Applied Computer Science 2026, vol. 22, no. 2, pp. 67–84
https://doi.org/10.35784/acs_8124

Submitted: 2025-07-19 | Revised: 2026-03-22 | Accepted: 2026-04-13



Keywords: content-based image retrieval (CBIR) systems, particle swarm optimization (PSO) algorithms, VGG16, k-means

*Mohsin Hasan HUSSEIN* [1,2*], *Ali Mohsin Ahmed AL-SABAAWI* [3], *Zakaria A. Hamed ALNAISH* [4]

[1] University of Kerbala, Karbala, Iraq, mohsin.h@uokerbala.edu.iq
[2] University of Warith Al-Anbiyaa, Iraq,
[3] Ninevah University, Iraq, ali.mohsin@uoninevah.edu.iq
[4] University of Mosul, Iraq, zakriahamoalnaish@uomosul.edu.iq
[*] Corresponding author: mohsin.h@uokerbala.edu.iq

# Improving image retrieval using CNN with PCA and optimized k-means clustering

**Abstract**

*Content-based image retrieval (CBIR) systems play an important role in many applications, including object recognition, digital forensics, and biomedical research. Handling large volumes of digital images in our daily lives requires the development of an efficient CBIR system. In this paper, a comprehensive and unified approach for image retrieval has been introduced. It brings together deep learning, feature reduction, and clustering optimization in a way that hasn't been explored before. The pre-trained VGG16 model has been used to extract rich, high-level features from images. To make the process more efficient and focused, Principal Component Analysis (PCA) is applied to reduce the number of features while retaining the most important information with minimal runtime. What sets our work apart is the use of hybrid K-means clustering with Particle Swarm Optimization (PSO). Instead of relying on random initialization, PSO helps find better starting points and improves the overall clustering by guiding K-means away from local minima. These well-known techniques are combined into a single, coherent framework called VGG16-PCA-PSO-K-means. The effectiveness of the proposed method has been evaluated on the Corel 1K and UC Merced Land Use datasets using mean average precision (mAP), clustering purity, recall, F-score, NDCG, and runtime. The results of the experiments conducted indicate that the proposed system achieves higher precision and clustering purity than state-of-the-art methods. The improvement ranged from 5% to 18% across different feature set sizes, with mAP@10 of 97.5% for the 10-class retrieval of the Corel 1K and mAP@10 of 96.7% for the 21-class retrieval of the UC Merced Land Use datasets, using only 30 features. Furthermore, the results of the Kendall's W and Friedman tests confirmed that the (VGG16-PCA-PSO-K-means) model achieved a higher rank than other methods in the literature, with a significant difference, highlighting its effectiveness and robustness.*

## 1. INTRODUCTION

Content-Based Image Retrieval (CBIR) has been extensively studied in computer vision for decades. The main objective of CBIR is to search for images by analyzing their visual contents, making image representation a crucial aspect of the process. The CBIR system performs retrieval by comparing the features of the query image with those of the images stored in the database.

Recently, Artificial Intelligence techniques (AI) have been used in various disciplines such as text mining (Blanchard et al., 2024), speech recognition (Lian et al., 2025), recommendation systems (Al-Mani et al., 2022), medical image classification (Alkahya et al., 2023; Dai et al., 2024), and target detection (Mon et al., 2017). Common Artificial Intelligence techniques used in the aforementioned disciplines include hybrid intelligence systems (Hasoon & Jasim, 2013), ensemble learning (Saeed et al., 2024), and optimization and machine learning algorithms (Alneamy & Alnaish, 2022). Deep learning, a part of AI, is widely used in developing the CBIR system (Gautam & Khanna, 2024). Feature extraction can be achieved using deep learning (Zhu et al., 2016), which is considered more efficient than traditional methods. Convolutional Neural Networks (CNNs), a deep learning algorithm, are among the most widely used techniques in computer vision.

It may be employed in feature categorization and extraction from digital images (Ghaleb et al., 2021). In their study, Ghaleb et al. (2021) proposed a Convolutional Neural Network (CNN) model that is capable of evaluating the mean retrieval accuracy for a dataset comprising 10 object images and 10-digit images, achieving accuracy rates of 92.9% and 99.8%, respectively. In Babenko et al. (2014), the authors used three Convolutional Neural Network (CNN) models for feature extraction, namely, D-Leaf, fine-tuned AlexNet, and pre-trained AlexNet. The obtained results have been compared with five machine-learning techniques in terms of accuracy. The fine-tuned AlexNet model achieved the best retrieval accuracy (95.54%).

However, the high-dimensional deep features extracted by CNNs (convolutional neural networks) limit the retrieval efficiency and make it difficult to satisfy the requirements of existing image retrieval (Chen et al., 2020). Therefore, many researchers suggested to use the dimension reduction technique, such as principal component analysis (PCA), to get a smaller number of features (Cao et al., 2018; Chen et al., 2020; Ghaleb et al., 2022). In Abdi and Williams (2010), the authors suggested compressing CNN features using discriminative dimensionality reduction and principal component analysis (PCA). In Pardede et al. (2019), the study utilizes deep features from a CNN model in combination with an XGBoost classifier. However, the method does not include any clustering, optimization, or feature reduction strategy. This limits its scalability to large, unlabeled image datasets. In Desai et al. (2021) and Kannagi and Lanke (2022), the authors combined VGG16 features with a Support Vector Machine (SVM) classifier. Although this hybrid improves classification performance, it still does not incorporate dimensionality reduction or any optimization process. This may lead to high computational costs and reduced generalizability.

In Cao et al. (2018), the authors used multilinear PCA (MPCA) instead of PCA for dimension reduction of image features. By combining pertinent tensor properties, this approach can directly reduce the dimensionality of tensors and accomplish the goal of dimension reduction. The PCA reduces the extracted features to a reasonable level and then utilizes them to retrieve images. Moreover, after reducing the extracted features, the similarity technique plays a vital role in comparing the query with the entire images in the dataset and retrieving the most similar images. However, the comparison requires a long search process and the consumption of storage resources as well. Furthermore, searching for an image in an entire dataset may negatively affect the retrieval accuracy. Therefore, it is more effective to categorize the images into groups based on their features. Thus, images with similar characteristics are grouped together, while those with differing features are placed in separate groups. This approach is known as clustering (Younus et al., 2015).

Clustering is a widely used technique in machine learning, with numerous studies applying it across various domains to group data objects into distinct clusters (Al-Sabaawi et al., 2021; Hussein et al., 2017; Mohammed et al., 2020). In image retrieval, the clustering approach has been employed in many studies (Sudhish & Nair, 2024; Wang et al., 2023). By utilizing clustering, the search time is reduced, and the accuracy of retrieval is significantly improved. One of the most common clustering techniques is the K-Means algorithm. Yet, one of its main limitations is its sensitivity to the initial selection of cluster centers (centroids). If the centroids have been set randomly, this may lead to accuracy reduction. Furthermore, the optimal solution will not be attained and will mostly be trapped in a local optimum. Thus, using optimization algorithms is regarded as a promising solution in determining the centroids to solve the K-Means problem. In addition, the particle swarm optimization (PSO) algorithm has better efficiency compared to other optimization algorithms, such as the Genetic algorithm (GA) and the Firefly algorithm (FA) (Bimantara & Widiartha, 2023).

After reviewing previous studies, especially (Desai et al., 2021; Hassan et al., 2023; Kannagi & Lanke, 2022; Pardede et al., 2019), it can be concluded that the majority of CBIR studies using deep learning models for feature extraction still extract high-dimensional feature vectors directly from their results without conducting any analysis of feature redundancy or dimensionality reduction problems. The standard k-means method, which clustering-based retrieval systems depend on, needs preset centroids for its operation. Yet, it results in both unstable results and inefficient cluster creation. Research studies aiming to connect dimensionality reduction methods with metaheuristic centroid optimization techniques for improved clustering stability and better retrieval performance remain scarce. Our proposed method directly addresses these issues through a unified, fully unsupervised framework that integrates VGG16 for deep feature extraction, PCA for dimensionality reduction, and a hybrid PSO–K-means algorithm for robust clustering. This combination not only eliminates reliance on labeled data but also improves clustering quality by reducing noise and guiding K-means toward more optimal solutions. This offers a theoretically justified and computationally efficient alternative to prior methods. Accordingly, the proposed model achieves the following contributions:

The VGG16-based convolutional neural network extracts high-level semantic features from images, thereby improving image discrimination compared to traditional handcrafted descriptors.

The dimensionality of deep features was reduced using Principal Component Analysis (PCA). The study used three candidate dimensions (10, 15, and 30) to determine the best balance between compact representation and accurate retrieval. This differs from existing research that applies fixed-dimension reductions.

