Nonlinear Functional Analysis and Applications
Vol. 31, No. 2 (2026), pp. 613-636
ISSN: 1229-1595 (print), 2466-0973 (online)

https://doi.org/10.22771/nfaa.2026.31.02.20
http://nfaa.kyungnam.ac.kr/journal-nfaa

APPLICATIONS OF BREGMAN DISTANCE FUNCTION TO IMPROVE THE PERFORMANCE OF THE PLANT LEAF IMAGE RETRIEVAL SYSTEM

B. H. Phong[1] and P. N. Anh[2]

[1]Faculty of Information Technology, Hanoi Architectural University, Hanoi, Vietnam
e-mail: phongbh@hau.edu.vn

[2]Lab of Applied Mathematics and Computing,
Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
e-mail: anhpn@ptit.edu.vn

Abstract. The retrieval of plant leaf images has shown a crucial role in the smart agriculture. The accurate and high performance development of plant leaf image retrieval systems allows agricultural users to search relevant results. In this paper, we propose a new approach to improve the performance of the plant leaf image retrieval system via using Bregman distance function. *Firstly*, we apply various advanced feature extraction techniques to plant leaf images. *Next*, using the Bregman distance is proposed to retrieve the images with high performance. The experimental results demonstrate that combining feature extraction methods with the Bregman distance significantly enhances the performance of the image retrieval system. We have evaluated the performance of the proposed method on two public datasets of plant leaf images. The obtained accuracy for the retrieval of plant leaf images and the time complexity analysis have shown the effectiveness of our proposed method.

1. Introduction

In recent years, image retrieval has become a well-known computer vision issue that focuses on finding images that are similar to an input query from a large image database [4, 5, 25]. There exist two types of the development methods of image retrieval systems [14]. The first one is the text-based image

[0]Received December 19, 2025. Revised March 28, 2026. Accepted April 16, 2026.
[0]2020 Mathematics Subject Classification: 46E35, 49K20, 76D05, 35Q30.
[0]Keywords: Image retrieval, deep neural networks, feature extraction, distance metric, plant leaf classification.
[0]Corresponding author: P. N. Anh(anhpn@ptit.edu.vn).

retrieval and the second one is the "content-based image retrieval" (CBIR). Developing the text-based image retrieval systems requires much time and human-efforts [19]. The annotation of a large amount of text for the text-based image retrieval is a time-consuming task. The performance of text-based image retrieval highly depends on the specific languages. Therefore, the CBIR approaches have been considered to provide the advantages compared to the text-based image retrieval ones [2].

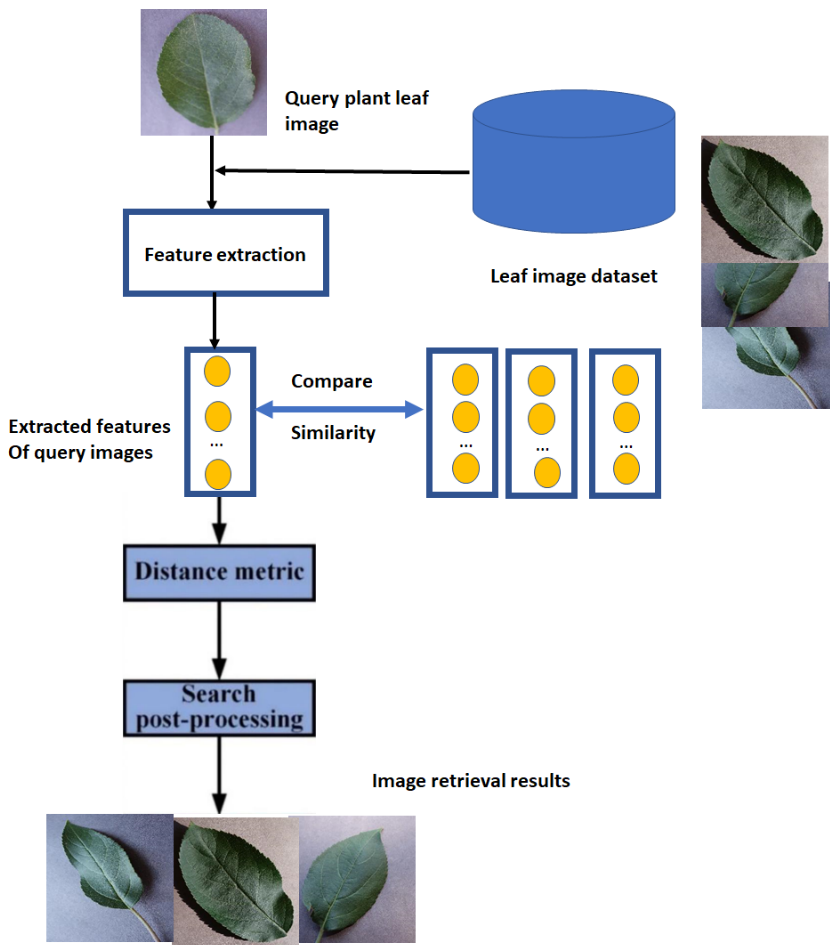

Figure 1. Flowchart of a plant leaf image retrieval system.

The development of a CBIR system normally consists of two phases. The first phase extracts visual features of images. The second phase computes the differences in features of images to rank the results [2]. In recent years, the image retrieval of plant leaves have shown many real applications in smart agriculture such as sustainable development, green production and environmental protection [8]. In reality, users may input query plant leaf images to find the species of plants. Retrieval systems can return the most similar images of a specific species related to searching images. Fig.1 shows the key steps of the development of a plant leaf image retrieval system. The image retrieval systems help reduce the search time for users.

Although advanced progress has been achieved for the task over the past years, the task still has several difficulties. The low quality of the query images may cause the wrong retrieval. In addition, retrieving images from a large database may consume much time. Therefore, the feature extraction of leaf images and the efficient distance calculation between the extracted features need to be further improved for the retrieval systems. Traditional methods extracted low-level features of images for the retrieval purposes [9]. In recent years, deep neural networks (DNN) have shown the leading performance for the feature extraction [27]. DNNs are able to extract a large number of visual features of images automatically. After the feature extraction, the distance metrics (e.g., Euclidean, Cosine) are widely applied to compute the similarity between features of images. The main drawback of existing methods is the high time complexity in retrieving images in a large number of images [21, 28].

In the paper, we propose the combination of the Bregman distance [34] with various feature extraction methods including handcrafted and deep neural network feature extraction to improve the performance of plant leaf image retrieval system. To the best of our knowledge, the solution has been proposed for the information retrieval of plant leaf images for the first time. Specifically, the integration of Bregman distance into the CBIR framework is motivated by its theoretical advantages in capturing non-Euclidean relationships and distribution-sensitive similarities, which are not effectively handled by traditional metrics such as Euclidean and Cosine distances. Unlike these conventional measures, Bregman divergence provides a more flexible formulation that can better model the underlying structure of feature spaces, particularly for high-dimensional and heterogeneous image representations. The proposed method allows us to achieve high accuracy and reduce the complexity in retrieving plant leaf images. In addition, we provide a comprehensive analysis of how Bregman distance affects similarity ranking behavior, which, to the best of our knowledge, has not been thoroughly explored in prior CBIR studies.

