PROCEEDINGS OF SCIENCE

Nonlinear fusion of multiple efficient manifold rankings in content-based medical image retrieval

Tran Văn Huy [a], Dao Van Tuyet [b,*], Ngo Nguyen Khoi [c], Pham Thi Kim Dzung [d], Ngo Hoang Huy [e], Bui Dinh Chien [d] and Tran Cong Hung [b]

a Hong Duc University,
Thanh Hoa City, Vietnam

b Saigon International University,
Ho Chi Minh City, Viet Nam

c Conestoga College – Cambridge,

Ontario, Canada

d FPT University,
Ha Noi City, Viet Nam

e CMC University,
Ha Noi City, Viet Nam

E-mail: tranlehuy@hdu.edu.vn, daovantuyet@siu.edu.vn, ngonguyenkhoi98@gmail.com,dungptk8@fe.edu.vn, nhhuy@cmc-u.edu.vn, chienbd@fe.edu.vn, tranconghung@siu.edu.vn

The efficient manifold ranking (EMR) algorithm has been widely utilized in content-based image retrieval (CBIR). In this algorithm, each image is represented by low-level features that describe color, texture, and shape. However, low-level features have limitations in capturing semantic meaning. To enhance EMR performance in CBIR, this research proposes a fusion method called CoEMR.

CoEMR combines multi-rankings on low-level features with CNN features extracted from a CNN model to enhance the discriminative power of a query image compared to dataset images. Furthermore, CoEMR generates a similarity score between two input images, constructing a similarity learning model. Experiments demonstrate the effectiveness of the proposed methods in improving EMR quality. Additionally, the potential integration of CBIR with Large Language Models in Medical Image Diagnosis Systems is discussed.

Key words: Content-based medical image retrieval, Medical image, Efficient manifold ranking, Deep Metric Learning, Contrastive loss, Triplet loss, EMR learning, LLMs, Semantic similarity

International Symposium on Grids and Clouds (ISGC2024)
24 -29 March, 2024
Academia Sinica Computing Centre (ASGC), Institute of Physics, Academia Sinica
Taipei, Taiwan

PoS(ISGC2024)002

1. Introduction

In recent times, in the context of economic and social issues, the demand for disease diagnosis through image and deep disease recognition in crop plants in the field of agriculture has been increasing. Therefore, the classification and search for diseases based on content-based image retrieval (CBIR) techniques play an important role in practice [1-2].

In a CBIR system, images are often represented by feature vectors such as low-level features [6]. In addition, exploiting the excellent image discriminative features of CNN-based classification models, in [7-8], the authors used pre-trained networks to extract features (CNN) that achieved very high accuracy in image classification problems in the trained domain. However, CNN models often require a large amount of training data for the network to adapt to changes in image orientation and structure. Therefore, when deploying practical CBIR systems, we often need to combine CNN features and low-level features (color, texture, and shape descriptors) on image datasets that lack diversity in orientation and structure.

In image recognition and retrieval, in general, similarity measures such as Euclidean distance are not suitable for measuring the similarity between images because many related images can be different even in visual appearance. In complex cases, feature vectors of images are often considered as data points in some manifold to apply similarity measures or distances based on manifold approximation. Yang [9] has pointed out the existence of relationships between manifold learning and distance metric learning.

Regarding manifold data, manifold ranking (MR) [10] is known as a graph-based model that has been successfully applied to content-based image retrieval (CBIR) using low-level features [11-13] and CNN features. MR is very effective in measuring the similarity between query images and image datasets where the relevance is not easily detected by visual appearance but has a "latent" nature, only detected when similarity spreads through a graph structure [16]. In this article, we propose two new methods to rank image databases based on query images. The first is a nonlinear combination method of multiple manifold ranking results, and the second is a manifold ranking distance metric learning method to exploit the advantages of manifold ranking with the property of propagating similarity on a graph that represents the local geometric structure of the feature vector set representing images.

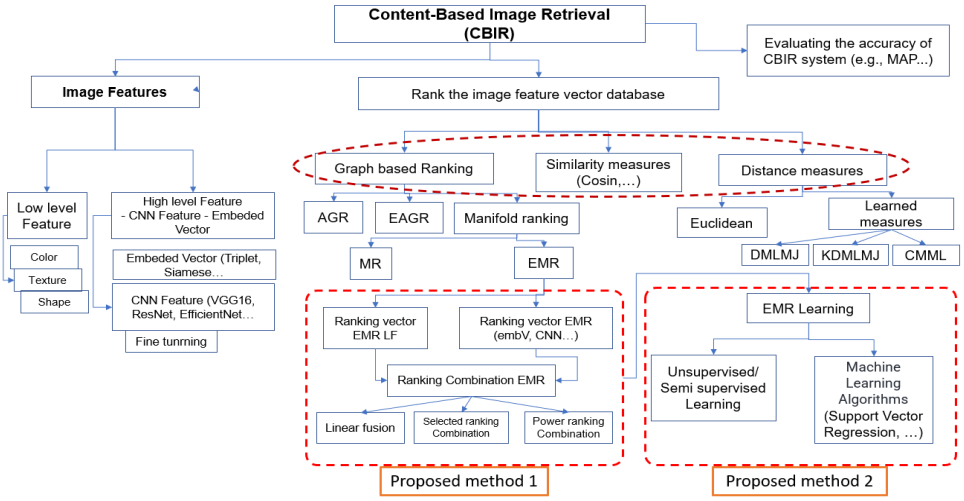

Figure 1. Overview of the research problem.

