JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 02
2025

Research Article
https://doi.org/10.56979/902/2025

Content Based Image Retrieval using VGG19 KPCA and ELM

Anosha Iqbal[1], and Zahid Mehmood[1*]

[1]Department of Computer Engineering, University of Engineering and Technology, Taxila, 47080, Pakistan.
[*]Corresponding Author: Zahid Mehmood Email: Zahid.mehmood@uettaxila.edu.pk

Received: July 04, 2025 **Accepted:** August 20, 2025

Abstract: Content-Based Image Retrieval (CBIR) systems aim to retrieve visually similar images based on low-level image features. However, a major challenge in conventional CBIR methods is the semantic gap, disconnect between low-level visual features and the high-level semantic meaning perceived by humans. This research tackles the semantic gap and retrieval inefficiency by proposing a novel CBIR framework based on deep learning. The proposed method works like this: First, it uses a well-known convolutional neural network called VGG19 to extract rich, abstract visual features from images. Next, it uses an algorithm called Kernel Principal Component Analysis (KPCA) to reduce the dimensionality of the feature vectors, which are huge and difficult to manage. Finally, it uses another machine learning algorithm called Extreme Learning Machine (ELM) to classify images based on the features extracted by VGG19. The experimental details of the proposed technique indicate that it outperforms state of the art CBIR methods in terms of the performance evaluation metrics.

Keywords: Content-Based Image Retrieval (CBIR); Deep Feature Extraction; Kernel Principal Component Analysis (KPCA); Extreme Learning Machine (ELM)

1. Introduction

Content-Based Image Retrieval (CBIR) is a technology employed to search and retrieve images from a database having a large collection of images by examining their visual content instead of text-based metadata such as tags or descriptions. Visual features such as color, texture, shape, and spatial relationships are extracted from images in CBIR and utilized to compare and search for similar images. Unlike keyword-based searches done traditionally, which rely on manually provided descriptions, CBIR is based on computer-aided analysis of image data, and thus it becomes more accurate and scalable for big data. CBIR technology is found in a variety of applications such as medical imaging, digital libraries, surveillance, e-commerce, and remote sensing, in which accurate matching of images is important [1]. The CBIR procedure typically contains three primary steps: feature extraction, feature representation, and similarity matching. In the process of feature extraction, the system identifies and extracts significant features from the distance measures like Euclidean distance, cosine similarity, or Canberra distance. Methods like k-Nearest Neighbors (kNN) or XGBoost classifiers are generally applied for ranking and retrieving most similar images [2]. CBIR Suggested methodology combines state-of-the-art methods to maximize image retrieval accuracy and efficiency. It starts with the feature extraction using the VGG16 deep learning model, which extracts strong and distinctive feature aspects like color, texture, and shape. These high-dimensional feature aspects are optimized through Linear Discriminant Analysis (LDA), which minimizes dimensionality but retains maximum discriminative information, enhancing computational efficiency. These improved features are classified using XGBoost, a strong gradient boosting algorithm famous for precision and scalability. Finally, similarity metrics like cosine similarity or Euclidean distance are utilized to compare the query the image with database images, and returning and ranking the top relevant ones. This hybrid model overcomes the semantic gap by integrating deep learning and statistical