The contributions of the paper are threefold.

- Both hand-made and deep learning feature extraction methods are applied to extract features of leaf images efficiently. The handcrafted method extracts visual features such as the shape, color of leaf images with high performance. Whereas, the Alexnet [16, 17] and Resnet[11] neural networks are employed to extract a large number of features of leaf images automatically.

- After the feature extraction, we propose to use the Bregman metric to compute the similarity between the images with high performance. Existed methods widely applied the Euclidean and Cosine distance metrics [14, 27] to compute the similarity between images. Compared to existed methods, the Bregman metric reduces the computational time for the retrieval of images.
- We evaluated the performance of the proposed method on the public datasets of plant leaf images and compared with various existing methods to analyze the strengths and weaknesses of the proposed method.

The remainder of the paper is structured as follows. Section 2 reviews the significant approaches for image retrieval. Section 3 describes our proposed framework for the retrieval of leaf images. The experimental settings and numerical results are shown in Section 4. The last section concludes our contributions and draws some future work.

2. Literature review

The section reviews and analyzes a numerous approaches for the development of image retrieval systems. The image retrieval can be performed using the content-based and text-based techniques [25, 31]. The content-based retrieval systems are based on the feature extraction of images. After that, the similarity between images are determined using the distance metrics. Meanwhile, text-based retrieval systems apply sequential searching techniques [8]. The text-based retrieval systems require a lot of time and human-effort to insert text and meta-data for images [25]. In the paper, we focus on the retrieval of leaf images using the content-based searching method.

2.1. **Image retrieval systems.** The development of image retrieval systems have attracted a large amount of research in the past twenty years [31, 8]. Basically, the existing methods have proposed the different feature extraction methods of images. Then, various distance metrics have been applied to compute the similarity scores between the query image and related images in the database.

The traditional work in [26] proposed a color and shape extraction methods in 2D dimensional space to develop the image retrieval system. The work in [39] proposed the generalized biased discriminant analysis for the retrieval of images. The work in [37] fuses the color and shape features to develop the retrieval of images.

The work in [38] firstly applied the Daubechies wavelet transformation to represent the features of images. Then, the multiple support vector machines (SVM) is proposed to improve the performance of the retrieval of images. The work in [36] extracts the color and texture features of images. After that, the particle swarm optimization (PSO) is combined with K-means clustering algorithm to improve the accuracy of the retrieval of images.

The work in [29] analyzes and compares a huge number of parameters of deep neural networks that have been applied for the retrieval of remote sensing images. To reduce the number of extracted features of remote sensing images, the approximate nearest neighbor searching algorithm is applied in the work.

The recent research in [21] applied the Alextnet neural network to extract rich features of images. The principal component analysis (PCA) is applied to extracted features to obtain the principle features of images. Then, the Euclidean distance is used to calculate the similarity scores between the query images and images stored in the database.

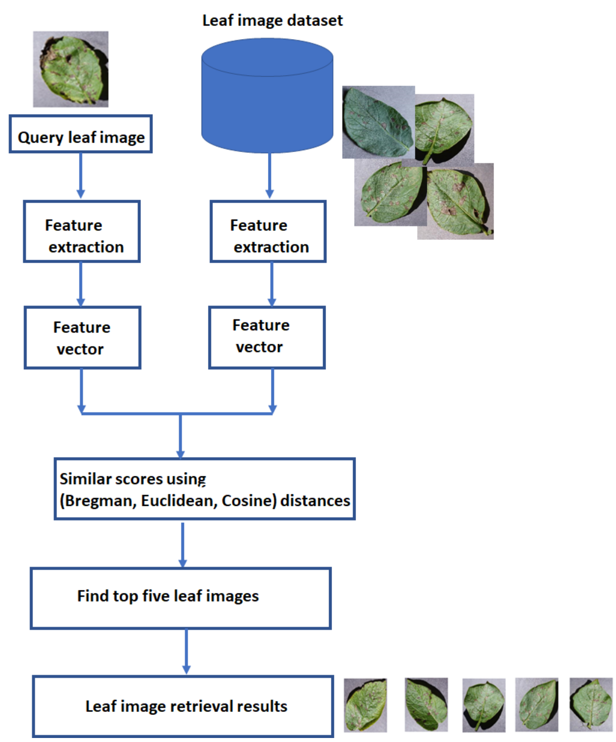

FIGURE 2. Flowchart of the proposed plant leaf image retrieval system using feature extraction and distance metrics.

2.2. **Retrieval of plant leaf images.** The early work in [32] extracts the shape features of leaf images. Then, the distance curve is applied for the retrieval system. After that, the work in [24] develops a mobile application to retrieve information of plant leaf images. Firstly, the visual features (e.g., color, shape and texture) of mobile-captured images are extracted. Then, the K-Nearest Neighbor algorithm is applied to search images.

The work in [1] extracts handcrafted visual features of foliage plant leaves. Then, the work applies various distance metrics (e.g., Euclidean, Canberra, Bray-Curtis distances) to compute the similarity scores of extracted feature vectors. Based on the computed similarity scores, the work finds top five images that are most similar with the query image. The method in [7] proposes approximate nearest neighbor search to retrieve images in a large set.

In recent years, the Deep Neural Networks (DNNs) have been widely applied to extract features of images efficiently. The work in [23] applied deep neural networkssuch as VGG-16, InceptionV3, Xception, and ResNet50along with similarity metrics like Manhattan and Euclidean distances to develop an image retrieval system. The work in [40] proposes an introspective deep metric learning (IDML) framework for uncertainty-aware image comparisons. However, the method requires labeled data and additional training complexity.

The work in [10] applied the convolutional neural network (CNN) to extract features of leaf images. The fully connected and softmax layers are applied to classify leaf images. Then, the optimized mean average precision values are used for the retrieval of images. The work in [25] applies several DNNs to develop an image retrieval system of plant leaf images. Firstly, the YOLOv5 network is applied to detect leaf regions in images. Then, the Resnet-50 backbone is applied to extract features of detected leaf images. Finally, the Cosine and Euclidean distances are applied to find similar images.

As in the above analysis of existing methods for the development of retrieval systems for plant leaf images, the high time complexity and the low accuracy of retrieval systems need to be further investigated. In the paper, the application of Bregman distance and various feature extractions have been proposed to improve the performance and reduce the time complexity of the retrieval system.

3. Proposed system for the retrieval of plant leaf images

The proposed system is described in Fig.2. The system consists of three main modules:

(1) A large number of plant leaf images are collected for the development of the retrieval system. The plant leaf images of potatoes and apples are collected

from public resources [12]. Then, the query images and collected images in database are processed and normalized. The input images are resized for the feature extraction by DNNs.

(2) The feature extraction of images are performed using the handcrafted and deep neural networks. After the feature extraction, we obtain the feature vectors of the query image and collected images in database. The handcrafted feature extraction obtains the shape, color, skewness and vein features [1] of plant leaf images. Moreover, in the work, we apply the Alexnet [17], VGG-16 [16] and Resnet-50 [11] neural networks to extract features of images automatically.