The remaining part of the paper is structured as follows. Section II presents the related research and the proposed methods. Section III presents the experiments on the Leaf2k and COVID-19 chest X-ray datasets. The discussion is in Section IV and the conclusion is in Section V.

2. Related work and proposed method

2.1 Metric learning

In Content-Based Image Retrieval (CBIR) systems, both similarity and dissimilarity measures are crucial for the quality of search results. Machine learning algorithms are known to be sensitive to these measures [3]. Traditional metrics like Euclidean distance often fail to fully represent the semantic complexity inherent in some problems. In response, various metric learning algorithms have been developed to better grasp these complex semantic spaces [4]. Among them, Mahalanobis distance metric learning stands out as a popular choice [4], effectively transforming the Euclidean distance to better fit the data structure.

For metric learning to be effective, it should ensure that similar items are measured as closely together while dissimilar items are farther apart. A good metric learning method should meet three criteria: it should accurately reflect the true similarities between items, be computationally efficient, and be versatile enough to handle various data types and learning scenarios.

Standard metric learning algorithms, often referred to as single or global metric learning algorithms, include Neighbourhood Component Analysis (NCA), Large Margin Nearest Neighbor (LMNN), and others. Although effective for simple data structures, they struggle with complex, nonlinear datasets. To overcome this, strategies such as kernel methods, deep embeddings, and multiple metric learning have been proposed [20]. Kernel methods can separate data effectively in a high-dimensional space, but they are computationally demanding and choosing the right kernel is challenging. Deep embeddings offer strong feature representation yet lack interpretability and require considerable preprocessing.

For a detailed overview of distance metric learning, references [20] are recommended, with examples like LMNN, ITML, and NCA. With the advent of deep learning, new algorithms such as FaceNet [5] and energy-based neural networks [20] have emerged. These deep metric learning algorithms have taken advantage of traditional methods to improve performance.

In classification applications, algorithms specifically designed to enhance the K-NN classifier have gained attention due to better performance. These can be broadly categorized into two groups. The first includes triplet-based methods, which rely on prior information to construct triplets. While effective, these methods can be overly restrictive, narrowing the learned metric's search space and potentially complicating training. They also face challenges with the large number of triplets required, which can be computationally prohibitive [20].

The second group does not require prior information about sample neighborhoods. An example is NCA, which aims to minimize the expected leave-one-out training error of the K-NN classifier. However, these methods might overlook valuable discriminative information because they tend to focus on the nearest neighbors at the expense of those further away.

Considering distance metric learning as an optimization of the empirical risk for the K-NN classifier, there's a gap between traditional loss functions and the K-NN's empirical risk. This has motivated the development of algorithms that target the empirical risk directly. Nonetheless,

optimizing this risk is challenging due to the non-continuous nature of the K-NN decision function in relation to the distance metric, making it more complex than optimizing for SVMs, linear regression, or softmax regression [20].

2.2 Graph-based Ranking

Graph-based ranking models have been widely applied in content-based image retrieval (CBIR) query, including MR (Manifold Ranking), AGR (Anchor Graph Regulrization), EAGR (Efficient Anchor Graph Regulrization), EMR (Efficient Manifold Ranking), and more.

2.2.1 Anchor Graph Regulrization (AGR)

Anchor Graph Regularization (AGR) is a graph-based ranking model that utilizes anchor points to construct a graph representing the relationships between data points and anchor points [15]. Specifically, the effectiveness of AGR lies in two steps:

1. Building the adjacency relationships between points in the same class based on anchor points, instead of computing all pairwise relationships between data points.

2. Labeling data points based on the adjacency relationships between their respective classes and anchor points.

2.2.2 Efficient Manifold Ranking - EMR

Unlike MR algorithms in CBIR, the original EMR algorithm [10] focuses on selecting the number of anchor points and then ranking the data points in the image database without using label information. Ranking images based on the neighborhood relationships in the EMR graph is represented through these anchor points and maintains an expandable graph structure. Thus, EMR has the capability to efficiently rank large databases and thereby enhance the performance of CBIR systems [10].

A feature vector E_i is considered to be adjacent to E_j if $i \neq j$ and there exists a common anchor pointAc (c = 1, 2, ..., C; is the number of anchor points) such that A_c is a neighboring anchor point of both E_i and E_j. For each symbol E_i, we denote Nb(i; s) as a set consisting of s feature vectors of the nearest anchor points to Ei (s is a testing parameter, for example, s = 5), and the maximum distance between E_i and s is represented by:

$$d_s = \max_{l \in N_b(i,s)} \{d(E_i, A_l)\} \quad (1)$$

The multidimensional measure is a ranking vector $r_Q = (r_i)$ constructed by solving the objective function:

$$EMR_Q(r) = \frac{1}{2}\left(\sum_{i,j} w_{ij} \left\| \frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{ij}}} \right\|^2 + \mu \sum_i \left\| r_i - r_{0,i} \right\|^2 \right) \to \min \quad (2)$$

EMR is considered more effective than other methods. To demonstrate its effectiveness in classification and determining anchor points when constructing the EMR graph, we compare it with the ranking method based on the AGR graph, as shown in the experimental section.