It investigates the performance of integrating the K-means clustering algorithm with the Particle Swarm Optimization (PSO) algorithm for solving the clustering problem in CBIR.

The study employed the non-parametric Kendall's W test to evaluate the statistical significance of performance results, necessitating more comprehensive validation than simple metric-based comparisons.

The purity metric demonstrated that combining optimization techniques with clustering methods yields better results than using clustering methods alone, as evidenced by the enhanced cluster homogeneity observed during testing.

The rest of the paper is organized as follows: Section 2 represents the method, Section 3 includes the Results and Discussions, Section 4 represents the challenges and limitations, and finally, Section 5 concludes the paper.

## 2. METHODOLOGY

The method proposed in this paper includes two phases. The first phase represents feature extraction and selection processes, depending on a hybrid VGG-16 CNN with PCA. The second phase represents implementing the PSO algorithm for selecting the cluster center (centroid) in the K-means algorithm to enhance its performance in retrieving the most similar images. Fig. 1 represents the methodology of the proposed method. Further, the details of each phase are explained in the following subsections.

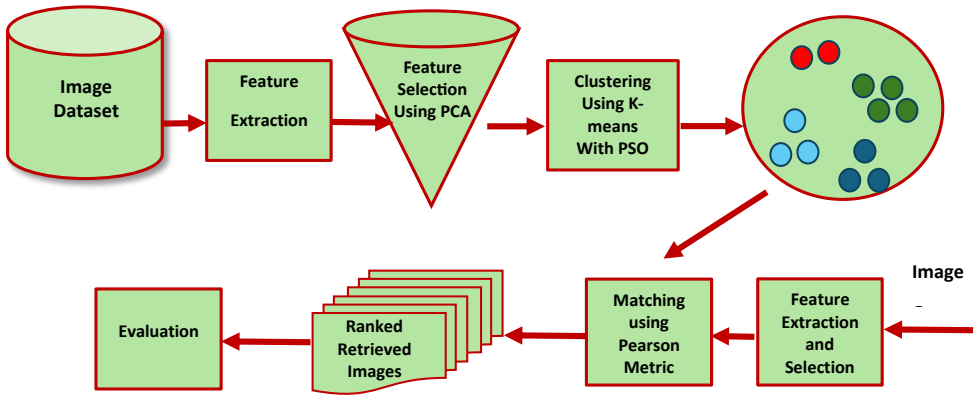


**Fig. 1. The methodology of the proposed method**

### 2.1. The implemented VGG

The VGG-16 model is a 16-layer convolutional neural network that Simonyan and Zisserman (2014) and his colleagues from Oxford University's Visual Geometry Group Lab proposed in 2014. Only three 3x3 convolutional layers are placed on top of one another in this simple architecture. The initial two convolutional layers consist of 64 feature kernel filters, each with a size of 3×3. When an RGB image with a depth of 3 is fed into the network, it is processed through these layers, resulting in an output dimension of 224×224×64. The output is then passed to a max pooling layer with a stride of 2. The third and fourth convolutional layers utilize 128 feature kernel filters of the same 3×3 size. After these layers, another max pooling layer with a stride of 2 reduces the dimensions to 56×56×128. Three additional convolutional layers (fifth, sixth, and seventh) follow the initial layers. Each of these layers uses 256 feature maps and a 3x3 kernel size. A second max pooling layer with a stride of two follows each convolutional layer. The architecture then includes two groups of convolutional layers (eighth through thirteenth), where each layer contains 512 3×3 kernel filters. A max pooling layer with a stride of 1 follows each set. Each of the fourteenth and fifteenth layers has 4096 units and is made up of fully connected hidden layers. Lastly, a SoftMax output layer (sixteenth layer) with ten units completes the network. Thus, the features are extracted from fc7 of the VGG16 network. Fig. 2 represents the structure of the VGG-16 CNN used.

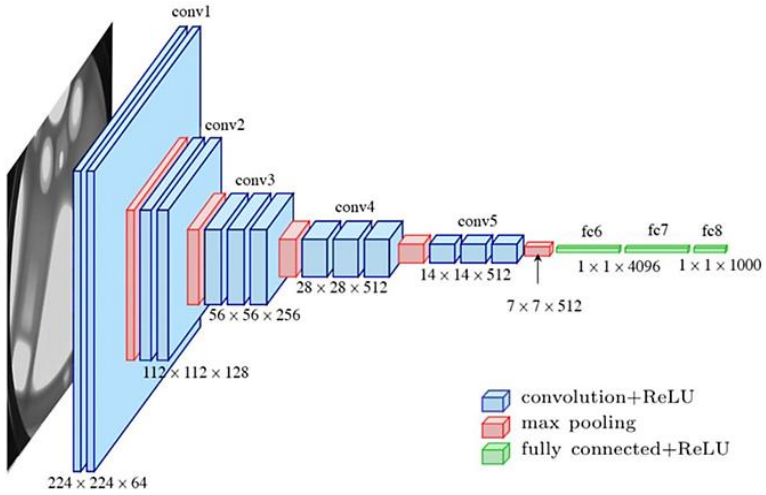


**Fig. 2. The structure of the VGG-16 CNN used**

## 2.2. Principal component analysis

The deep features extracted from VGG16 have a high-dimensional structure, producing 4096-dimensional vectors per image. The high dimensionality leads to increased computational expenses and potential feature redundancy problems. The researchers used Principal Component Analysis (PCA) as a solution to this problem because it enabled them to decrease data dimensions while maintaining the crucial information that represented maximum data variance (Greenacre et al., 2022). The dimensionality was reduced from 4096 to 10, 15, and 30 principal components. PCA transforms the original correlated feature space into principal components through its statistical method, which creates a new set of orthogonal components. The covariance matrix of the feature set is used to derive the components through eigenvector analysis, which identifies maximum variance directions. The components are ordered according to their associated eigenvalues, and only the top components that capture the majority of data variance are retained.

PCA achieves its goal of reducing redundant elements together with unneeded components through the process of projecting deep feature vectors onto the reduced subspace while retaining essential discriminative information. The process of dimensionality reduction results in improved computational performance together with better stability of clustering results during the following K-means–PSO optimization process.

## 2.3. Hybrid K-means and particle swarm optimization (PSO) algorithms

One of the most common unsupervised learning algorithms is K-Means, which is usually used for solving clustering problems. One of its common weaknesses is sensitivity to the center selection of the cluster (centroid). Randomly determining the centroids in the K-Means algorithm leads to decreased accuracy and mostly gets trapped in a local optimum (Bimantara & Widiartha, 2023). Thus, several optimization algorithms, such as the firefly algorithm (FA) (Maazalahi & Hosseini, 2024), equilibrium optimizer algorithm (Al-Kababchee et al., 2023), Particle Swarm Optimization (PSO) algorithm (Wang et al., 2018), and the genetic-fuzzy ant colony optimization algorithm (Ran et al., 2024), have recently been used to enhance the performance of the K-Means algorithm. In Bimantara and Widiartha (2023), the authors integrated the Particle Swarm Optimization with the K-Means algorithm. The main objective of the hybridization of the Particle Swarm Optimization (PSO) and K-means is to combine their strengths to improve clustering performance and solve the aforesaid weaknesses by determining the optimum centroids of the clusters (Bimantara & Widiartha, 2023; Younus et al., 2015). Mainly for this reason, it has been adopted in this paper. Further, VGG16 was selected as the feature extraction backbone for the content-based image retrieval system due to its proven ability to capture rich spatial feature representations. Besides, VGG16 provides structured convolutional descriptors that are particularly effective for similarity-based retrieval tasks (Saritha et al., 2019). In addition, Palla and Karra (2025) have compared the performance of the VGG-16 model with EfficientNet-B0, DenseNet-201, and AlexNet for CBIR, and the results have demonstrated that the VGG-16 model performs well.

The following steps have been followed to get the optimal centers of the appointed clusters for the K-means algorithm using the PSO algorithm.

1. Set the initial values for PSO parameters: population size (N), Maximum iteration, dimensions ($d$) (no. of clusters)
2. Generate the population (N) randomly, where the population (N) includes many particles, and each particle is represented by a 2-dimensional matrix ($d*a$). dimensions ($d$) represent the number of clusters, and a indicates the number of attributes.
3. Find the fitness value for all the individuals (i.e., best centers) in the population using the fitness function based on the Squared Euclidean Distance for each cluster, as shown in Equation (1).

$$dist(p,q) = \sum_{i=1}^{n}(p_i - q_i)^2 \quad (1)$$

Where $dist$ ($p$, $q$) is the distance between $p$ and $q$. $n$ represents the number of data attributes. $p_i$ represents the data for i-th attribute. $q_i$ indicates value of the i-th coordinate of the centroid.