(3) Various distance metrics (e.g., Euclidean, Cosine and Bregman distance) are applied to compute the similar scores between the query and images in database. Based on obtained scores, the system returns the most similar images. In the paper, we propose to employ the Bregman distance to obtain the higher performance of the leaf retrieval compared to other distance metrics. The proposal of efficient distance allows us to improve the performance of the system.

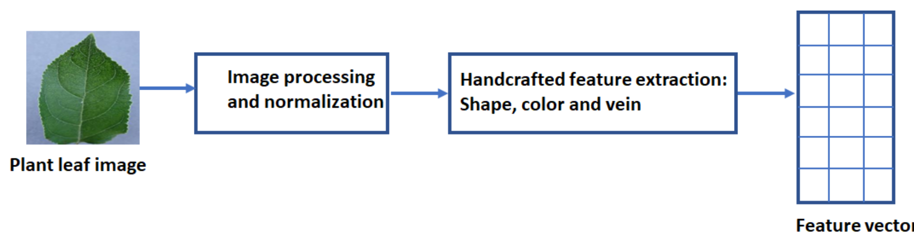

Figure 3. Steps of handcrafted feature extraction of leaf images.

3.1. **Feature extraction techniques.** Feature extraction is a crucial step for the accurate retrieval of images. Features are the specific signatures that represent and describe images. In the section, we apply two approaches to extract image features. The first one is the handcrafted feature extraction. The second one is using the deep neural networks. In general, the handcrafted feature extraction techniques are designed to work with high speed. In contrast, deep neural networks have been applied to extract a large number of features automatically. The deep neural networks obtain higher accuracy for the feature extraction [30]. Deep neural networks require a large volume of data to train the models efficiently.

3.1.1. *Handcrafted feature extraction techniques.* Handcrafted feature extraction methods are designed to extract low-level visual features of plant images

[33]. For leave images, the shape, color and vein features describe accurately the images. Given a searching image X of size $m \times n$, we compute the following features of the image: shape, color, vein and skewness features [1]. Fig.3 illustrates the steps of the handcrafted feature extraction of leaf images.

(1) Shape features of leaf images.

Shape features of leaf images represent the physical appearance of leaf images. Fore the identification of leaves, shape features consists of the following features:

The eccentricity feature aims to extract the elongated or stretched shapes of images. The feature is often represented as the ratio of the major axis length to the minor axis length of the bounding ellipse of shapes. The feature is calculated as follows:

$$E = \frac{d}{l}, \tag{3.1}$$

where d is the length of the leaf minor axis and l is the length of the leaf major axis.

The roundness feature represents the perimeter of the leaf. The feature is calculated as follows:

$$R = \frac{s}{r^2}, \tag{3.2}$$

where s is the area of the leaf and r is the perimeter of the leaf.

(2) Vein features of leaf images.

The vein features represent the morphological opening of leaf images [8], [33]. The features can be considered as the fingerprint of the plant leaf. The features are important characteristics to discriminate species of plants [18]. Therefore, the features are widely extracted to discriminate the leaves. An example of the vein extraction of a leaf image is shown in Fig.4.

(a) Apple leaf image (b) Gray leaf image (c) Vein of the leaf image

FIGURE 4. Examples of the vein extraction of a leaf image. Color leaf image (a) is converted to the gray (b) image and the vein of the image is extracted (c).

The vein features are computed as follows:

$$V_1 = \frac{p_1}{p}, \tag{3.3}$$

$$V_2 = \frac{p_2}{p}, \tag{3.4}$$

where V_1 and V_2 represent features of the vein of leaf images, p_1 and p_2, are total pixels of the veins, and p denotes total pixels on the part of the input leaf image.

(3) Color features of leaf images.

Color features allow us to extract the differences between colors in leaf images. Color features are key properties to discriminate plant leaf images [25]. For each plant leaf image, the mean, standard deviation, and skewness color features are extracted. The color features of an image P of size $m \times n$ are computed as follows.

The mean color feature (μ) is described as follows:

$$\mu = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij}. \tag{3.5}$$

The standard deviation (σ) color feature is described as follows:

$$\sigma = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (P_{ij} - \mu)^2. \tag{3.6}$$

The skewness (θ) color feature is described as follows:

$$\theta = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (P_{ij} - \mu)^3}{M \times N \times \sigma^3}, \tag{3.7}$$

where P_{ij} is the value of pixel at the positions i and j. Table 1 shows examples of the calculation of the extracted features in a leaf image.

Table 1. Example of a color feature vector of an image that has shown in Fig. 4.

Color	Mean	Median	Standard deviation	Skewness
Red	0.58	0.58	0.15	0.20
Green	0.52	0.49	0.12	0.18
Blue	0.56	0.55	0.12	0.05

Seven features of shape, color and vein of leaf images are extracted for leaf images using the handcrafted feature extraction method.

Algorithm 1 Plant leaf image retrieval algorithm using handcrafted feature extraction and Bregman distance

---

Input a query image X and a dataset of n image Y_i
Output top k images in the dataset that have the most similar scores with the query image X
$F \leftarrow Handcrafted\_Feature\_Extraction(X)$ ▷ Feature extraction of image X using the handcrafted feature extraction techniques
for $i \leftarrow 1$ to n **do**
    $G\_i \leftarrow Handcrafted\_Feature\_Extraction(Y\_i)$ ▷ Feature extraction of image Y_i using the handcrafted feature extraction techniques
    $R\_i \leftarrow Distance(G\_i, F)$ ▷ Calculate the distance between feature F vector of image X and G_i vector of image Y_i using the Bregman distance
end for
$R \leftarrow Rank(R)$ ▷ Rank obtained distance scores in vector R
$k \leftarrow SelctTop\_k(R)$ ▷ Select the most k lowest scores of vector R
Return k most similar images with the query image X in the dataset

---

3.1.2. *Feature extraction using the deep neural networks.* Given a searching image X of size $m \times n$, we apply the following deep neural network to extract the following features of the image. Input leaf images are resized at [224x224x3] to be suitable for the deep neural networks. Fig.5 shows the process of feature extraction using CNNs.

The Alexnet network [17] extracts 4096 visual features automatically using the *fc7* layer. Input images are resized of [227x227x3]. The structure of the Alexnet network is shown in Table 2. The network consists of eight layers.

VGG-16 network [16] consists of 16 layers. The network consists of 13 convolutional layers and 3 fully connected layers. Input images are resized of [224x224x3]. The network extracts 4096 features of images. The features are extracted at fully connected layer *fc7*. Structure of the network is described in Table 3.

Layers
Convolution 1
Maxpooling
Convolution 2
Maxpooling
Convolution 3
Convolution 4
Convolution 5
Maxpooling
Classification layer
Total: 8 layers

Table 2. Structure of the Alexnet neural network.

Layers
Convolution 1_1
Convolution 1_2
Maxpooling
Convolution 2_1
Convolution 2_2
Maxpooling
Convolution 3_1
Convolution 3_2
Convolution 3_3
Maxpooling
Convolution 4_1
Convolution 4_2
Convolution 4_3
Maxpooling
Convolution 5_1
Convolution 5_2
Convolution 5_3
Maxpooling
Fully connected layer 6
Fully connected layer 7
Total: 16 layers

Table 3. Structure of the VGG-16 neural network.