2.3 Proposed Methods for Efficient Manifold Ranking

Image features can be divided into low-level and high-level features, where low-level features contain the image's characteristics such as color, texture, shape, etc. They are represented in a high-dimensional vector with less semantics but do not lose detailed image information. In

contrast, high-level image features can represent more semantics but are constrained by reduced color and texture details.

The following diagram illustrates the steps of the image retrieval algorithm that we propose:

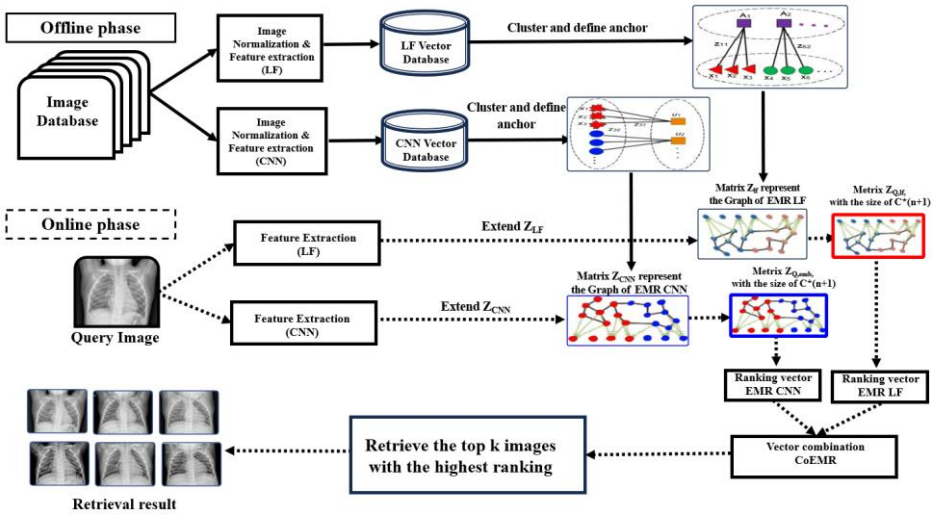

Figure 2. Fusing EMR Rankings in a CBIR System

Using high-level features represented by CNN feature vectors extracted from the EfficientNet deep learning network, querying images can yield relatively high accuracy results [16]. Assuming a dataset of original images, where each image has a corresponding unique feature vector, and for a query image IQ, the system selects rankings based on the low-level feature image ranking $r^*_{lf.v.Q,i}$ combined with the high-level feature image ranking $r^*_{CNN.Q,i}$, where the ranking vectors are calculated using EMR. The following is the detailed algorithm that we propose.

Algorithm. CoEMR (The algorithm combines two separate EMR rankings on low-level features and CNN).

Input: $\{I_i\}_{1\le i\le n}$ is the original image dataset, and IQ is the query image.

MxN the image size standardized for training using the EfficientNet.

d: the dimension of an embedded vector.

C: the number of anchors in the EMR algorithm, parameter $\in (0,1)$ (a $\approx$ 1),

α, β: To linearly combine the ranking weights of two EMR algorithms, α, β >0, α + β = 1.

Output: $r = \{r_i\}_{1\le i\le n}, r_i \in [0,1]\forall i = \overline{1,n}$ is the similarity ranking of image IQ with image Ii in the image database E.

Step 1 (offline): Build the EMR graph

Step 1a. Train the image dataset E using the EfficientNet models with the standardized input image size of M*N, resulting in:

1.1. Use the model parameter W_CNN.

1.2. Passing each image $\{I_i\}_{1\le i\le n}$ through the model yields a set of embedded vectors $\{v.E_i\}_{1\le i\le n}$ with dimension dCNN=1280.

1.3. Determine C anchor points $\{A_c\}_{1\le c\le C}$ for the image dataset $\{I_i\}_{1\le i\le n}$ based on an improved version of the FCM algorithm using the CNN vectors $\{v.E_i\}_{1\le i\le n}$

1.4. Determine the adjacency matrixW=(wij) for the EMR algorithm using the CNN vectors $\{v.E_i\}_{1\le i\le n}$.

1.5. Determine the weight matrix Z with dimensions C*n for EMR.

Step 1b. Low-level feature extraction: Color Moments, LBP, Gabor Wavelets Texture, Edge, and GIST are extracted, resulting in a feature vector set with dLF=809 dimensions. Repeat steps 1.3 - 1.5 as in step 1a with the low-level features.

Step 2 (online): Combine the ranking vectors

2.1. Normalize the image IQ về kích thước M*N, then pass it through the EfficientNet model defined in step 1.1, and obtain the CNN vector v.EQ with dimension d.

