinguishable color characteristics of the collected soil samples. However, texture descriptors provide complementary structural information that improves robustness under certain feature combinations.

Comparison with existing methods

Traditional machine learning approaches such as KNN offer lower computational complexity and easier interpretability compared to deep learning models. However, modern convolutional neural network-based approaches may provide improved feature learning capability for large-scale datasets. Future work will include comparative evaluation using support vector machine and convolutional neural network architectures to further validate the proposed framework.

Limitations of the study

Although the proposed framework achieved high classification accuracy, the research has certain limitations. The dataset was collected from a limited geographical region and contains only four soil categories. In addition, augmented images may introduce similarity between training and testing samples if not carefully managed. The controlled imaging conditions may also contribute to higher classification accuracy. Future work will focus on collecting more heterogeneous soil datasets from multiple environmental conditions and validating the framework using independent benchmark datasets.

Conclusions

This paper presents a comprehensive framework for soil image feature extraction and classification by integrating RGB color, luminance, and GLCM-based texture features. The proposed approach effectively captures both the color and spatial characteristics of soil, leading to high classification performance using the KNN algorithm. Experimental results demonstrate that the appropriate selection of distance metrics significantly improves accuracy, with strong performance achieved across individual and combined feature sets. The minimal gap between training and testing results confirms good generalization capability. Overall, the framework demonstrates promising performance for soil image classification and has potential applications in precision agriculture and soil monitoring systems. Future work will focus on larger benchmark datasets, advanced deep learning models, real-time deployment, and multimodal soil analysis integrating spectral and environmental information.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Girish D. Chate, S. S. Bhamare

Acquisition, analysis, or interpretation of data: Girish D. Chate, S. S. Bhamare

Drafting of the manuscript: Girish D. Chate, S. S. Bhamare

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. *Cureus J Comput Sci* 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

Critical review of the manuscript for important intellectual content: S. S. Bhamare

Supervision: S. S. Bhamare

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Motwani A, Patil P, Nagaria V, Verma S, Ghane S: [Soil analysis and crop recommendation using machine learning](#). 2022 International Conference for Advancement in Technology (ICONAT), Goa, India. 2022, 1-7. [10.1109/ICONAT53423.2022.9725901](#)
2. Pant J, Pant HK, Bhatt A, Pant D, Bhatt J: [Intelligent machine learning modelling for soil analysis and pH prediction](#). 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India. 2024, 1195-1198. [10.1109/ICCPCT61902.2024.10673053](#)
3. Jafar Sathick Ali M, Jane Rubel Angelina J, Sita Ramaiah K, Ravi Prakash C, Siva Naveen M: [An analytical approach for soil classification using image processing](#). International Research Journal on Advanced Engineering Hub (IRJAEH). 2024, 2:761-765. [10.47392/irjaeh.2024.0107](#)
4. Sivabalaselvamani D, Rahunathan L, Nanthini K, Harshini T, Hariprasath C: [Soil classification using deep learning techniques](#). 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAIS), Trichy, India. 2023, 582-586. [10.1109/ICAIS58487.2023.10250710](#)
5. Pham TH, Acharya P, Bachina S, Osterloh K, Nguyen KD: [Deep-learning framework for optimal selection of soil sampling sites](#). Computers and Electronics in Agriculture. 2024, 217:108650. [10.1016/j.compag.2024.108650](#)
6. Gyasi EK, Purushotham S: [Advancements in soil classification: an in-depth analysis of current deep learning techniques and emerging trends](#). Air, Soil and Water Research. 2023, 16:10.1177/11786221231214069
7. Sharma HK, Kumar S: [Soil classification & characterization using image processing](#). 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India. 2018, 885-890. [10.1109/ICCMC.2018.8488103](#)
8. Rahman SAZ, Mitra KC, Mohidul Islam SM: [Soil classification using machine learning methods and crop suggestion based on soil series](#). 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh. 2018-12, 1-4. [10.1109/ICCITECHN.2018.8631943](#)
9. Bhattacharya B, Solomatine DP: [Machine learning in soil classification](#). Neural Networks. 2006, 19:186-195. [10.1016/j.neunet.2006.01.005](#)
10. Krizhevsky A, Sutskever I, Hinton GE: [ImageNet classification with deep convolutional neural networks](#). Communications of the ACM. 2017, 60:84-90. [10.1145/3065386](#)
11. Shaha M, Pawar M: [Transfer learning for image classification](#). 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India. 2018, 656-660. [10.1109/ICECA.2018.8474802](#)
12. Sofou A, Evangelopoulos G, Maragos P: [Soil image segmentation and texture analysis: a computer vision approach](#). IEEE Geoscience and Remote Sensing Letters. 2005, 2:394-398. [10.1109/lgrs.2005.851752](#)
13. Shenbagavalli R, Ramar K: [Classification of soil textures based on Laws features extracted from preprocessing images on sequential and random windows](#). Bonfring International Journal of Advances in Image Processing. 2011, 1:15-18.

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>

[10.9756/bijaip.1004](https://doi.org/10.9756/bijaip.1004)

14. Zhang X, Younan NH, King RL: [Soil texture classification using wavelet transform and maximum likelihood approach](#). IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), Toulouse, France. 2003, 2888-2890. [10.1109/IGARSS.2003.1294621](https://doi.org/10.1109/IGARSS.2003.1294621)
15. Kiran Pandiri DN, Murugan R, Goel T: [Smart soil image classification system using lightweight convolutional neural network](#). Expert Systems with Applications. 2024, 238:122185. [10.1016/j.eswa.2023.122185](https://doi.org/10.1016/j.eswa.2023.122185)

How to cite this article:

Chate G D, Bhamare S S (May 18, 2026) Comprehensive Soil Image Features Extraction Using Color, Luminance, and Gray-Level Co-Occurrence Matrix Analysis. Cureus J Comput Sci 3 : es44389-026-00078-y. DOI <https://doi.org/10.7759/s44389-026-00078-y>