The Resnet-50 [11] network extracts 2048 visual features automatically. Input images are resized of [224x224x3]. The extracted features are obtained at the output of the global pooling layer *pool5* of the network. The network extracts 2048 features of input images. The structure of the Resnet-50 network is shown in Table 4. The network consists of 50 layers.

Table 4. Structure of the Resnet-50 network.

Layers
Block 1
Block 2
Block 3
Block 4
Block 5
Classification layer
Total: 50 layers

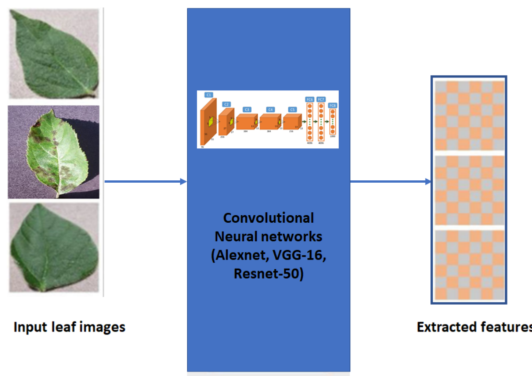

Figure 5. Flowchart of feature extraction of images using CNNs.

To extract features of leaf images efficiently, the DNNs are trained with collected leaf images. The training process of the DNNs is an optimization [35] problem. The problem can be defined as follows:

$$\min_{w} \mathcal{L}(w) = \frac{1}{n} \sum_{i=1}^{n} h\left(f\left(x_i, w\right), y_i\right). \tag{3.8}$$

We let $h\left(f\left(x_i, w\right), y_i\right)$ denote the loss function of the DNNs. We let $f\left(x_i, w\right)$ denote the output of the network. We let $\{x_i, y_i\}$ denote the training samples of each iteration of the training of the DNNs.

In the paper, the cross-entropy loss function [22] is applied to train the DNNs efficiently. Moreover, the stochastic gradient descent (SGD) optimization algorithm [15] is applied for the DNNs. The (SGD) optimization algorithm is commonly used for the DNNs to obtain the high performance during the training process [15].

During the training of the DNNs to extract features of leaf images, the gradient descent and stochastic gradient descent (SGD) algorithm [15] is applied to minimize the loss values. The momentum of the SGD algorithm is set as 0.9. Input leaf images are resized as the requirement of the DNNs. The initial learning rate is set as 0.001. The implementation of the DNNs is supported by the Python environment with the Keras library. The networks were trained from scratch using two plant leaf datasets. The training from scratch allows the model to learn features tailored specifically to leaf patterns and textures.

3.2. **The Bregman distance operator.** Let C be a nonempty, convex and closed subset of $\mathbb{R}^n$. Recall that the proximal operator of a proper lower semicontinuous and convex function $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with respect to the set C is formulated as follows:

$$Prox_C^g(u) = \operatorname{argmin}\left\{ g(x) + \frac{1}{2}\|x-u\|^2 : x \in \mathbb{R}^n \right\}, \quad \forall u \in C.$$

In the particular case when $g = \iota_C$, where ι_C is the indicator function of the set C, then the function $Prox_C^{\iota_C}$ is the metric projection onto C.

Definition 3.1. Let C be a closed subset in $\mathbb{R}^n$ and $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function such that $C \cap \operatorname{dom} g \neq \emptyset$. Then the function g is said to be prox-convex on C with prox-convex value $\gamma > 0$ if $Prox_C^g(u) \neq \emptyset$ for every $u \in C$, and for $\bar{x} \in Prox_C^g(u)$, we have

$$g(\bar{x}) - g(x) \leq \gamma\langle \bar{x} - u, x - \bar{x}\rangle, \ \ \forall x \in C.$$

Prox-convex functions, introduced recently for algorithm purposes, have called the attention of several authors, specially from nonconvex mixed variational inequalities and its applications.

Definition 3.2. Let $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function and $h: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a strongly convex and differentiable function such that $\operatorname{intdom}(h) \neq \emptyset$. The Bregman distance operator of g with respect to a set $\emptyset \neq K \subset \mathbb{R}^n$ is the operator given by

$$\operatorname{Prox}_{g+\delta_K}^h(u) := \arg\min\{g(x) + \delta_K(x) + D_h(x,u) : x \in \mathbb{R}^n\}, \ \forall u \in \operatorname{int}(\operatorname{dom}(h)).$$

Theorem 3.3. *Let K be a closed convex subset in $\mathbb{R}^n$. If g is Bregman-prox-convex on K with respect to h and $\gamma > 0$, then the function $\frac{1}{\alpha}g$ is also Bregman-prox-convex on K with respect to h and 1. Moreover, we have*

$$\operatorname{Prox}_{\frac{1}{\alpha}g+\delta_K}^h = \operatorname{Prox}_{g+\delta_K}^h.$$

Proof. Since g is (h,γ)-prox-convex on K, by definition, for every $u \in K$ and any $x \in K$, letting $\bar{x}(u) = \operatorname{Prox}_{g+\delta_K}^h(u)$, we have

$$g(\bar{x}(u)) - g(x) \leq \gamma\langle \nabla h(u) - \nabla h(\bar{x}(u)), \bar{x}(u) - x\rangle. \tag{3.9}$$

It follows that

$$\frac{1}{\gamma}[g(\bar{x}(u)) - g(x)] \leq \frac{1}{\gamma}\gamma\langle \nabla h(u) - \nabla h(\bar{x}(u)), \bar{x}(u) - x\rangle$$
$$= \langle \nabla h(u) - \nabla h(\bar{x}(u)), \bar{x}(u) - x\rangle,$$

which means that $\frac{1}{\gamma}g$ is Bregman-prox-convex on K with respect to h and 1.

Let $\bar{x}(u) = \mathrm{Prox}^h_{g+\delta_K}(u)$. Using (3.9) and the identity

$$\langle \nabla h(x) - \nabla h(y), y - z\rangle = D_h(z, x) - D_h(z, y) - D_h(y, x)$$

with $x = u$, $y = \bar{x}(u)$, $z = x$, we obtain

$$g(\bar{x}(u)) - g(x) \leq \gamma\big(D_h(x, u) - D_h(x, \bar{x}(u)) - D_h(\bar{x}(u), u)\big).$$

Dividing both sides by γ gives

$$\frac{1}{\gamma}\big(g(\bar{x}(u)) - g(x)\big) \leq D_h(x, u) - D_h(x, \bar{x}(u)) - D_h(\bar{x}(u), u),$$

which implies that

$$\frac{1}{\gamma}g(\bar{x}(u)) + D_h(x, \bar{x}(u)) \leq \frac{1}{\gamma}g(x) + D_h(x, u) - D_h(\bar{x}(u), u) \leq \frac{1}{\gamma}g(x) + D_h(x, u)$$

for all $x \in K$. Therefore, $\bar{x}(u) = \mathrm{Prox}^h_{\tilde{g}+\delta_K}(u)$. This completes the proof. □

3.3. **Distance calculation using the Bregman distance operator.** The Bregman distance metric has shown the powerful performance of the unsupervised machine learning [6] . The distance [34] is defined as follows:

$$d(x, y) = \begin{cases} \frac{1}{2}\|x - y\|^2 + \beta \sum_{i=1}^n y_i^2 \left(\frac{x_i}{y_i}\log\frac{x_i}{y_i} - \frac{x_i}{y_i} + 1\right), & \text{if } x > 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{3.10}$$

where $\beta \in (0, 1)$. We set $\beta = 0.9$ for the calculation between X and Y sets of features of images. x and y are the elements in these feature vectors.