2.2. Expand the Z matrix according to EMR [7] (refer to formula (4) above): Using the C distance values between EQ and the anchor points in Ac, we obtain a new weight matrix ZQ with dimensions C*(n+1).

2.3. Set $r_Q = \{r_i\}_{1\le i\le n+1}, r_i = 0 \forall i = \overline{1,n}, r_{n+1} = 1.0$. From the weight matrix ZQ, we determine the ranking vector $r^*_{CNN.Q,i}$ using the EMR algorithm[7].

Step 3: Repeat steps 2.1 - 2.3 with low-level features lf.Ei we obtain the ranking values là $r^*_{lf.v.Q,i}$

Step 4: To combine the ranking vectors of the two EMR methods

- Linear fusion $CB(r_1,r_2) = \{\alpha r^*_{lf.v.Q,i} + \beta r^*_{CNN.Q,i}\}_{1\le i\le n}$.

- Selected ranking Combination: $CB(r_1,r_2) = \begin{cases} r_2 \\ r_1 \end{cases} \quad if\ r_2 \ge th$

- Power ranking Combination: $CB(r_1,r_2) = \begin{cases} \sqrt[3]{\frac{\{r_1^3+r_2^3\}}{2}} \\ min\ (r_1,r_2) \end{cases} \quad if\ r_1 + r_2 \ge 0$

Output: r is represents the similarity ranking of image IQ with image Ii in the image database E.

CoEMR consists of 2 phases, offline and online, performed by linearly combining 2 EMR rankings. Offline phase: O(C*n*d) + O(C3); Online phase: O(C*n*d), where C is the number of clusters, n is the number of samples, and d is the feature vector dimensionality.CoEMR is combines selected rankings of 2 EMRs, **CoEMR** complexity is: O(C*n*d) + O(C3 has the same computational complexity as the original EMR [10].

2.4 Building image similarity measure based on manifold ranking

EMR has been used to determine the similarity between any two vectors [11]. However, there is a significant limitation when applying EMR to images outside the dataset [10]. To address this issue, we propose integrating machine learning by training a Support Vector Machine Regression (SVR) model to learn the ranking results of EMR for each image pair in the training dataset.

Step 1. Defining the sample set of feature vector pairs:

The training image feature vector pairs (Iv1,Iv2) are selected as follows: Iv1 is chosen from a random υ% of images in the image feature vector database.

For each Iv1, we perform the following two steps:

+ Select Iv2 from the N_2 nearest images to Iv1 based on Euclidean distance.

+ Rank all feature vectors in the database with the query vector Iv1 using EMR.

We obtain the input-output pairs as: { (Iv1-Iv2, rankIv1(Iv2))}

Step 2. Splitting the sample set into training and testing sets.

Step 3. Training SVR regression on the training set, where the input is the difference vector of the two image feature vectors, and the output is the real-valued similarity score (according to EMR).

Step 4. Evaluating image similarity.

For an image pair (I_1, I_2) with corresponding feature vector pair (I_{V1}, I_{V2}), the learned machine model determines the similarity score $\Omega(I_{v1} - I_{v2})$.

- The computational complexity of the similarity calculation phase is the sum of the complexity of building EMR on n images and the complexity of training SVR. O(SVR) = O(n2*d). So: O(Ω) = O(C*n*d) + O(C3) + O(n2*d).

- The computational complexity of the similarity prediction phase is: O(nsv * d) wherensv is the number of support vectors obtained after training the SVR model.

We observe that O(nsv * d)<< O(C*n*d).

The main contributions of our proposal are as follows:

(1) Providing an effective combination of low-level features and CNN features, thereby increasing the accuracy of image query results in CBIR.

(2) Experimentally demonstrating on the Leaf2k and COVID-19 chest X-ray datasets that measuring similarity on CNN feature vectors using EMR is more effective than using AGR.

(3) Proposing the integration of machine learning through the SVR regression model to improve accuracy in the problem of evaluating image similarity using EMR.

3. Experiments

3.1 Image Datasets

To conduct the experimental part, we first select suitable datasets to demonstrate the rationality of the proposed parameters and algorithms in this article. The datasets should be large, complex, and unlabeled. Therefore, we have chosen the following two datasets:

3.1.1 Leaf2k Dataset

The Leaf2k dataset contains 2,600 leaf images of 10 plant species compiled from the PlantVillage, Leaf Disease, Leaf Images datasets. It is divided into 20 classes in order of healthy leaf images, diseased leaf images, with each class containing 130 images and a total size of 185MB. This dataset is temporarily named Leaf2k. The image names are assigned by the folder name with the sequential number of the image within the folder.

3.1.2 COVID-19 chest X-ray Dataset

The X-Ray Lung COVID-19 Image Dataset [17] consists of a database of lung X-ray images for COVID-19 positive cases, as well as normal images and viral pneumonia images. The dataset contains a total of 21,165 images, categorized into the following classes: Normal: 10,192 images, Lung_Opacity: 6,012 images, COVID-19: 3,616 images, Viral_Pneumonia: 1,345 images.