1. Apply the PSO algorithm by updating the velocity and the positions for each individual (proposed solution) by Equation (2) and Equation (3), respectively.

$$v_i^{t+1} = wv_i^t + c_1 * r_1\left(P_{besti}^t - x_i^t\right) + c_2 * r_2\left(g_{besti}^t - x_i^t\right) \quad (2)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (3)$$

Where $x_i^t$ represents the values of centroids, $\boldsymbol{v}$ denotes the velocity vector, $\boldsymbol{w}$ represents the inertia weight that is used to balance the exploration and exploitation, $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are random numbers generated within the interval [0, 1] $^{d}$ (where $d$ represents the dimensionality of the search space, which is equal to the number of clusters). $c1$ and $c2$ are positive constants called "acceleration coefficients". $P_{besti}^t$ represents the best solution (i.e., best particle which contains the best centroids) at iteration $t$, $g_{besti}^t$ represents the best solution (i.e., the best particle which contains the best centroid values) of all the iterations.

2. Repeat steps 3–6 until the stop condition is satisfied (maximum iteration).
3. Group the data into clusters where the total distance of the points from the centroid is the smallest. Fig. 3 demonstrates the workflow of the hybrid PSO-K-means algorithm.

In the proposed algorithm, the population consists of 150 particles. The algorithm begins by randomly initializing the centroids. In Step 3, the fitness values are computed by calculating the distances between all data points in the dataset and each centroid. Based on these fitness values, the personal best ($P_{best}$) and global best ($g_{best}$) parameters are updated accordingly. Subsequently, the particle velocities are updated, followed by the position updates. These steps are iteratively repeated until the stopping criterion is satisfied. Finally, the global best ($g_{best}$) solution represents the optimal centroids of all clusters across all particles, corresponding to the minimum distance (i.e., the best clustering quality).

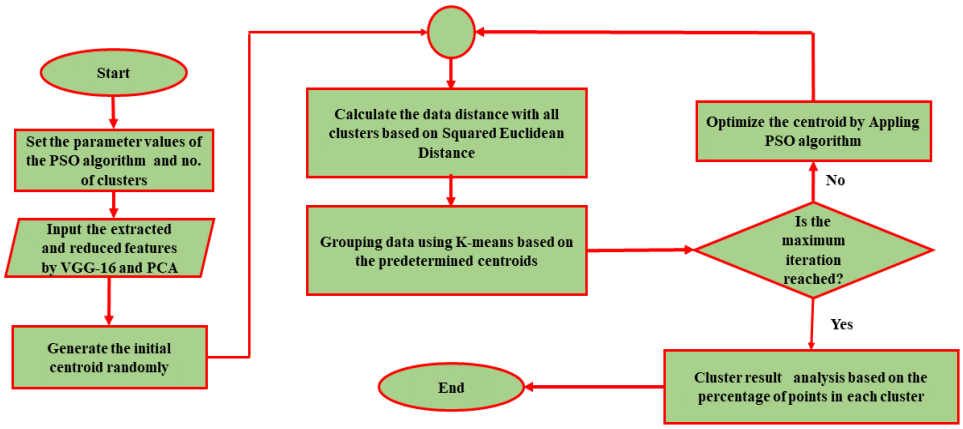


**Fig. 3. The workflow of the hybrid K-means-PSO algorithms**

## 3. RESULTS AND DISCUSSIONS

### 3.1. Dataset

In this study, the dataset utilized is called Corel 1k (Krizhevsky & Hinton, 2009). This dataset involves 1000 images with 256*384 or 384*256. The images fall into 10 categories; each category has 100 images. Although the images within the same class exhibit nearly identical global structures, there are noticeable variations in several of their individual properties. However, there are also significant similarities between the pictures in the various classes (for example, the bus and building limits). In addition, we conducted experiments using the UC Merced Land Use Dataset to assess the proposed approach in contemporary, complex scenario-based settings. The model evaluation requires assessment using two datasets, which provide distinct tests for evaluating model performance across different data distribution tests. The evaluation process obtains balanced results through the two benchmark tests, which enable comparison with existing research while testing model performance against contemporary dataset conditions. The dataset comprises 2,100 images with 256*256 distributed across 21 classes, with 100 images per class. The categories include diverse scene types such as Agriculture, Airplane, Parking plot, Freeway, and other land-use classes. Fig. 4 and Fig. 5 depict a sample of images.

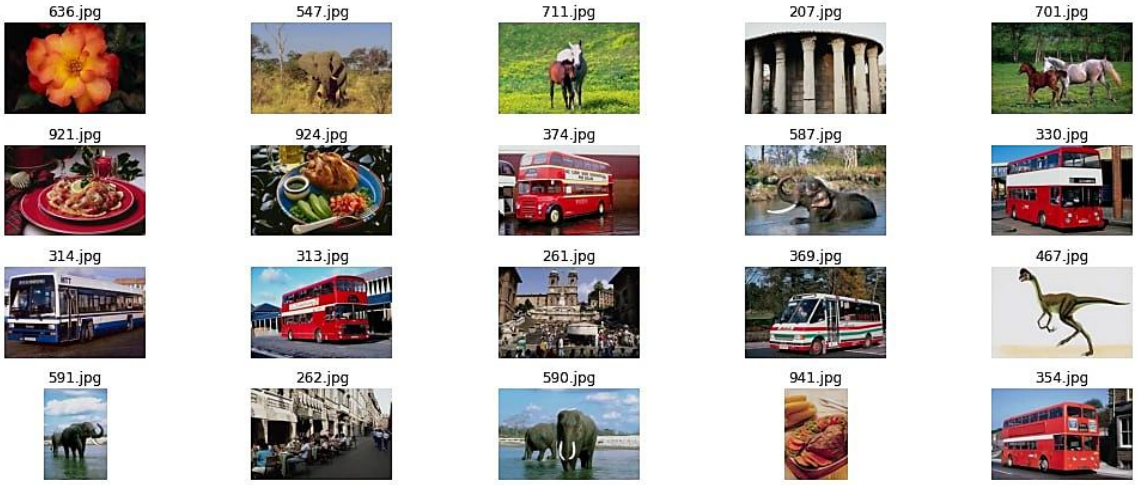


**Fig. 4. Samples of the Corel 1k dataset**


**Fig. 5. Sample of the UC-Merced land Use dataset**

### 3.2. Evaluation performance

Precision and mean average precision (*mAP*) are used to evaluate the performance of the CBIR. To calculate precision, divide the total number of recovered photos by the total number of relevant images. It demonstrates how a system can only provide pertinent images, as shown in Equation 4. *AP* refers to the average precision (Equation 5). The *mAP* indicates the mean of the average precision for entire clusters (Equation 6).

$$p = \frac{number\ of\ relevant\ retrieved\ images}{total\ number\ of\ the\ retrieved\ images\ of\ a\ specific\ class} \tag{4}$$

$$Ap = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{5}$$

$$mAp = \frac{1}{m}\sum_{k=1}^{m} Ap_k \tag{6}$$

Where $p_i$ is the *p* value of *i-th* class, *n* is the number of classes, $Ap_k$ refers to the *Ap* of class k, *m* is the number of implementations. In addition, recall and f-score metrics were used to evaluate the model's performance. Equations 7 and 8 show the computation process of these metrics.

$$recall = \frac{number\ of\ relevant\ retrieved\ \ images}{total\ number\ of\ the\ \ relevent\ images\ of\ a\ specific\ class} \tag{7}$$

$$f - score = \frac{precision@10 * recall@100}{precision@10 + recall@100} * 2 \tag{8}$$

Also, the Kendall *W* test has been used. It is a nonparametric statistical method for quantifying the level of concordance among multiple values. It evaluates the consistency depending on the rank given by different models to a collection of subjects. In this test, each instance stands for a judge or rater, and every factor signifies the entity or individual under evaluation. The steps used in computing the score for the Kendall *W* test (Legendre, 2005) are explained below:

1. Suppose the object *(i)* is the CBIR model (i.e. object (*i*) refers to one method of *n* methods that would be ranked), the ranked objects have obtained the rank ($r_{ij}$) depending on the raters (*j*) represented by the datasets used, where a total number of objects is (*n*) and a total number of raters is (*m*). Equation (9) represents the calculation formula to find the rank (*R*) of the object (*i*).

$$R_i = \sum_{j=1}^{m} r_{i,j} \tag{9}$$

2. Calculate the average value of (R) using Equation 10.

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{10}$$

3. Calculate the summation of squared deviations (*S*) using Equation 11.

$$S = \sum_{i=1}^{n}(R_i - \bar{R})^2 \tag{11}$$

4. Calculate the Kendall's *W* coefficient using Equation 12.

$$W = \frac{12*S}{m^2(n^3-n)} \tag{12}$$

The value of a Kendall *W* test score will be between zero and one, and the significance of the decision will rely on one of the following conditions (Landis & Koch, 1977; Alnaish & Hasoon, 2023):

- If w >=0.00 and < 0.20 then Slight agreement
- If w >= 0.20 and < 0.40 then Fair agreement
- If w >= 0.40 and < 0.60 then Moderate agreement
- If w>= 0.60 and < 0.80 then Substantial agreement
- If w > = 0.8, then it is considered a perfect agreement

Moreover, the Friedman test has been applied. It is a non-parametric test for comparing the performance of multiple algorithms across several datasets. Its results show whether there is a significant difference among the performance of the compared models or not (Jan & Shieh, 2025).