Bregman distance introduces a fundamentally different mechanism for measuring similarity compared to conventional metrics such as Euclidean and Cosine distances, which directly impacts ranking behavior in CBIR systems. Unlike metric-based distances, Bregman divergence is defined with respect to a convex function and measures the discrepancy between a point and the first-order approximation of that function. As a result, it is inherently asymmetric and distribution-aware, enabling it to better capture the geometry of feature spaces that deviate from isotropic assumptions.

The algorithm of the retrieval of plant leaf images using the Bregman distance and the handcrafted feature extraction is described in the Algorithm 1. The algorithm of retrieval of plant leaf images using the Bregman distance and deep neural networks is described in Algorithm 2.

Algorithm 2 Plant leaf image retrieval algorithm using deep learning feature extraction and Bregman distance

Input a query image X and a testing dataset of n image Y_i
Output top k images in the dataset that have the most similar scores with the query image X
$F \leftarrow DeepLearning\_Feature\_Extraction(X)$ ▷ Feature extraction of image X using the trained Resnet-50
for $i \leftarrow 1$ to n **do**
  $G\_i \leftarrow DeepLearning\_Feature\_Extraction(Y\_i)$ ▷ Feature extraction of image Y_i using the trained Resnet-50
  $R\_i \leftarrow Distance(G\_i, F)$ ▷ Calculate the distance between feature F vector of image X and G_i vector of image Y_i using the Bregman distance
end for
$R \leftarrow Rank(R)$ ▷ Rank obtained distance scores in vector R
$k \leftarrow SelctTop\_k(R)$ ▷ Select the most k lowest scores of vector R
Return k most similar images with the query image X in the testing dataset

In the algorithms, all distance measures operate in $\mathcal{O}(d)$ time per comparison, where d is the feature dimensionality.

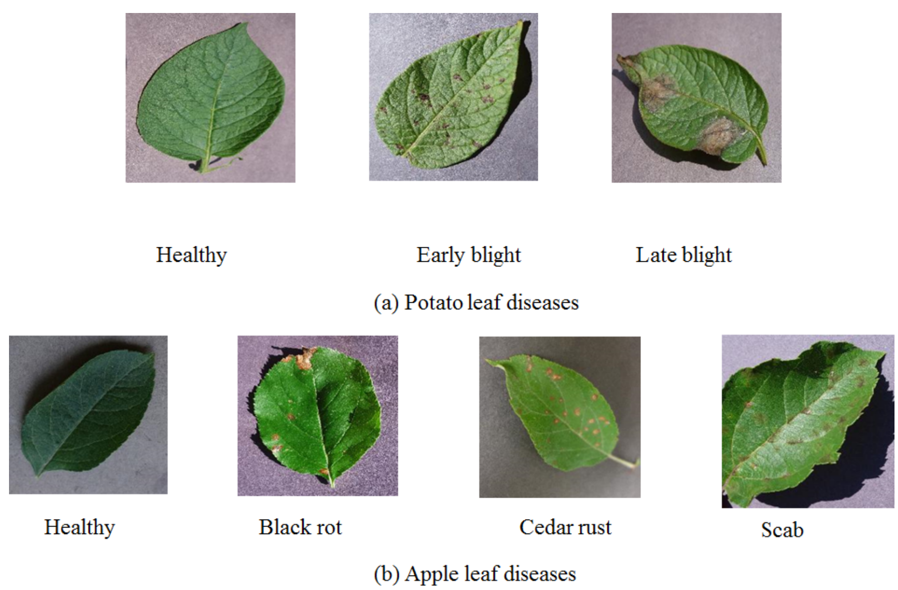

FIGURE 6. Examples of potato and apple plant leaf images.

4. Experimental results

4.1. **Datasets and evaluation metric.** We evaluate the proposed system on two public datasets. The datasets consist of apple and potato leaf images [12]. Information of the datasets is shown in Table 6. The dataset contains three classes of potato leaf images: late blight, early blight, and healthy, with 1000 images in each class. In contrast, the apple leaf dataset comprises four classes. The four classes of potato leaves are: black rot, scab, cedar rust and healthy. Each class of apple leaf consists of 1600 images. Eighty percent of the images are used to train the DNN models, while the remaining images are used to test the trained models. Fig.6 shows examples of plant leaf diseases images in the datasets.

To evaluate the proposed method, the accuracy metric [25] that is commonly used for the image retrieval is adopted. As in traditional evaluation methods [14], the retrieval system selected the top one, three, and five images with the highest similarity scores to the query image.

The accuracy metric of the retrieval system can be calculated as follows:

$$\text{Accuracy} = \frac{\text{Correct\_retrieved\_images}}{\text{Total\_retrieved\_images}}. \tag{4.1}$$

Moreover, the paper applies the Mean Average Precision (mAP) that is a popular metric for evaluating the image retrieval system.

4.2. **Performance evaluation.** The performance comparison of the searching images using Euclidean, Cosine and Bregman distances are shown in table 7 and 8. The accuracy of the searching method using Euclidean distance is better than that of Bregman. Examples of similar leaf images are shown in Fig.7. Table 5 illustrates the similarity scores using the distances for the plant leaf image retrieval that are shown in Fig.7. The similarity score obtained by the Euclidean distance is better than that of Bregman and Cosine distances. However, the execution time of the method using the Bregman distance is better than methods using Euclidean and Cosine distances. Moreover, the method using the deep neural network has a higher accuracy compared to those using the handcrafted feature extraction techniques. The obtained accuracy of the method on the apple leaves is higher compared to that on the potato leaves because the image quality of apple leaves is better. The resolution of apple leaf images is higher than that of potato leaf images. Table 7 and 8 show the accuracy of the retrieval method on the apple and potato leaf images, respectively. Table 9 shows the comparison of execution time between the methods using the Bregman, Euclidean and Cosine distances.

Compared to the handcrafted feature extraction method, the leaf retrieval using the deep neural networks obtains significant higher accuracy. The higher accuracy of the deep neural networks because the DNNs extract more feature extractions than handcrafted feature extraction method. The performance of the Resnet-50 is higher compared to Alexnet because the Resnet-50 consists of more layers that allows the network extract features more efficiently than Alexnet.

Fig.8 illustrates examples of the query leaf image and the top five retrieved images. The proposed system is developed in the computer that consists of the 8GB of RAM memory and Core-i5 processor. The programming language used in the paper is Python v3.7 with Tensorflow library.

Table 5. Examples of similarity scores of two plant leaf images that are shown in Fig.7 using the Euclidean, Cosine and Bregman distances.