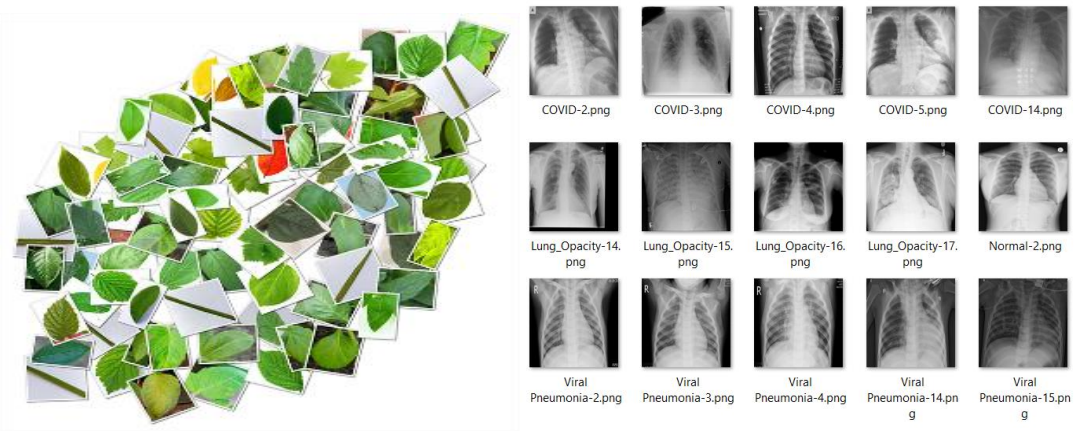

Figure 3: Some images from the Leaf2K (a) and the X-Ray Lung COVID-19 (b) image dataset.

3.2 Feature Extraction

The low-level features (LF) comprise five sets: Color Moments, LBP, Gabor Wavelets Texture, Edge, and GIST, to describe an image. All these features of the Leaf2k and COVID-19 chest X-ray datasets are normalized so that each vector component of each image falls within the range [-1, 1]. They are then concatenated into a vector with a size of dLF=809 [10].

In parallel, each image in these datasets is resized to 256x256 and passed through the EfficientNet [14] model (with the last layer removed), resulting in a corresponding CNN feature vector set with a size of dCNN=1280.

3.3 Evaluation Formula

3.3.1 The evaluation index for label recognition accuracy

The ERR (Error Rate Reduction) index was used to evaluate the results of image label recognition [21] as follows:

$$Accuracy(q;S,N,DS) = \begin{cases} 1: l_q = DS(argminN - top(S(q,E_i))) \\ 0 \end{cases}$$

$$Acc(S,N,DS) = \frac{1}{|Q|}\sum_{q=1}^{|Q|} Accuracy(q;S,N,DS).$$

$$ERR(S,N,DS) = 100 - 100 * Acc(A,N,DS) \quad (3)$$

3.3.2 The evaluation index for query accuracy

For each query image q∈Q, using the similarity scores provided by EMR, we chose N = 100 is the number of images in a class. The accuracy value is the average ratio between the number of relevant images N_q^+ within the retrieved images N and the similarity of each image q. Let the set of related elements to the query $q \in Q$ be $\{d_1, d_2, \ldots, d_{mj}\}$, then mAP(q) represents the mean Average Precision for a query q and the accuracy for all queries is calculated as follows: $mAP_{(q)} = \frac{N_q^+}{N} * 100$ (4), and $mAP = \frac{1}{|Q|}\sum_{q=1}^{|Q|} mAP_{(q)}$ (5)

The ranking results of EMR for each query image q can be considered as a similarity measure of the image, indicating the degree of similarity between the query image q and the image Ei in the database, which is assigned as rankq(Ei).

3.4 Experimental Results

We conducted several experiments including:

- Experiment 1: Comparing the classification results using the AGR and EMR algorithms.
- Experiment 2: Evaluating the accuracy in terms of mAP when retrieving images of leaf diseases and lung diseases using the CoEMR algorithm, which includes the following components:
  - Using EMR for low-level feature ranking.
  - Using EMR for CNN feature ranking.
  - Using the proposed method of combining EMR rankings.
- Experiment 3: Comparing the results of EMR with DMLMJ and CMML algorithms for the given image datasets.
- Experiment 4: Learning similarity measures.

3.4.1 Comparing EMR and AGR when utilizing label information from the image database for classification

We compare based on the ERR index, which is the average error rate of AGR and EMR on the Leaf2k and COVID-19 chest X-ray datasets. The experimental results in the table below demonstrate that EMR often achieves higher accuracy in many cases, highlighting the effectiveness of improving the modeling of relationships in the image retrieval process.

Table 1: Comparing the average error rate ERR (Error Rate Rate)

Index	**Method**	**Error Rate (%)**	
		Leaf2k (C=300, nb=100)	COVID-19 Chest X-ray (C= 300, nb=100)
1.	Anchor Graph Regulrization (AGR)	26.92±0.85	15.72±0.85
2.	Efficient Manifold Ranking (EMR)	13.5±0.71	9.28±0.50

3.4.2 Evaluation of accuracy based on mAP when querying images with the CoEMR algorithm

We set the common parameters for all experiments as follows: parameter a = 0.99, anchor points C = 200, r = 5, nbest for each query nb = 100, and randomly sampling 20% of the query samples. The parameters for combining the two rankings in CoEMR are $\alpha = 0.2$ and $\beta = 0.8$.