Furthermore, the Normalized Discounted Cumulative Gain (NDCG) metric is usually used to assess the results of information retrieval and recommender systems, as well as image retrieval. The system evaluates ranked lists through this method by measuring how effectively they rank their most important items. The NDCG at position K (or NDCG@K) is represented by equations (13) and (14) (Jeunen et al., 2023).

$$\text{NDCG@k} = \frac{\text{DCG@k}}{\text{IDCG@k}} \tag{13}$$

$$DCG@K = \sum_{i=0}^{k} \frac{rel_i}{log_2(i+1)} \tag{14}$$

Where $k$ = number of retrieved results, $rel_i$ = relevance score of images at rank $i$, $i$ = position in the ranked list.

### 3.3. Experimental results

In this paper, the experiments are implemented using SPYDER (Python 3.12) with the sklearn library and executed on a laptop with an Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz and 16 GB RAM. The preprocessing stage resized all input images to a fixed resolution of 224 × 224 pixels to ensure compatibility with the VGG16 architecture and maintain uniform input dimensions across the dataset.

The parameters values of the PSO algorithm in this paper are set as follows: the number of particles ($np$) set to 150 particles, the number of PSO maximum iterations set to 100 iterations, the values of both $c1$ and $c2$ set to 1, number of clusters =10 or 21 depending on the dataset used, and the values of $w$ set to 0.5.

In addition, the PCA algorithm was exploited to reduce dimensionality. In this study, three dimensions were applied: 10, 15, and 30. Table 1 and Table 2 show the average precision results for these three dimensions of Corel-1K and UC Merced Land Use datasets, respectively.

**Tab. 1. Average Precision % of the proposed method in terms of mean average precision of Corel-1K dataset**

| Classes | Dim =10 | Dim = 15 | Dim =30 |
|---|---|---|---|
| Africa | 0.92 | 0.92 | 0.94 |
| Beaches | 0.94 | 0.95 | 0.95 |
| Building | 0.97 | 0.97 | 0.98 |
| Bus | 1 | 1 | 1 |
| Dinosaur | 1 | 1 | 1 |
| Elephant | 1 | 1 | 1 |
| Flowers | 1 | 1 | 1 |
| Horses | 0.99 | 0.99 | 1 |
| Mountains | 0.98 | 0.98 | 0.97 |
| Food | 0.92 | 0.93 | 0.92 |
| *mAP* | 0.972 | 0.974 | 0.975 |

**Tab. 2. Average Precision % of the proposed method in terms of mean average precision of UC-Merced Land Use Dataset**

| Classes | Dim =10 | Dim = 15 | Dim =30 |
|---|---|---|---|
| Agriculture | 0.92 | 0.98 | 0.96 |
| Airplane | 0.92 | 0.92 | 1 |
| Baseball | 0.94 | 0.95 | 0.97 |
| Beach | 0.96 | 1 | 0.99 |
| Building | 0.89 | 0.90 | 0.96 |
| Chaparral | 0.95 | 0.95 | 0.955 |
| Denseresidential | 0.96 | 0.95 | 0.96 |
| Forest | 0.97 | 0.97 | 1 |
| Freeway | 0.87 | 0.89 | 0.95 |
| Golfcourse | 0.93 | 0.94 | 0.96 |
| Harbor | 0.95 | 0.95 | 0.98 |
| Intersection | 0.88 | 0.91 | 0.95 |
| Mediumresidential | 0.87 | 0.90 | 0.95 |
| Mobilehomepark | 0.95 | 0.95 | 0.95 |
| Overpass | 0.94 | 0.94 | 0.97 |
| Parkinglot | 0.97 | 0.97 | 1 |
| River | 0.95 | 0.95 | 0.95 |
| Runway | 0.90 | 0.91 | 0.94 |
| Sparseresidential | 0.92 | 0.94 | 0.97 |
| Storagetanks | 0.95 | 0.95 | 0.99 |
| Tenniscourt | 0.88 | 0.91 | 0.97 |
| *mAP* | 0.927 | 0.939 | 0.967 |

As demonstrated in Tables 1 and 2, the precision values are nearly perfect in most cases, indicating the effectiveness of the proposed method in retrieving images belonging to the same category as the query image. Specifically, the precision values for the classes of the Corel-1K dataset of "Bus," "Dinosaur," "Elephant," "Horses," and "Flowers" are consistently 1.0 across all dimensions. In contrast, the precision values of the same dataset for the "Buildings" and "Mountains" classes are close to 1.0, reflecting high performance with minimal variation. However, the lowest precision values are observed in the "Africa" and "Food" classes, ranging between 0.92 and 0.94. Similarly, in the second dataset (Table 2), the feature dimension of 30 outperformed the lower-dimensional configurations (10 and 15), indicating the importance of retaining a higher number of discriminative components. Specifically, the mean Average Precision (*mAP*) reached 0.967 when the dimensionality is set to 30, compared to 0.939 and 0.927 for 15 and 10 dimensions, respectively.

Further, class-wise evaluation reveals that most of the categories achieved very high precision values, reaching 1.0 or close to 1.0 for classes such as Agriculture, Airplane, Chaparral, and Forest.

Overall, the experimental results demonstrate that the integration of VGG16 for feature extraction, Principal Component Analysis (PCA) for feature reduction, and optimized K-means clustering forms an effective framework for image retrieval/classification. The findings indicate that all evaluated dimensional configurations (10, 15, and 30 components) achieved comparable performance levels. However, a slight improvement is observed when the dimensionality is set to 30, for both datasets. These observations highlight the dataset-dependent impact of dimensionality selection. Therefore, the results obtained with dimension 30 can be considered reliable, offering the advantage of reduced dimensionality.

The study investigates distinctiveness together with intra-class variation. In some classes, such as dinosaurs and buses, the structural patterns show simple and precise characteristics that enable better discrimination from other categories. The consistent shapes and clear object boundaries in these categories reduce ambiguity during feature extraction and similarity matching. In contrast, other classes exhibit more complex structural layouts and diverse texture patterns, resulting in higher intra-class variability and increased inter-class similarity. These characteristics introduce greater challenges for the retrieval process and consequently lead to lower precision values.

Moreover, Tables 3 and 4 present the evaluation results in terms of Precision@10, Recall@100, and F-score for dimensionality settings of 10 and 30 principal components. The evaluation of Recall at small K values (5 or 10) shows non-representative results because each class contains 100 images. The study reports Precision@10 to measure early retrieval performance, while Recall results are obtained at K =100, which represents the total relevant images in each class. F1-score results are obtained by combining the Precision@10 results with the Recall@100 results.

**Tab. 3. The results of precision @10, recall@100, and f-score of Corel-1k dataset**

| **Dataset** | **Dim =10** | | | **Dim =30** | | |
|---|---|---|---|---|---|---|
| | **precision@10** | **Recall @100** | **F-score** | **precision @10** | **Recall @100** | **F-score** |
| Africa | 0.92 | 0.95 | 0.934 | 0.94 | 0.96 | 0.94 |
| Beaches | 0.94 | 0.91 | 0.924 | 0.95 | 0.91 | 0.92 |
| Building | 0.97 | 0.96 | 0.964 | 0.98 | 0.96 | 0.96 |
| Bus | 1 | 0.99 | 0.99 | 1 | 0.99 | 0.99 |
| Dinosaur | 1 | 0.99 | 0.99 | 1 | 0.99 | 0.99 |
| Elephant | 1 | 0.98 | 0.98 | 1 | 0.99 | 0.99 |
| Flowers | 1 | 0.99 | 0.99 | 1 | 0.99 | 0.99 |
| Horses | 0.99 | 0.99 | 0.99 | 1 | 0.99 | 0.99 |
| Mountains | 0.98 | 0.94 | 0.95 | 0.97 | 0.94 | 0.95 |
| Food | 0.92 | 0.96 | 0.93 | 0.92 | 0.96 | 0.93 |
| Av | 0.972 | 0.966 | 0.964 | 0.975 | 0.968 | 0.965 |