methods to provide enhanced accuracy and quicker retrieval. The method is planned for efficient management of large sets of data while offering accurate and scalable image retrieval for varied applications. The semantic gap, which is one of the key challenges in CBIR, denotes the dissimilarity between the low-level characteristics (like pixel-based attributes) that computers are able to comprehend and the high-level abstractions (like object classes) that humans perceive. For instance, two images can be of similar colors and forms but depict quite different scenes. Recent methods solve this problem through the integration of several techniques in order to enhance feature extraction and representation. While traditional CBIR techniques have relied on handcrafted features such as color histograms and texture descriptors, these methods often struggle to capture high-level semantics. This limitation has motivated a shift toward deep learning approaches, which offer more robust and semantically expressive feature representations. Recent CBIR strategies have increasingly integrated deep learning models and classical statistical techniques to overcome limitations such as the semantic gap and computational inefficiency. For instance, Fusion Net-Remote fuses CNNs with Random Forests to boost classification accuracy in remote sensing, whereas Spiking Neural Networks (SNNs) replicate human-like edge detection to enhance retrieval accuracy. Swin Transformers and contrastive learning architectures such as ResNet have been promising in identifying spatial hierarchies and fine-grained features, particularly in medical image analysis. Methods such as Adaptive GSM with Swin Notwithstanding these developments, problems like high computational expense, variability sensitivity (e.g., lighting, resolution, noise), and system sophistication remain impediments to real time scalability. Our suggested CBIR framework attempts to fill these niches by integrating deep semantic features with computationally inexpensive dimensionality reduction and classification models, presenting a scalable solution appropriate for large-scale image retrieval applications. In this article, we propose an effective and efficient. Content-Based Image Retrieval (CBIR) system that aims to bridge the semantic gap and enhance retrieval accuracy by leveraging a combination of deep learning and machine learning techniques. First, deep features are extracted from input images using the VGG19 convolutional neural network, which captures high-level semantic representations. To tackle the curse of dimensionality and enhance the quality of features, Kernel Principal Component Analysis (KPCA) is applied for non-linear dimensionality reduction. These optimized features are then classified using the Extreme Learning Machine (ELM), a fast and generalized single-layer feedforward network that ensures efficient model training and high classification performance. Finally, a similarity measure is employed to retrieve images most relevant to the query image. This combined pipeline efficaciously reduces feature redundancy, improves classification accuracy, and enhances semantic understanding in the retrieval process. When the proposed framework is employed for image classification and content-based image retrieval (CBIR), it provides the following benefits, which can be summarized as follows:

Transformers and Gradient-Structures Histogram (GSH) extend accuracy further by drawing on biological and statistical perspectives [3].

a) We have proposed an effective CBIR framework using VGG19 for feature extraction, KPCA for reducing the dimensionality, and ELM for fast and accurate classification.
b) Addressed the semantic gap by utilizing deep semantic features and robust classification techniques.
c) Demonstrated the effectiveness of the proposed method through comprehensive experiments on benchmark datasets, outperforming existing CBIR approaches in mean Average precision, recall, F1-score, and Confusion Matrix.

The remaining section of this article are organized as follows: section 2 presents related works of the state-of-the-art CBIR methods. The detail Methodology of the proposed technique is presented in section 3. Section 4 presents details of the datasets, performance metrices, experimental details, and performance comparison of proposed technique with state-of-the-art CBIR techniques. Section 5 conclude the proposed technique

2. Related works

Kayhan and Fekri *et al.*, [1] suggested methodology that combines texture and color features via a weighted decision model to enhance image retrieval accuracy. The procedure is split into two phases: feature extraction and similarity matching. At the first phase, Modified Local Binary Patterns (MLBP) and

Local Neighborhood Difference Patterns (LNDP) are utilized to extract the texture features, and a quantized color histogram is used to capture color details. Gaussian filters are applied to enhance texture representation for different luminance and noise scenarios. In the second phase, the system applies the Canberra distance measure to compare color and texture features independently and takes a weighted fusion strategy to fuse these similarities in order to improve matching. The proposed approach attained superior precision and recall when compared to current state-of-the-art methods. It solves the semantic gap problem by merging several feature types and assigning greater significance to texture, as perceived by humans. The method, however, raises computational cost due to the incorporation of several feature extraction techniques and fine tuning of the weighting parameters.

Alyahyan *et al.,* [4] proposes the Fusion Net-Remote model, which combines Convolutional Neural Networks (CNNs) with Random Forest (RF) for improved image classification in remote sensing tasks. The process has three steps: first, CNNs derive spatial features from multispectral image data; second, the features are fed to an RF classifier for decision; and third, performance is tuned by Hyperparameters adjustment, namely, learning rate, batch size, and dropout rate adjustments. The model performed best leading to the best testing accuracy, balanced F1-score among datasets. The paper resolves the issue of merging deep learning and ensemble learning to improve remote sensing image classification. Yet, challenges exist in terms of high computational needs, memory requirements, and latency brought about by the double-layer complexity of CNNs and RF, which poses a challenge to real-time deployment despite the use of optimization techniques.