Experiment 1: Using EMR for low-level features.

Using EMR for low-level feature ranking: The results on the Leaf2k dataset achieved an accuracy of 65.05%, while on the COVID-19 chest X-ray dataset, the accuracy was 68.51%.

Experiment 2: Using EMR for CNN features.

Using EMR for CNN feature ranking: The results on the Leaf2k dataset achieved an accuracy of 70.72%, while on the COVID-19 chest X-ray dataset, the accuracy was 73.27%.

Evaluation: The results obtained from the experiments indicate that using CNN features is more effective than using low-level features. However, there are specific cases where the retrieval results are poor, even though the query image has a highly accurate CNN feature vector. The main reason for this is the difference between the pretraining data and the Leaf2k and COVID-19 chest X-ray datasets. This is a common challenge when applying Deep Learning in CBIR, as pretraining models are typically used.

Experiment 3: Using the proposed method of combining EMR rankings.

In this part, we used two EMRs to rank the low-level feature vectors and the CNN feature vectors extracted from EfficientNet, and combined these rankings using the proposed CoEMR algorithm. This achieved the highest accuracy results with a power ranking combination of 81.70% on the Leaf2k dataset and 82.81% on the COVID-19 chest X-ray dataset. The image query results for (a) Cherry_healthy_0076.jpg and (b) COVID_0302.png using the combined EMR ranking are depicted in the following image.

(a) (b)

Figure 4. The query results for (a) Cherry_healthy_0076.jpg; (b) COVID_0302.png using the EMR ranking fusion.

3.4.3 Comparing the results of EMR with DMLMJ and CMML algorithms for the given image datasets

We utilized DMLMJ and CMML on the Leaf2k and COVID-19 chest X-ray datasets, resulting in fairly good performance. Specifically, for the Leaf2k dataset, the DMLMJ accuracy was 57.91% and the CMML accuracy was 71.74%. For the COVID-19 chest X-ray dataset, the DMLMJ accuracy was 57.79% and the CMML accuracy was 70.26%.

We summarized the results from the experiments in Table 2 based on our arguments and the cases presented in this article. The results demonstrated that EMR performed well when applied individually and even better when combined on both the Leaf2k and COVID-19 chest X-ray datasets. The accuracy was measured using mAP and is as follows:

Table 2: The experimental results

Index	**Method**	**Leaf2k**	**COVID-19 chest X-ray**
		C=200, nb=100	C=200, nb =100
1	EMR for low-level features (Color Moments, LBP, Gabor Wavelets Texture, Edge, and GIST)	65.05%	68.51%
2	EMR for CNN features	70.72%	73.27%
3	DMLMJ	57.91%	57.79%
4	CMML	71.74%	70.26%
5	Combine EMR		
5.1	Linear fusion	79.87%	80.01%
5.2	Selected ranking Combination	80.28%	80.93%
5.3	Power ranking Combination	81.70%	82.81%

3.4.4 EMR learning

The experimental parameters were set as follows: $\upsilon = 50$, $N_2 = 20$. The set (Iv1, Iv2) divided into 80% randomly selected pairs for model training and 20% pairs of image feature vectors were used for training validation. The SVR model utilized the RBF kernel.

The test results on the training dataset showed a high correlation of 0.94, while on the testing dataset, the correlation was 0.92. These high correlation coefficients demonstrate the effectiveness of the unsupervised machine learning model for learning image similarity measures.

The SVR model can effectively learn to accurately predict the similarity between two images based on the deviation of their input feature vector pairs. Here, we assume that the images belong to the same domain as the images in the database, with the feature vectors equipped with a multi-dimensional manifold ranking.

4. Discussion

In the field of content-based image retrieval (CBIR), a major challenge is determining the semantic similarity between images. To improve search efficiency, we need to evaluate both visual content similarity and textual description similarity.

Methods for determining low-level and high-level visual feature similarity (CNN and Manifold Ranking) have shown good results thanks to the propagation mechanism of the Manifold Ranking algorithm. Using EMR or CoERM, we can determine the most similar image set for each query image, thereby obtaining the corresponding textual description set from the annotated image dataset. We can then utilize techniques like EMR learning and natural language processing to assess the semantic similarity between images to be compared.

With large language models (LLMs), we can estimate the semantic similarity between textual descriptions [24]. In the field of medical image diagnosis, LLMs can analyze information from medical records to assist doctors in diagnosis.