**Tab. 4. The results of precision @10, recall@100, and f-score of UC-merced land use dataset**

| Dataset | Dim =10 | | | Dim =30 | | |
|---|---|---|---|---|---|---|
| | precision@10 | Recall @100 | F-score | precision@10 | Recall @100 | F-score |
| Agriculture | 0.92 | 0.91 | 0.914 | 0.96 | 0.93 | 0.944 |
| Airplane | 0.92 | 0.92 | 0.929 | 1 | 0.94 | 0.969 |
| Baseball | 0.94 | 0.52 | 0.669 | 0.97 | 0.76 | 0.852 |
| Beach | 0.96 | 0.98 | 0.969 | 0.99 | 0.98 | 0.984 |
| Building | 0.89 | 0.44 | 0.588 | 0.96 | 0.44 | 0.603 |
| Chaparral | 0.95 | 0.96 | 0.954 | 0.955 | 0.96 | 0.957 |
| Denseresidential | 0.96 | 0.55 | 0.699 | 0.96 | 0.60 | 0.738 |
| Forest | 0.97 | 0.98 | 0.974 | 1 | 0.98 | 0.989 |
| Freeway | 0.87 | 0.49 | 0.626 | 0.95 | 0.84 | 0.891 |
| Golfcourse | 0.93 | 0.54 | 0.683 | 0.96 | 0.68 | 0.796 |
| Harbor | 0.95 | 0.91 | 0.929 | 0.98 | 0.99 | 0.984 |
| Intersection | 0.88 | 0.34 | 0.490 | 0.95 | 0.38 | 0.542 |
| Mediumresidential | 0.87 | 0.29 | 0.435 | 0.95 | 0.43 | 0.592 |
| Mobilehomepark | 0.95 | 0.46 | 0.619 | 0.95 | 0.36 | 0.522 |
| Overpass | 0.94 | 0.43 | 0.590 | 0.97 | 0.51 | 0.668 |
| Parkinglot | 0.97 | 0.79 | 0.870 | 1 | 0.32 | 0.484 |
| River | 0.95 | 0.47 | 0.628 | 0.95 | 0.71 | 0.812 |
| Runway | 0.90 | 0.35 | 0.504 | 0.94 | 0.56 | 0.701 |
| Sparseresidential | 0.92 | 0.53 | 0.672 | 0.97 | 0.79 | 0.870 |
| Storgetanks | 0.95 | 0.78 | 0.856 | 0.99 | 0.71 | 0.826 |
| Tenniscourt | 0.88 | 0.42 | 0.56 | 0.97 | 0.42 | 0.58 |
| AV | 0.927 | 0.62 | 0.72 | 0.967 | 0.68 | 0.77 |

As shown in Table 3 (Corel-1k dataset), the values of precision and recall are very close for both dimensionality configurations (Dim = 10 and Dim = 30), with a slight improvement observed when the dimensionality is increased to 30. The F-score values are also nearly identical in both cases, indicating that reducing the dimensionality to 10 components preserves most of the discriminative information required for effective clustering. However, retaining 30 principal components provides marginal yet consistent performance gains, suggesting improved feature representation stability.

In Table 4 (UC-Merced Land Use dataset), recall is computed based on the cluster size of 100 images. The results demonstrate that precision values remain high under both dimensional settings, with noticeable improvement when the dimensionality is increased to 30. This confirms that preserving a larger number of principal components maintains more discriminative visual information extracted from VGG16 features. Similarly, recall improves when using 30 components compared to 10 components.

Nevertheless, when comparing recall with precision, recall values are generally lower, particularly in the UC-Merced dataset. This discrepancy indicates the higher complexity and intra-class similarity of aerial scene images. The larger number of clusters and the visual similarity among land-use categories make the clustering task more challenging compared to the Corel-1k dataset.

Finally, the F-score is adopted as a balanced evaluation metric to reflect the trade-off between precision and recall, providing a comprehensive assessment of clustering performance across both datasets. Additionally, the average run time for the proposed methods with and without using PCA is based on 10, 15, and 30 features, as shown in Table 5.

**Tab. 5. Comparative results of the proposed method, VGG 16-PCA-Kmeans-PSO, in terms of run time/seconds**

| # Features / Datasets | 10 features | 15 features | 30 features | All features without PCA |
|---|---|---|---|---|
| Corel 1K (1000 samples) | 123.2616 | 123.6849 | 124.6463 | 2561.6011 |
| UC Merced Land Use Dataset (2100 samples) | 240.0335 | 240.5932 | 242.3821 | 2636.27 |

As shown in tables (1), (2), and (5), for the Corel 1K dataset that contains 1000 samples with ten classes, when the number of features was 10, 15, and 30, it required only around 123.2616, 123.6849, and 124.6463 seconds, respectively. The obtained *mAP* for 10, 15 and 30 features were 0.972 %, 0.974%, and 0.975 %, respectively. On the other hand, the run time for the proposed method without using PCA is very high (around 2562 seconds).

Regarding the UC Merced Land Use Dataset that contains 2100 samples with twenty-one classes, when the number of features was 10, 15, and 30, it required only around 240.0335, 240.5932, and 242.3821seconds, respectively. The obtained *mAP* for 10, 15 and 30 features were 92.7%, 93.9%, and 96.7%, respectively. On the other hand, the run time for the proposed method without using PCA is very high (around 2636.27 seconds). It is very clear that using the PCA algorithm enhances the proposed method's performance in terms of runtime.

Table 6 presents the performance results of the proposed method evaluated using the NDCG metric on the Corel-1K and UC-Merced Land Use (UCM) datasets. The table consists of three columns. The first column represents the cutoff rank (k), indicating the number of retrieved images considered in the evaluation. The second and third columns report the corresponding NDCG values for the Corel-1K and UCM datasets using the proposed method, respectively.

**Tab. 6. The results of NDCG for the proposed method**

| K | NDCG of Corel-1k dataset | NDCG of the UCM dataset |
|---|---|---|
| 5 | 0.91 | 0.89 |
| 10 | 0.86 | 0.82 |
| 20 | 0.82 | 0.72 |

As shown in Table 6, the NDCG values decrease as the cutoff rank (k) increases for both datasets. For the Corel-1K dataset, the highest NDCG value (0.91) is achieved at k = 5, while the lowest value (0.82) is observed at k = 20. For the UCM dataset, the highest NDCG value (0.89) is achieved at k = 5, while the lowest value (0.72) is observed at k = 20. This decline can be attributed to the fact that retrieval accuracy typically decreases as more images are included in the ranked list. In other words, the top-ranked images are generally more similar to the query image than those appearing at lower ranks. Furthermore, the NDCG scores for the Corel-1K dataset are consistently higher than those for the UCM dataset across all values of k, indicating that the proposed method achieves better relevance-based ranking performance on the Corel-1K dataset.

Furthermore, the computational complexity of the proposed method, VGG16-PCA-PSO-K-means, is represented by two parts. The first part includes the complexity of VGG16-PCA. The complexity of VGG16 for all convolution layers is represented by $O(\sum_{l=1}^{13} H_l W_l K^2 C_l^{in} C_l^{out})$, where H and W are the spatial dimensions of the feature map, K is the size of the kernel, $C_{in}$ and $C_{out}$ are the numbers of input channels and the number of output filters, respectively. The complexity of PCA is represented by $O(d^3) + O(ndm)$, where n represents the number of images, d represents the total number of feature dimensions, and m is the reduced dimension. Thus, the total complexity for VGG16-PCA is:

$$O(\sum_{l=1}^{13} H_l . W_l . K^2 . C_l^{in} . C_l^{out}) + O(d^3) + O(n.d.m) \quad (15)$$

The second part includes the complexity of PSO-K-means. The computational complexity for PSO-K-means is dominated by the PSO algorithm. At each iteration in PSO, the fitness value for each particle is evaluated depending on *N* samples, and *K* clusters with *D* dimensions. Thus, the total complexity of PSO-K-means is: O (*T.P.N.K.D*), where *T* represents the PSO iterations, *P* represents the population size, *N* is the number of samples in the dataset, *K* represents the number of clusters, and *D* represents the dimensions of the features. Fig. 6 shows examples of image retrieval from the Corel-1k dataset based on a query.