Retrieval method	Distance scores
Using Resnet and Euclidean distance	0.111
Using Resnet and Bregman distance	0.118
Using Resnet and Cosine distance	0.128

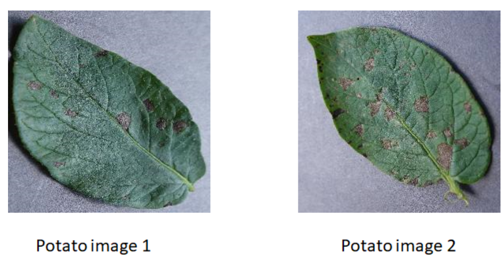

Figure 7. Examples of two similar potato leaf images.

Table 6. Statistic information of apple and potato leaf disease datasets [12].

Dataset	Number of images
Potato (early blight)	1000
Potato (late blight)	1000
Potato (healthy)	1000
Apple (black rot)	1600
Apple (black scab)	1600
Apple (cedar rust)	1600
Apple (healthy)	1600

Table 7. Accuracy of the proposed retrieval methods on the apple plant leaves. The highest scores are in bold.

Method	top 1	top 3	top 5	mAP
Handcrafted feature, Euclidean metric	**70%**	**76%**	**78%**	**74.67%**
Handcrafted feature, Bregman metric	51%	56%	58%	55.00%
Handcrafted feature, Cosine metric	49%	54%	55%	52.67%
Alexnet and Euclidean metric	79%	82%	**91%**	**84.00%**
Alexnet and Bregman metric	**81%**	**83%**	84%	82.67%
Alexnet and Cosine metric	75%	78%	79%	77.33%
VGG-16 and Euclidean metric	82%	83%	**92%**	**85.67%**
VGG-16 and Bregman metric	**83%**	**84%**	85%	84.00%
VGG-16 and Cosine metric	70%	78%	80%	76.00%
Resnet and Euclidean metric	80%	83%	**94%**	85.67%
Resnet and Bregman metric	**81%**	**84%**	85%	83.33%
Resnet and Cosine metric	71%	78%	81%	76.67%

Table 8. Accuracy of the proposed retrieval methods on the potato plant leaves. The highest scores are in bold.

Method	top 1	top 3	top 5	mAP
Handcrafted feature, Euclidean metric	**72%**	**77%**	**79%**	**76.00%**
Handcrafted feature, Bregman metric	55%	58%	60%	57.67%
Handcrafted feature, Cosine metric	51%	55%	57%	54.33%
Alexnet and Euclidean metric	81%	83%	91.50%	**85.17%**
Alexnet and Bregman metric	**82%**	**84%**	84.50%	83.50%
Alexnet and Cosine metric	75%	78%	80%	77.67%
VGG-16 and Euclidean metric	80%	83%	93%	**85.33%**
VGG-16 and Bregman metric	**82%**	**85%**	86%	84.33%
VGG-16 and Cosine metric	73%	78%	81%	77.33%
Resnet and Euclidean metric	83%	**86%**	95%	**88.00%**
Resnet and Bregman metric	**84%**	85%	86%	85.00%
Resnet and Cosine metric	72%	77%	82%	77.00%

Table 9. Execution time (in milliseconds) of the proposed methods for the retrieval plant leaf images. The highest scores are in bold.

Model	**Execution time**
Handcrafted feature and Euclidean metric	30
Handcrafted feature and Bregman metric	**28**
Handcrafted feature and Cosine metric	34
Alexnet and Euclidean metric	41
Alexnet and Bregman metric	**38**
Alex and Cosine metric	43
VGG-16 and Euclidean metric	42
VGG-16 and Bregman metric	40
VGG-16 and Cosine metric	44
Resnet and Euclidean metric	45
Resnet and Bregman metric	43
Resnet and Cosine metric	47

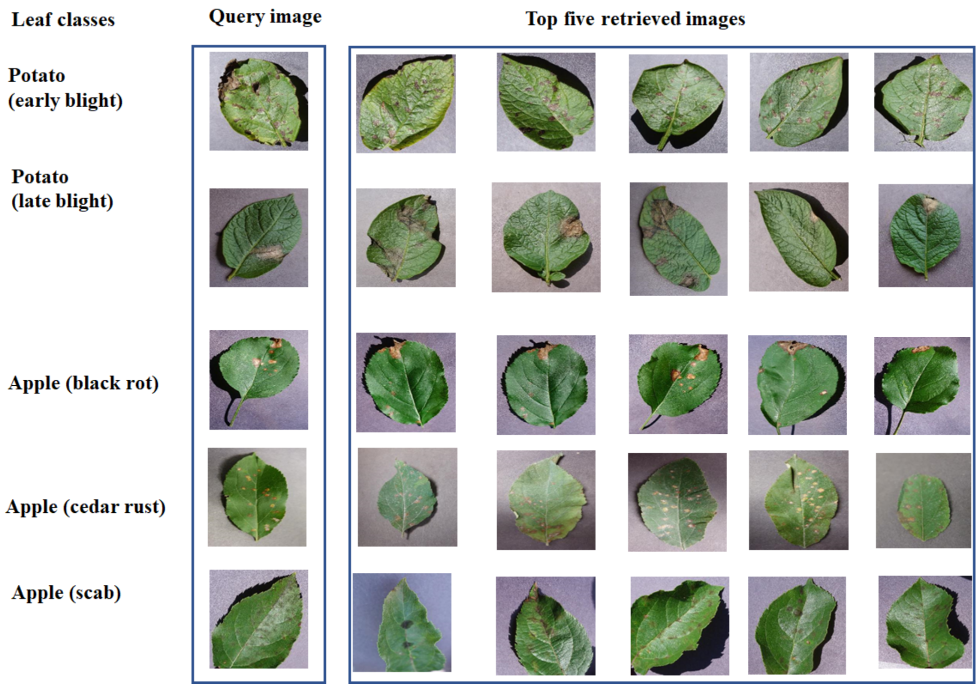

FIGURE 8. Examples of the query leaf image and top five retrieved results.

TABLE 10. Performance comparison between the proposed and existing methods on apple plant leaves. The highest scores are in bold.

Method	Accuracy
Retrieval method using the Inceptionnet V3 [10]	68 %
Using the Alexnet and Euclidean distance [27]	79 %
Using the handcrafted method and Euclidean distance[1]	64 %
Using the deep neural network and Cosine distance[13]	65 %
Proposed method using Bregman metric	**86 %**

TABLE 11. Performance comparison between the proposed and existing methods on potato plant leaves. The highest scores are in bold.

Method	Accuracy
Retrieval method using the Inceptionnet V3 [10]	67 %
Using the Alexnet and Euclidean distance [27]	77 %
Using the handcrafted method and Euclidean distance[1]	64 %
Using the deep neural network and Cosine distance[13]	63.5 %
Proposed method using Bregman metric	**85 %**

Table 12. Comparison of execution time (in milliseconds) between the proposed and existing methods. The highest scores are in bold.