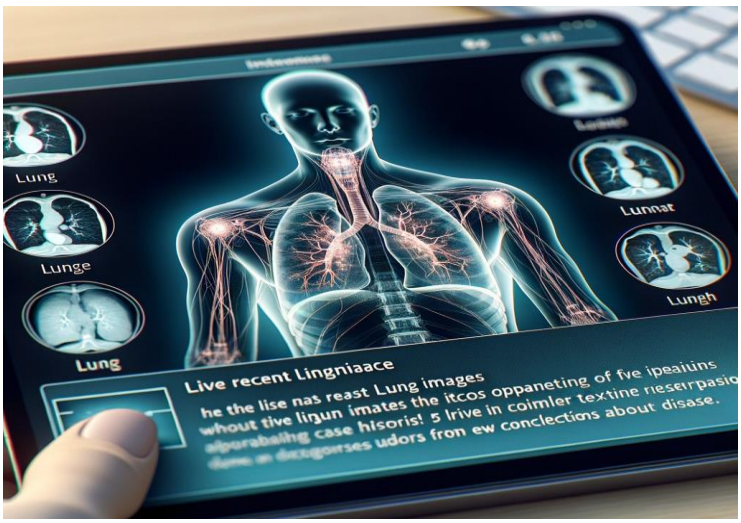

Figure 5. An application for disease diagnosis using CBIR combined with LLMs.

Therefore, to comprehensively evaluate the semantic relationship between images, we can combine tools for measuring visual similarity from EMR and textual similarity from LLMs. This

integrated approach allows a more complete and accurate assessment of semantic relationships (see Fig.6).

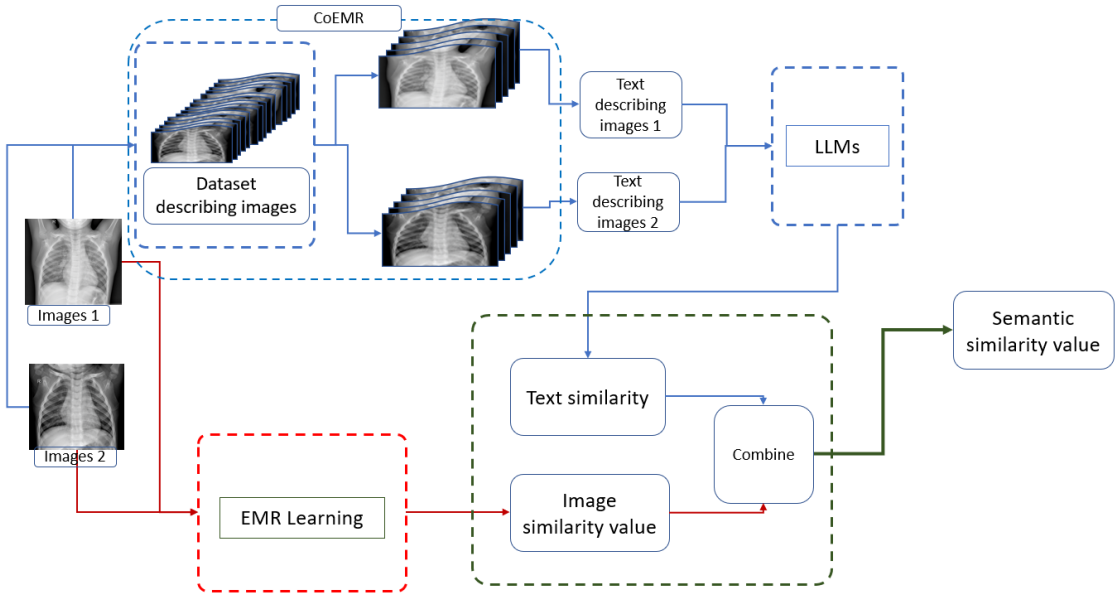

Figure 6. Estimate the semantic similarity of images based on EMR learning and LLMs

5. Conclusion

The paper proposes two important suggestions to improve the efficiency and accuracy of the EMR algorithm in content-based image retrieval.

Firstly, combining EMR rankings in different ways. Experimental results show that this combined approach is superior to other graph-based ranking methods (AGR, EMR, etc.) and distance metric learning methods between feature vectors.

Secondly, using the Support Vector Machine Regression (SVR) model to learn the rankings of EMR. This method allows estimating the similarity between two images that may not be in the image dataset. The effectiveness of the proposed EMR ranking learning method has been verified through experiments on image datasets in agriculture and healthcare.

The above results demonstrate the great potential of applying the proposed methods to enhance the quality of CBIR systems. In future work, we will continue to research combining natural language processing and AI methods based on large language models to improve the ability to predict and retrieve images by semantics. We hope this work will contribute to the development of the CBIR field and bring many useful applications in practice.

In summary, the key contributions of this paper are:

- Proposing effective combinations of EMR rankings to improve search accuracy
- Introducing EMR ranking learning using SVR regression for generalized similarity assessment
- Achieving state-of-the-art results on agricultural and medical image datasets
- Demonstrating the potential to integrate semantic analysis of images via natural language processing and large language models
- Providing novel techniques to advance content-based image retrieval systems for real-world.

References

[1] Bhagat, M., Kumar, D. *A comprehensive survey on leaf disease identification & classification,* Multimed Tools Appl 81, 33897–33925 (2022).

[2] Holger R. Roth, et al, *An application of cascaded 3D fully convolutional networks for medical image segmentation*, Computerized Medical Imaging and Graphics,(2018) 66, pp 90-99.