**Fig. 6. Examples of image retrieval based on a query of the Corel-1k dataset**

### 3.4. Benchmarking with previous methods

In this subsection, the proposed method is compared with four recent studies that also employed deep learning techniques. The first study used XGBoost as a classifier with a convolutional neural network (CNN) for feature extraction to improve CBIR performance (Pardede et al., 2019). The second study proposed a new hybrid approach as CBIR based on a pre-trained VGG16 deep neural network and a support vector machine (SVM) (Desai et al., 2021). The third study used a convolutional neural network (CNN) for feature extraction and a support vector machine (SVM) for classification as a new CBIR approach (Kannagi & Lanke, 2022). The fourth study developed a new CBIR technique represented by using CNN for feature extraction and Euclidean distance for similarity measurement and retrieving the first top 10 pictures (Hassan et al., 2023). From a theoretical standpoint, VGG16 offers high-level semantic representations that capture deep spatial and contextual features, which are often missed by conventional descriptors. PCA plays a key role in reducing noise and eliminating irrelevant variance in the feature space, making the clustering process more focused and efficient. Furthermore, while K-means clustering minimizes intra-cluster variance, it is inherently sensitive to initial centroid placement and can converge to suboptimal solutions. By integrating PSO, which is designed to perform global search over the solution space, the system is able to avoid poor local minima and find better clustering configurations.

Table 7 presents the comparative results. The first column lists the dataset, while the corresponding numerical values represent the average precision achieved for each class by the previous methods in the literature. The results of the proposed method were recorded with a dimensionality of 10.

**Tab. 7. Performance comparison with previous methods in terms of mean average precision of Corel 1K and UC-Merced Land Use datasets**

| **Dataset** | **Previous methods** | | | | |
|---|---|---|---|---|---|
| **Corel 1K (1000 samples)** | **Ref (Pardede et al., 2019)** | **Ref (Desai et al., 2021)** | **Ref (Kannagi & Lanke, 2022)** | **Ref (Hassan et al., 2023)** | **Proposed method** |
| No. of retrieved images | | 10 | 10 | 10 | 10 |
| Africa | 0.63 | 0.84 | 0.90 | 0.91 | 0.92 |
| Beaches | 0.53 | 0.84 | 0.96 | 0.65 | 0.98 |
| Buildings | 0.36 | 0.835 | 0.94 | 0.95 | 0.97 |
| Bus | 0.54 | 0.827 | 1 | 0.90 | 1 |
| Dinosaurs | 0.82 | 0.83 | 1 | 1 | 1 |
| Elephants | 0.63 | 0.838 | 0.86 | 0.75 | 1 |
| Flowers | 1 | 0.83 | 0.94 | 1 | 1 |
| Horses | 0.93 | 0.84 | 0.86 | 0.99 | 0.99 |
| Mountains | 0.60 | 0.838 | 0.90 | 1 | 0.99 |
| Food | 0.88 | 0.837 | 0.94 | 0.65 | 0.94 |
| *mAP* | 0.69 | 0.836 | 0.933 | 0.88 | 0.98 |
| UC Merced Land Use Dataset (2100 samples ) | MALTP Ref (Ugale & Railkar, 2025) | RESNet50 Ref (Ugale & Railkar, 2025) | MALTP+ RESNet50 Ref (Ugale & Railkar, 2025) | Proposed method | |
| *mAP* | 0.93 | 0.95 | 0.98 | 0.967* | |

For the Corel 1K dataset (1000 samples), the proposed method achieved the best results compared to the previous studies. Notably, the study referenced in (Kannagi & Lanke, 2022) recorded the second-best performance. Specifically, the results for classes such as Africa, Buildings, Dinosaurs, Flowers, Horses, and Mountains are relatively close between the proposed method and (Hassan et al., 2023), with a light enhancement toward the former. However, the proposed method demonstrates significant improvements in other classes, particularly Elephants and Beaches, where the achieved accuracies are 1.00 and 0.98, respectively. In contrast, the corresponding accuracies in Hassan et al. (2023) are 0.75 and 0.65, highlighting the superior performance of the proposed approach in these categories. Furthermore, the study presented in (Kannagi & Lanke, 2022) achieved the second-best results. Notably, in Africa, Beaches, Buildings, Bus, Dinosaurs, and Food classes, the performance of Kannagi and Lanke (2022) is closely comparable to that of the proposed method. However, the superiority of the proposed approach over (Kannagi & Lanke, 2022) is evident in the Elephants, Flowers, Horses, and Mountains classes.

For the UC Merced Land Use Dataset (2100 samples), the proposed method was compared to the results of a study by Ugale and Railkar (2025). In that study, five models were evaluated: Local Ternary Pattern (LTP), Multiscale Adaptive Local Ternary Pattern (MALTP), ResNet50, and hybrid combinations integrating MALTP with ResNet50. Further, two main configurations were investigated. These are ResNet50 for shape feature extraction, and a hybrid model combining texture features extracted using MALTP with shape features extracted using ResNet50.

Based on the reported results, the proposed model outperforms the standalone ResNet50 model. Specifically, it achieved a precision of 0.967, whereas ResNet50 achieved a precision of 0.95, indicating the effectiveness of the adopted feature representation strategy.

However, when texture and shape features are combined, the MALTP + ResNet50 hybrid model significantly outperforms the proposed method, achieving a precision of 0.98. This result highlights the advantage of multi-feature fusion, where integrating complementary texture and shape descriptors enhances discriminative capability and improves the overall classification performance.

Moreover, Fig. 7 compares the performance of the proposed method with previous approaches on the Corel 1K dataset. It is evident that the proposed method outperforms all prior studies in most cases. However, there are instances of equal performance in certain classes, such as Buses, Dinosaurs, Flowers, and Horses. Specifically, in the Bus class, the proposed method achieves results equal to those reported in (Kannagi & Lanke, 2022). For the Dinosaurs class, the proposed method shows parity with both (Kannagi & Lanke, 2022)

and (Hassan et al., 2023). Similarly, in the Flowers class, the proposed method's performance matches that of the studies Pardede et al. (2019) and Kannagi and Lanke (2022).

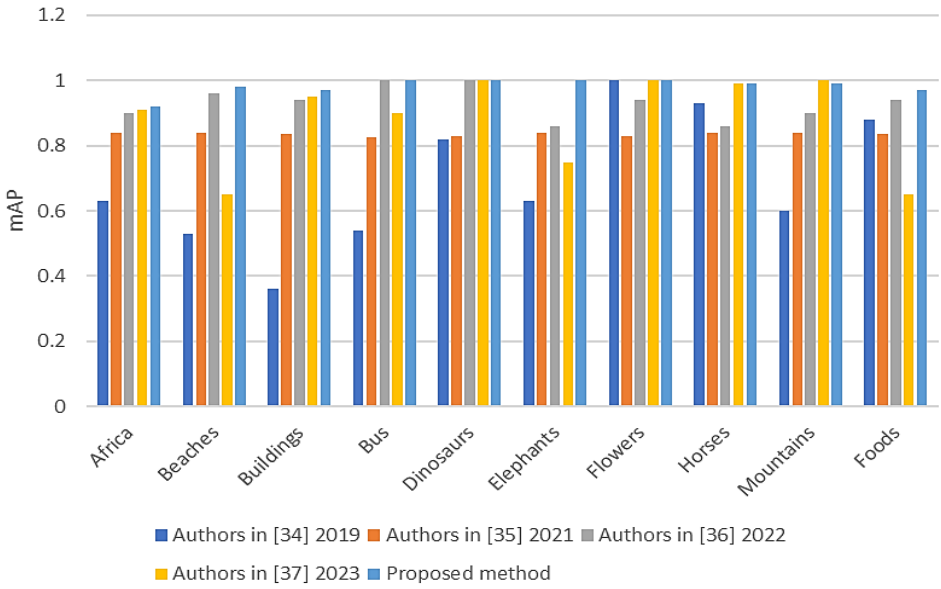


**Fig. 7. Performance comparison of the proposed method and previous approaches for the Corel 1K dataset**

In terms of Mean Average Precision (mAP) as a key metric, the proposed method scores are consistently high across all categories. They range from 0.92 to 1.0, which indicates robust generalization across different image types. The proposed method also demonstrates superior performance across all datasets by achieving the highest mAP (0.98) and outperforming all existing methods in every single category. It improved by 5.04%, 17.22%, 11.36% and 42.03% over Kannagi and Lanke (2022), Desai et al. (2021), Hassan et al. (2023) and Pardede et al. (2019), respectively.

The analysis reveals that while some previous studies outperformed others in specific classes, they underperformed them in different contexts. In contrast, the proposed method demonstrates consistent competitiveness across nine classes, either matching or surpassing the accuracy of prior studies. An exceptional case is observed in the Mountain class in (Hassan et al., 2023), where the accuracy reaches 1.00, slightly exceeding the proposed method's accuracy of 0.99. Overall, these findings highlight the robustness and superior performance of the proposed approach across diverse categories.

For more investigation, a Friedman test has been implemented to compare the performance of five competing methods across ten benchmark datasets that are mentioned in Table 3. So, the hypotheses are declared as follows:

Null hypothesis ($H_0$): There is no significant difference among the performances of the methods.