Method	Time
Retrieval method using the Inceptionnet V3 [10]	38
Using the Alexnet and Euclidean distance [27]	40
Using the handcrafted method and Euclidean distance[1]	32
Using the deep neural network and Cosine distance[13]	55
Proposed method using Bregman metric	**35**
Proposed method using Bregman metric	**28**

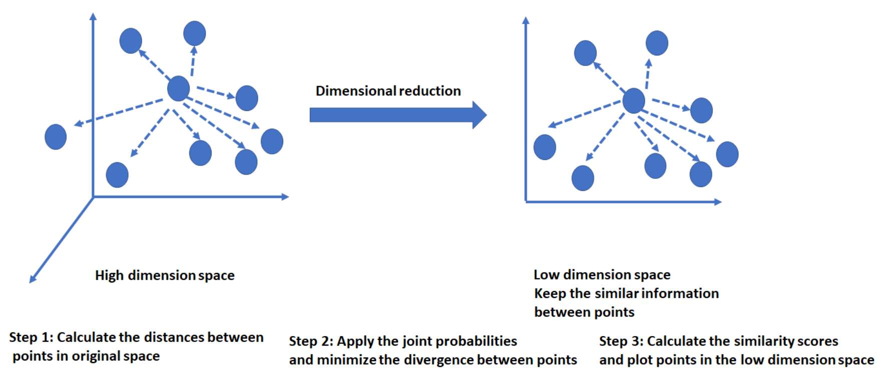

Figure 9. The t-SNE illustration of the reduction process from high to low dimension [20].

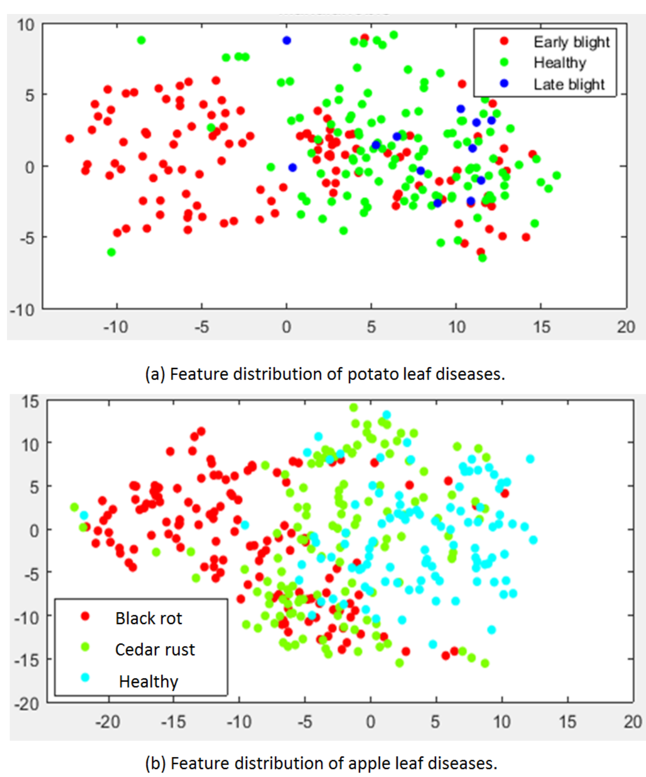

Figure 10. The demonstration of the extracted features using the Resnet-50 and the t-SNE. Features of potato (a) and apple diseases (b) are illustrated in red, green and blue circles, respectively.

4.3. **Performance comparison of the proposed with existing methods.** To deeply analyze the strengths and weaknesses of the proposed method, we compare the obtained results of the proposed method with existing methods. The proposed method obtained higher accuracy compared to existing methods because the Resnet-50 extracts features of plant leaf images efficiently. Moreover, the use of Bregman distance reduces the computational time compared with existing method. The results are shown in table 11 and 12, respectively. Other works applied the Euclidean and Cosine [13] distance metrics, therefore, the time complexity is not as efficient as our proposed method. Moreover, our method applied the Resnet-50 network that has a complex architecture to extract features of leaf images accurately. The work in [10] applied the InceptionNet V3 network to extract features of images. The Resnet-50 shows better performance than InceptionNet V3 network for the feature extraction purpose.

The work in [27] applied the Euclidean distance, therefore, the speed of the work is not high. The work in [1] applied the handcrafted feature extraction and Euclidean distance, therefore, the accuracy is lower than that of methods using the DNNs and Bregman distance. Notably, the Bregman metric achieves the highest scores for top-1 and top-3 image retrieval. The results have shown that the proposed system is not only correct, but also efficient and practical for real-world use.

4.4. **Visualization of feature extraction using the dimensional reduction techniques.** To demonstrate the efficiency of the features extracted by the DNNs, we visualize the extracted features using the dimensional reduction. The t-distributed stochastic neighbor embedding t-SNE [20] is applied to map high dimensional data points in the low dimensional space. The distances between the points are respected when we display the points in lower dimensional space. Alexnet and Resnet-50 extract 4096 and 2048 features of images, respectively. Using the dimensional reduction techniques, we reduce the extracted number of features to 2. Fig 10 demonstrates the distribution of features extracted by DNNs of potato (a) and apple diseases (b). The visualization shows the efficient feature extraction of leaf images of the DNNs.

The t-SNE is an unsupervised non-linear dimensionality reduction technique. The technique is powerful for data exploration and visualization. Nonlinear dimensionality reduction refers to the algorithms that separate data that cannot be separated by a straight line. Fig. 9 demonstrates the main idea of the t-SNE technique.

PCA (Principal Component Analysis) [3] is a linear technique that is suitable for data that have a linear structure. Both t-SNE and PCA are dimensional reduction techniques that have shown effectiveness for data visualization. PCA reduces linear dimensionality of data by identifying orthogonal axes of maximum variance. In contrast, t-SNE, keeps local relationships and emphasizes non-linear structures of the data. Compared to the PCA technique, the t-SNE obtains better efficiency for the visualization of extracted features of images using DNNs.

5. Conclusions and future works

The paper presents a method that applied feature extraction techniques and distance metrics to improve the performance of image retrieval system for plant leaves. The Bregman allows to improve the execution time compared to Euclidean and Cosine metrics in term of execution time. In term of accuracy, the Euclidean metric allows to obtain better performance. The feature extraction using DNNs obtains higher accuracy of handcrafted feature extraction. The comparison with existing methods shows that using ResNet-50 and Bregman distance achieves high accuracy in leaf image retrieval. In the future, the parallel computing can be applied to improve the processing time for the retrieval of a large number of images. Moreover, we can apply the metric learning to automatically select the suitable one for the image retrieval systems.

References

[1] K.L. Abdul and U.G. Mada, *Experiments of distance measurements in a foliage plant retrieval system,* Int. J. Signal Processing, Image Processing and Pattern Recognition , **5**(2) (2012), 47–60.

[2] F. Ali and A. Hashem, *Content based image retrieval (cbir) by statistical methods,* Baghdad Sci. J., **17** (2020).

[3] E. Ali, M.A. Hossain and M.R. Islam, *Analysis of pca based feature extraction methods for classification of hyperspectral image*, 2nd International Conference 23 on Innovation in Engineering and Technology (ICIET) (2019).

[4] P.N. Anh, *Relaxed projection methods for solving variational inequality problems,* J. Global Opti., **90** (2024), 909-930.

[5] P.N. Anh, A. Gibali and N.D. Truong, *PIPA - A parallel inertial proximal algorithm with applications to image recovery problems,* J. Nonlinear Convex Anal., **25**(11) (204), 2913–2931.