[3] Bac Nguyen, Carlos Morell, Bernard De Baets, *Supervised distance metric learning through maximization of the Jeffrey divergence*, Pattern Recognition, Volume 64, 2017, Pages 215-225

[4] Bac Nguyen, Francesc J. Ferri, Carlos Morell, Bernard De Baets, *An efficient method for clustered multi-metric learning*, Information Sciences, Volume 471, 2019, Pages 149-163

[5] Huy Tran Van, Dzung Pham Thi Kim, Huy Ngo Hoang and Quy Hoang Van, *Enhancing the performance of manifold ranking in image retrieval using combined rank on low-level features and embedded vectors*, J. Inf. Hiding Multim. Signal Process. 11(4), 172–186 (2020)

[6] Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Prasad Bhat, N., Shashank, N., Vinod, P.V.: *Plant disease detection using machine learning*. In: 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control.

[7] Ferentinos, K.P.: *Deep learning models for plant disease detection and diagnosis*. Comput. Electron. Agric. 145, 311–318 (2018).

[8] Goncharov, P. Ososkov, G., Nechaevskiy, A., Uzhinskiy, A., Nestsiarenia, I.: *Disease detection on the plant leaves by deep learning*. In: Kryzhanovsky, B., et al. (eds.), Neuroinformatics 2018, SCI 799, pp. 151–159 (2019)

[9] L. Yang, *The connection between manifold learning and distance metric learning*, Technical report, 2007.

[10] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai, and X. He. Emr: *A scalable graph-based ranking model for content-based image retrieval,* IEEE Transactions on Knowledge and Data Engineering, 27:102–14,2015.

[11] Zhao, B., Li, X. (2015). *Course Similarity Calculation Using Efficient Manifold Ranking*. In: Zhang, S., Wirsing, M., Zhang, Z. (eds) Knowledge Science, Engineering and Management. KSEM 2015. Lecture Notes in Computer Science(), vol 9403. Springer, Cham.

[12] H. X. Trung, D. V. Tuyet, N. H. Huy, S. Ablameyko, N. Q. Cuong and H. V. Quy, *A Novel Non-Gaussian Feature Normalization Method and its Application in Content Based Image Retrieval*, Nonlinear Phenomena in Complex Systems, vol. 22, no. 1, pp. 1-17, 2019.

[13] Q. H. Van, H. N. Hoang, T. D. Van, S. Ablameyko, H. T. Van, D. P. T. Kim and M. Ablameyko, *A modified Efficient Manifold Ranking Algorithm for Large Database Image Retrieval*, Nonlinear Phenomena in Complex Systems, 2020.

[14] Tan, M., and Le, Q. V.: *EfficientNet: Rethinking model scaling for convolutional neural networks*, 36th International Conference on Machine Learning, ICML 2019, 2019-June, pp. 10691-10700, 2019.

[15] W. Liu, J. He, and S.-F. Chang, *Large graph construction for scalable semi-supervised learning*, in Proceedings of the International Conference on Machine Learning (ICML), 2010, pp. 679–686

[16] Van Quy, H., Dzung, P.T.K., Huy, N.H., Van Huy, T. (2023). *Improved EfficientNet Network for Efficient Manifold Ranking-Based Image Retrieval*. In: Nguyen, T.D.L., Verdú, E., Le, A.N., Ganzha, M. (eds) Intelligent Systems and Networks. ICISN 2023. Lecture Notes in Networks and Systems, vol 752. Springer, Singapore.

[17] https://www.kaggle.com/code/omarhayekasfar/x-ray-classifier-with-cnn-keras-with-84-accuracy

[18] Yao Xiao, Shenglan Liu, Lin Feng and Xiuqi Hao, *A novel ranking algorithm based on manifold learning for CBIR system*, Proceeding of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 2014, pp. 1002-1009,

[19] Rassoul Hajizadeh, A. Aghagolzadeh, M. Ezoji, *Local distances preserving based manifold learning*, Expert Systems with Applications, Volume 139, 2020.

[20] https://arxiv.org/abs/1911.10674 Adaptive Nearest Neighbor: A General Framework for Distance Metric Learning

[21] Tran Van Huy, Ngo Hoang Huy, Dao Van Tuyet, Nguyen Van Doan, Hoang Trong Minh, Hoang Van Quy, Pham Thi Kim Dzung, Nguyen Thanh Y, Le Dinh Nghiep, *Học không giám sát độ đo tương tự trên dữ liệu đa tạp của các bộ mô tả hình ảnh*, REV-ECIT, 2023.

[22] O. Vakhno and L. Ma, *A Hybrid Architecture for Semantic Image Similarity Learning*, 2020 5th International Conference on Computational Intelligence and Applications (ICCIA), Beijing, China, 2020, pp. 94-97

[23] Korfhage, Nikolaus, Markus Mühling, and Bernd Freisleben. *ElasticHash: semantic image similarity search by deep hashing with elasticsearch*. Computer Analysis of Images and Patterns: 19th International Conference, CAIP 2021, Virtual Event, September 28–30, 2021, Proceedings, Part II 19. Springer International Publishing, 2021.

[24] https://research.google/blog/crisscrossed-captions-semantic-similarity-for-images-and-text/?m=1