Alternative hypothesis ($H_1$): There is a significant difference among the performances of the methods.

The results have shown that there is a significant difference among the evaluated methods across datasets. The obtained Friedman chi-square value of $\chi^2(4) = 22.75$ and a p-value of 0.000142, which is less than 0.05. The mAP of the proposed method was the highest one, so the Alternative hypothesis ($H_1$) is accepted, and the null hypothesis ($H_0$) is rejected.

Moreover, the Kendall W test has been applied to show the rank of the proposed method compared to the other methods in the literature in terms of mAP. As shown in Table 8, the proposed method has the highest rank with a w value equal to 0.569.

**Tab. 8. Ranking of the proposed method vs. other methods in terms of *mAP* for the Corel 1K dataset**

| Method | Mean Rank | W |
|---|---|---|
| Proposed method | 4.55 | 0.569 |
| Ref. (Kannagi & Lanke, 2022) | 3.4 | |
| Ref. (Hassan et al., 2023) | 3.35 | |
| Ref. (Desai et al., 2021) | 2 | |
| Ref. (Pardede et al., 2019) | 1.7 | |

### 3.5. Clustering results analysis

In this subsection, the accuracy of image clustering is computed. The dataset in this study is divided into 10 classes, each of which has 100 images. The proposed method clustered the entire dataset into 10 classes using an optimized k-means algorithm based on their selected features (10, 15, 30). Therefore, the metric of evaluation for such a type of clustering (labeled clustering) is called clustering purity. Purity quantifies the extent to which the cluster Ci points only to one ground truth partition. Purity lies between 0 and 1 (1 indicates the highest purity, whereas 0 is the lowest). The purity metric is computed based on Equation (16).

$$purity = \frac{1}{N}\sum_{i=1}^{k} max_j \left|c_i \cap t_j\right| \tag{16}$$

Where *N* indicates the number of data points, *k* refers to the number of clusters, $c_i$ is a cluster in C. and $t_j$ denotes the classification with the total number for the class $c_i$. Table 9 presents the clustering purity results across all dimensional settings (10, 15, and 30) for the datasets used. The evaluation is conducted using both the standard k-means algorithm and the optimized k-means variant enhanced with Particle Swarm Optimization (PSO) to provide a comparative analysis of their performance. It can be seen that there is a significant improvement in using optimized k-means over the standard k-means algorithm. For example, for the Corel 1K dataset, the clustering purity results of standard k-means are 88% and 79% for 15 and 30 features, respectively. Comparatively, the results of optimized k-means are 96% and 97% for the same features, respectively. In addition, the highest clustering purity is achieved when the number of features is 30, where the result is 97% in the optimized k-means algorithm for the Corel 1K dataset and 69% for the UC Merced Land Use dataset. The improvement reflects the effectiveness of applying the PSO algorithm for selecting the centers of the clusters more accurately than when using the standard k-means algorithm.

**Tab. 9. Results of clustering purity for the Corel 1K dataset and UC merced land use dataset**

| dataset | Dim=10 using Standard k-means | Dim=15 using Standard k-means | Dim=30 using Standard k-means | Dim=10 using optimized clustering | Dim=15 using optimized clustering | Dim=30 using optimized clustering |
|---|---|---|---|---|---|---|
| Corel 1K | 91% | 88% | 79% | 96% | 96% | **97%** |
| UC Merced Land Use | 60% | 61% | 66% | 60.6% | 62% | 69% |

## 4. CHALLENGES AND LIMITATIONS

The proposed framework implementation encountered multiple challenges during its execution. The first challenge arose because VGG16 deep features extraction produced high-dimensional data, which made clustering processes more resource-intensive. The problem was solved through the application of PCA, which enabled the reduction of feature dimensionality. The optimal principal component count required extensive testing because excessive component reduction resulted in information loss, while minimal reduction kept computational expenses at an elevated level. The combination of K-means and PSO resulted in higher computational requirements for the system. The PSO method establishes better initial centroid positions while improving the method's ability to search the entire space. Yet, the method's continuous improvement process requires more time than standard K-means algorithms. The hybrid system performance depended on the

specific PSO parameter values, which included swarm size and inertia weight, thus requiring scientists to conduct experiments for establishing stable convergence points. In addition, one possible limitation could be that the robustness of the proposed method has not been evaluated with respect to variations in image resolution, noise, or class imbalance.

## 5. CONCLUSIONS

An improved image retrieval method has been proposed for retrieving an image from a dataset by depending on a deep learning model called the VGG16 architecture. The VGG16 CNN model has been utilized to extract the features of an image and then represent the extracted features as vectors. Consequently, the extracted features are reduced using the principal component analysis (PCA) algorithm with three feature (dimension) settings: 10, 15, and 30. Next, PSO enhances K-means by providing better initial centroids and exploring the solution space more effectively to avoid local optima. From a theoretical standpoint, K-means is sensitive to initial centroid positions and can get stuck in poor local minima. By integrating PSO, which is a global optimization algorithm, with K-means, the clustering process is enabled to explore a wider range of solutions and converge to more stable and meaningful clusters. This synergy between local convergence (K-means) and global exploration (PSO) forms the basis of the improved clustering quality observed in our experiments. The reduced features were used to retrieve the most similar images to the query image, based on the Squared Euclidean Distance algorithm within each cluster. The performance of the proposed method has been evaluated based on the Corel 1k and UC-Merced Land Use datasets in terms of mean average precision (m*AP*), clustering purity, recall, f-score, NDCG, and run time. The results show that the VGG16-PCA-PSO-K-means model with 30 features outperforms the VGG16-PCA-K-means model, achieving a clustering purity of 97% on the Corel 1K dataset and 69% on the UC Merced Land Use dataset. The results have also demonstrated that the VGG16-PCA-PSO-K-means model with 15 features has yielded better performance than the VGG16-PCA -K-means model with a clustering purity of 96% compared to 88% for the Corel 1K dataset and 62% compared to 61% for the UC Merced Land Use dataset. In addition, the results show that the VGG16-PCA-PSO-K-means model with 10 features outperforms the VGG16-PCA-K-means model, achieving a clustering purity of 96% compared to 91% for the Corel 1K dataset and 60% compared to 60.6% for the UC Merced Land Use dataset. The results obtained have demonstrated that using the VGG16 pre-trained deep neural network for feature extraction with hybrid PSO-K-means algorithms has led to significant improvement in CBIR performance across various image categories, compared to other existing methods. The improvement ranged between 5% and 18% across different feature sizes, with *mAP*@10 of 97.5% for the 10-class retrieval of the Corel 1K and *mAP*@10 of 96.7% for the 21-class retrieval of the UC Merced Land Use datasets, depending on 30 features only. Moreover, the results of applying the Kendall W test and the Friedman test have demonstrated that the (VGG16-PCA-PSO-K-means) model attained an advanced rank among other methods in the literature, with a significant difference for the Corel 1K dataset. For future work, another optimization algorithm, such as the Artificial Humming algorithm (AHA), may be used to improve the performance of the K-means algorithm, thereby enhancing the performance of the CBIR system. Eventually, another dataset may be considered with more categories to enhance the generality of the CBIR system.


### Acknowledgments

*The authors would like to thank the College of Science/Mosul University and the College of Information Technology/ Ninevah University for their support in conducting this study.*


### Conflicts of Interest

*There is no conflict of interest.*

### REFERENCES


Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Al-Kababchee, S. G. M., Algamal, Z. Y., & Qasim, O. S. (2023). Enhancement of K-means clustering in big data based on the equilibrium optimizer algorithm. *Journal of Intelligent Systems*, *32*(1), Article 20220230. https://doi.org/10.1515/jisys-2022-0230

Alkahya, M. A., Alreahan, H. O., & Algamal, Z. Y. (2023). Classification of breast cancer histopathological images using adaptive penalized logistic regression with the Wilcoxon rank sum test. *Electronic Journal of Applied Statistical Analysis*, *16*(3), 507–518. https://doi.org/10.1285/i20705948v16n3p507

Al-Mani, I. A., Al-Sabaawi, A. M. A., & Hussien, M. H. (2022). A review paper of model-based collaborative filtering techniques. In *Proceedings of the 2022 International Conference on Data Science and Intelligent Computing (ICDSIC)*. IEEE. https://doi.org/10.1109/ICDSIC56987.2022.10076148

Alnaish, Z. A. H., & Hasoon, S. O. (2023). Hybrid binary whale optimization algorithm based on taper-shaped transfer function for software defect prediction. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, *13*(1), 74–79. https://doi.org/10.35784/iapgos.4569

Alneamy, J. S., & Alnaish, R. A. H. (2022). A comparative study among some natural-inspired optimization algorithms. In *Proceedings of the 2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCITM56309.2022.10031877