[6] A. Banerjee, S. Merugu, I.S. Dhillon and J. Ghosh, *Clustering with bregman divergences,* J. Machine Learning Res., **6** (2005), 1705-1749.

[7] J. Bastiaan, P.X.H. Boom and R.B. Fisher, *Approximate nearest neighbor search to support manual image annotation of large domain-specific datasets,* Conference: Proc. Int. Workshop on Video and Image Ground Truth in Comput. Vision Appl. At: St. Petersburg, Russia **4**(2013), 1–8.

[8] H.H. Bauschke and P.L. Combettes, *A review on content-based image retrieval system: present trends and future challenges,* Int. J. Comput. Vision and Robotics, **11**(5) (2021).

[9] Y. Censor, A. Gibali and S. Reich, *Texture features for browsing and retrieval of image data,* IEEE Tran. Pattern Anal. Machine Intel., **18**(8) (1996), 837–842.

[10] B.P. Gyires-Toth, D.P.M. Osvath and C. Szucs, *Deep learning for plant classification and content-based image retrieval,* Cybernetics and Information Technologies, **19**(1) (2019).

[11] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition,* Proc. IEEE Conferencece on computer vision and pattern recognition, (2026), 770-778.

[12] D.P. Hughes and M. Salath, *An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing,* Comput. Sci., Envir. Sci., Agricu. Food Sci., arxiv (2015).

[13] B. Hung and P. Phuong, *Content based image retrieval based on deep learning approach. Computer Vision and Robotics,* Proc. CVR **2021** (2022).

[14] M. Ibtihaal, S.H. Hameed and B.M. Mahmmod, *Content-based image retrieval: A review ofrecent trends,* Numer. Algor. **8**(1) (2021).

[15] I. Iqbal, C.A. Odesanmi and L. Liu, *Comparative investigation of learning algorithms for image classification with small dataset,* Appl. Artificial Intelligence, **35**(10) (2021), 697-716.

[16] C. Izuchukwu, I. Shehu and J.C. Yao, *Very deep convolutional networks for large-scale image recognition,* arXiv:1409.1556, (2014).

[17] A. Krizhevsky, I. Sutskever and G.E. Hinton, *Imagenet classification with deep convolutional neural.networks,* Adv. Neural Info. Processing Syst., **60**(6) (2012), 84-90.

[18] L. Li, W. Hu, J. Lu and C. Zhang, *Leaf vein segmentation with self-supervision,* Comput. Elect. Agricul., **203** (2022).

[19] Y. Liu, D. Zhang, G. Lu and W.Y. Ma, *A survey of content-based image retrieval with high-level semantics,* Pattern Recognit, **40** (2007), 262–282.

[20] L. Maaten, G.E. Hinton, *Visualizing data using t-sne*, Journal of Machine Learning Research, **9** (2008), 2579-2605.

[21] S. Mahmoud, A. Ahmed, K.A. Fathy and K.R. Kamal, *A deep learning content-based image retrieval approach using cloud computing,* Indonesian J. Elect. Eng. Comput. Sci., **29**(3) (2023), 1577–1589.

[22] A. Mao, M. Mohri and Y. Zhong, *Cross-entropy loss functions: theoretical analysis and applications,* Proc. the 40th Int. Confe. Machine Learning, Cornell Univ., **13** (2023), arXiv:2304.07288.

[23] A. Mathew, N. Sejal and R.K Venugopal, *Analysis of content based image retrieval using deep feature extraction and similarity matching*, Int. J. Adv. Comput. Sci., Appl., **13**(12) (2022), 646–655.

[24] Y. Nam, E. Hwang and D. Kim, *Clover: A mobile content-based leaf image retrieval system,* ICADL 2005. Lecture Notes in Comput. Sci. **3815** (2005), 139–148.

[25] Y. Peng and Y. Wang, *Leaf disease image retrieval with object detection and deep metric learning,* Front. Plant Sci., **13** (2022), /doi.org/10.3389/fpls.2022.963302.

[26] B.G. Prasad, S.K. Gupta and K.K. Biswas, *Color and shape index for region-based image retrieval*, Lecture Notes in Comput. Sci., **2059** (2001), 716–729.

[27] A. Qayyum, S.M. Anwar, M. Awais and M. Majid, *Medical image retrieval using deep convolutional neural network*, Neurocomputing, **266** (2017), 8–12.

[28] G. Sumbul, J. Kang and B. Demir, *Deep Learning for Image Search and Retrieval in Large Remote Sensing Archives,* arXiv:2004.01613 (2020).

[29] G. Sumbul, J. Kang and B. Demir, *Deep Learning for Image Search and Retrieval in Large Remote Sensing Archives,* Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences, Wiley, (2021), 150–160.

[30] E. Tsalera, A. Papadakis, M. Samarakou and I. Voyiatzis, *Feature extraction with hand-crafted methods and convolutional neural networks for facial emotion recognition,* Appl. Sci., **12**(17) (2022).

[31] R. Vishraj, S. Gupta and S. Singh, *A comprehensive review of content-based image retrieval systems using deep learning and hand-crafted features in medical imaging, Research challenges and future directions*, Comput. Elect. Eng., **104** (2022), 108450.

[32] Z. Wang, Z. Chi, D. Fend and Q. Wang, *Leaf image retrieval with shape features,* VISUAL 2000, LNCS, **1929** (2003), 477-487.

[33] S.G. Wu, F.S. Bao, E.Y. Xu, Y.X. Wang, Y.F. Chang and Q.L. Xiang, *A leaf recognition algorithm for plant classification using probabilistic neural network,* IEEE Int. Symp. Signal Proces. Inform. Tech., (2007), 10.1109/ISSPIT.2007.4458016.

[34] L. Wu, C.H. Hoi, R. Jin, J. Zhu and N. Yu, *Learning Bregman Distance Functions and Its Application for Semi-Supervised Clustering*, IEEE Transactions on Knowledge and Data Engineering (TKDE) Systems **24**(3) (2012), 478–491.

[35] Y. Xu and H. Zhang, *Convergence of deep convolutional neural networks,* Neural Networks, **153** (2022), 553-563.

[36] Z.D. Younus, D. Mohamad, T. Saba, M.H. Mlkawaz, A. Rehman, M. AlRodhaan and A. Al-Dhelaan, *Content-based image retrieval using pso and k-means clustering algorithm,* Arab. J. Geosci., **8**(8) (2015), 6211-6224.

[37] L.J. Yue and Z. Fu, *Content-based image retrieval using color and texture fused features,* Math. Comput. Model, **54**(3) (2011), 1121-1127.

[38] L.J. Yue and Z. Fu, *Efficient content-based image retrieval using multiple support vector machines ensemble*, Expert Syst. Appl., **39**(3) (2012), 2385-2396.

[39] L. Zhang, L. Wang and W. Lin, *Generalized biased discriminant analysis for content-based image retrieval,* IEEE Trans. Syst., **42**(1) (2011), 282-290.

[40] W. Zheng, C, Wang, J. Zhou and J. Lu, *Introspective deep metric learning for image retrieval,* Cornell Univ., arXiv2205.04449 (2023).