Al-Sabaawi, A. M. A., Karacan, H., & Yenice, Y. E. (2021). SVD++ and clustering approaches to alleviating the cold-start problem for recommendation systems. *International Journal of Innovative Computing, Information and Control*, *17*(2), 383–396. https://doi.org/10.24507/ijicic.17.02.383

Babenko, A., Slesarev, A., Chigorin, A., & Lempitsky, V. (2014). Neural codes for image retrieval. In *Computer Vision – ECCV 2014: 13th European Conference*, *Proceedings, Part I* (pp. 584–599). Springer. https://doi.org/10.1007/978-3-319-10590-1_38

Bimantara, I. M. S., & Widiartha, I. M. (2023). Optimization of K-Means clustering using particle swarm optimization algorithm for grouping traveler reviews data on TripAdvisor sites. *Jurnal Ilmiah Kursor, 12*(1), 1–10. https://doi.org/10.21107/kursor.v12i01.269

Blanchard, E. E., Oner, B., Allgood, A., Peterson, D. T., Zengul, F. D., & Brown, M. R. (2024). Evolution of simulation scholarship: A text mining exploration. *Clinical Simulation in Nursing*, *96*, Article 101620. https://doi.org/10.1016/j.ecns.2024.101620

Cao, Z., Shaomin, M., Yongyu, X., & Dong, M. (2018). Image retrieval method based on CNN and dimension reduction. In *Proceedings of the 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. IEEE. https://doi.org/10.1109/SPAC46244.2018.8965601

Chen, R., Pan, L., Zhou, Y., & Lei, Q. (2020). Image retrieval based on deep feature extraction and reduction with improved CNN and PCA. *Journal of Information Hiding and Privacy Protection*, *2*(2), 67–75. https://doi.org/10.32604/jihpp.2020.010472

Dai, L., Johar, M. G. M., & Alkawaz, M. H. (2024). Semi-supervised medical image segmentation via frequency attention with DCT and data exchange: The FAS-Net approach. *Journal of Logistics, Informatics and Service Science*, *11*(11), 178–195. https://doi.org/10.33168/JLISS.2024.1111

Desai, P., Pujari, J., Sujatha, C., Kamble, A., & Kambli, A. (2021). Hybrid approach for content-based image retrieval using VGG16 layered architecture and SVM: An application of deep learning. *SN Computer Science*, *2*(3), Article 170. https://doi.org/10.1007/s42979-021-00529-4

Gautam, G., & Khanna, A. (2024). Content-based image retrieval system using CNN-based deep learning models. *Procedia Computer Science, 235*, 3131–3141. https://doi.org/10.1016/j.procs.2024.04.296

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2021). Content-based image retrieval based on convolutional neural networks. In *Proceedings of the 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICICIS52592.2021.9694146

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2022). Image retrieval based on deep learning. *Journal of System and Management Sciences*, *12*(2), 477–496.

Greenacre, M., Groenen, P. J. F., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), Article 100. https://doi.org/10.33168/JSMS.2022.0226

Hasoon, S. O., & Jasim, Y. A. (2013). Diagnose window problems based on hybrid intelligence systems. *Journal of Engineering Science & Technology (JESTEC), 8*(5), 566–578.

Hassan, R. Q., Sultani, Z. N., & Dhannoon, B. N. (2023). Content-based image retrieval based on the Corel dataset using deep learning. *International Journal of Artificial Intelligence*, *12*(2), 8938.

Hussein, M. H., Nawaf, H. N., & Bhaya, W. S. (2017). Exploiting the shared neighborhood to improve the quality of social community detection. In *Proceedings of the 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)*. IEEE. https://doi.org/10.1109/NTICT.2017.7976121

Jan, S. L., & Shieh, G. (2025). An improved nonparametric test and sample size procedures for the randomized complete block designs. *Sankhya B*, 1–26. https://doi.org/10.1007/s13571-025-00362-2

Jeunen, O., Potapov, I., & Ustimenko, A. (2023). *On (normalised) discounted cumulative gain as an off-policy evaluation metric for top-n recommendation*. *ArXiv, abs/2307.13555*. https://doi.org/10.48550/arXiv.2307.13555

Kannagi, A., & Lanke, R. (2022). Image retrieval based on deep learning-convolutional neural networks. In *Proceedings of the 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*. IEEE. https://doi.org/10.1109/IIHC55949.2022.10060450

Krizhevsky, A., & Hinton, G. (2009). *Learning multiple layers of features from tiny images* [Technical report]. University of Toronto.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics, 10*(2), 226–245. https://doi.org/10.1198/108571105X46642

Lian, X., Xia, N., Dai, G., & Yang, H. (2025). An efficient joint training model for monaural noisy-reverberant speech recognition. *Applied Acoustics*, *228*, Article 110322. https://doi.org/10.1016/j.apacoust.2024.110322

Maazalahi, M., & Hosseini, S. (2024). K-means and meta-heuristic algorithms for intrusion detection systems. *Cluster Computing*, *27*(8), 10377–10419. https://doi.org/10.1007/s10586-024-04510-7

Mohammed, N. T., Hussein, M. H., & Rashid, A. J. (2020). PAM clustering aided Android malicious apps detection. In *IOP Conference Series: Materials Science and Engineering*, *928*, 032041. IOP Publishing. https://doi.org/10.1088/1757-899X/928/3/032041

Mon, A. N., Pa, W. P., & Thu, Y. K. (2017). Exploring the effect of tones for Myanmar language speech recognition using a convolutional neural network (CNN). In *Proceedings of the International Conference of the Pacific Association for Computational Linguistics*. Springer. https://doi.org/10.1007/978-981-10-8438-6_25

Palla, M., & Karra, R. (2025). A comparative study of pre-trained CNN models with transfer learning for content-based image retrieval. *Engineering, Technology & Applied Science Research*, *15*(4), 25820–25826. https://doi.org/10.48084/etasr.11496

Pardede, J., Sitohang, B., Akbar, S., & Khodra, M. L. (2019). Improving the performance of CBIR using an XGBoost classifier with deep CNN-based feature extraction. In *Proceedings of the 2019 International Conference on Data and Software Engineering (ICoDSE)*. IEEE. https://doi.org/10.1109/ICoDSE48700.2019.9092754

Ran, X., Suyaroj, N., Tepsan, W., Ma, J., Zhou, X., & Deng, W. (2024). A hybrid genetic-fuzzy ant colony optimization algorithm for automatic K-means clustering in urban global positioning systems. *Engineering Applications of Artificial Intelligence*, *137*, Article 109237. https://doi.org/10.1016/j.engappai.2024.109237

Saeed, A. Q., Aldulaimi, M. H., Ismail, I. A., Ahmed, I. M., Yahya, Y. A., Kharma, Q. M., & Ghazal, T. M. (2024). Integrating three machine learning algorithms in an ensemble learning model for improving content-based spam email recognition. *Journal of Soft Computing and Data Mining*, *5*(2), 188–196. https://doi.org/10.30880/JSCDM.2024.05.02.014

Saritha, R. R., Paul, V., & Kumar, P. G. (2019). Content-based image retrieval using a deep learning process. *Cluster Computing*, *22*(2), 4187–4200. https://doi.org/10.1007/s10586-018-1731-0

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. *ArXiv, abs/1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Sudhish, D. K., & Nair, L. R. (2024). Content-based image retrieval for medical diagnosis using fuzzy clustering and deep learning. *Biomedical Signal Processing and Control*, *88*, Article 105620. https://doi.org/10.1016/j.bspc.2023.105620

Ugale, P., & Railkar, P. (2025). Content-based remote sensing image retrieval based on two-way feature representation using ResNet50 and modified multiscale local ternary pattern. *Mathematical Modelling of Engineering Problems*, *12*(2), 657–668. https://doi.org/10.18280/mmep.120902

Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: An overview. *Soft Computing*, *22*(2), 387–408. https://doi.org/10.1007/s00500-016-2474-6

Wang, X., Du, Y., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., & Han, X. (2023). RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval. *Medical Image Analysis*, *83*, Article 102645. https://doi.org/10.1016/j.media.2022.102645

Younus, Z. S., Mohamad, D., Saba, T., Alkawaz, M. H., Rehman, A., Al-Rodhaan, M., & Al-Dhelaan, A. (2015). Content-based image retrieval using PSO and the k-means clustering algorithm. *Arabian Journal of Geosciences*, *8*(8), 6211–6224. https://doi.org/10.1007/s12517-014-1584-7

Zhu, X., Zhang, Y., & Wang, F. (2016). Reducing the dimensionality of data with sparse autoencoder network. *Journal of Shenyang Ligong University*, *35*(1).