İnçetaş and Arslan *et al.,* [5] proposes a bioinspired Spiking Neural Network (SNN) edge detection model for CBIR systems that simulates the human visual system (HVS) to provide more precise image retrieval. The process involves two stages: edge detection and image retrieval. The first stage utilizes a three-layer SNN model to obtain edge features by simulating how the HVS detects edges via synaptic connections. This model employs a 3×3 receptive field, minimizing computational complexity by 2.5 times than traditional SNN models. In the second stage, the extracted edges are incorporated into three CBIR approaches (GSH, EQBTC, and DLTCoP-CH), taking the place of traditional edge detection algorithms such as Sobel and Canny. The model enhances retrieval precision, with improvement in mean precision on the dataset over traditional techniques. It solves the problem of noise sensitivity in traditional edge detectors and improves accuracy by concentrating on biologically motivated edge detection. The model's limitation, however, is that it relies on linear measurements for feature vectors, keeping it from registering the performance levels of state-of-the-art machine-learning-based methods.

Kittichai *et al.,* [6] suggests a deep contrastive learning-based method for Content-Based Image Retrieval (CBIR) with pre-trained CNN backbones (ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNeXt50). The methodology contains two significant phases: image preprocessing and deep learning model-based feature extraction first, and Uniform Manifold Approximation and Projection (UMAP) for reducing the dimension of high-dimensional feature vectors for visualization second; the second phase performs the CBIR task by conducting a k-nearest neighbor (kNN) search (k=12) to find images according to feature vector similarity. The method showed better performance with regard to accuracy and precision, especially in detecting Anaplasma marginale infections. It resolves the issue of precise diagnosis from microscopic images in veterinary medicine but suffers from limitations due to dataset variance and visual class differentiation.

Rajender and Gopalachari *et al.,*[2] presents a hybrid approach that integrates Adaptive Granular Statistical Method (Adaptive-GSM) for dimensionality reduction and a Swin Transformer-based Deep Neural Network (DNN) for classification. The approach is split into two phases: Feature Extraction and Dimensionality Reduction. During the first stage, raw image data undergoes preprocessing in the form of RGB color histograms and Histogram of Oriented Gradients (HOG) to extract meaningful features. During the second stage, Adaptive GSM is utilized to decrease the dimensionality while maintaining important information. The truncated features are then passed through a Swin Transformer-based DNN, which enforces hierarchical learning and shifted window methods for precise image classification. It surpasses conventional techniques such as PCA and LDA, with a classification accuracy. The method overcomes the difficulty of processing high-dimensional data while enhancing precision and computational efficiency. Nevertheless, it suffers from some drawbacks like higher computational complexity with the transformer-based learning and needs fine-tuning of Hyperparameters for the best performance. Yuan and Liu et *al.,* [3]

introduces a Gradient-Structures Histogram (GSH) method for content-based image retrieval (CBIR) by mimicking the human brain's orientation selection and color perception processes. The methodology is based on two steps: feature extraction and image representation. In the first step, a gradient-structure detector extracts local features such as edges and bars of multiple widths and orientations in opponent-color space (red-green, blue-yellow, white-black). In the second step, these features are quantized into color, intensity, and orientation maps and mixed together into a 130-bin histogram without needing independent weight coefficients. The suggested approach attained accuracy on the dataset, surpassing the conventional methods such as LBP, BOW, and PUD. It solves the issue of correct description of spatial and structural features with ensuring computational efficiency. However, it addresses that balancing color, intensity, and edge orientation parameters remains difficult, and the method does not fully simulate the complexity of human visual processing.

3. Methodology

The proposed methodology employs a deep learning and machine learning framework for efficient image retrieval. Initially, VGG19 is used for feature extraction, leveraging its deep convolutional layers to capture detailed and high-level semantic features from both training and query images. These extracted features are often high-dimensional, so Kernel Principal Component Analysis (KPCA) is applied to reduce the dimensionality while preserving the non-linear structure of the data, resulting in optimal and compact feature representations. These reduced features are then fed into an Extreme Learning Machine (ELM), which acts as a fast and efficient classifier to distinguish between different image categories. Finally, the similarity between the query image and the database images is measured based on the classified features, and the most relevant images are retrieved. This proposed methodology effectively addresses the challenge of accurate and scalable image retrieval by combining the powerful representation capabilities of VGG19 with the speed and simplicity of KPCA and ELM. As illustrated in Figure 1, the flow diagram outlines the core stages of the proposed CBIR pipeline, including feature extraction, dimensionality reduction, classification, and retrieval.

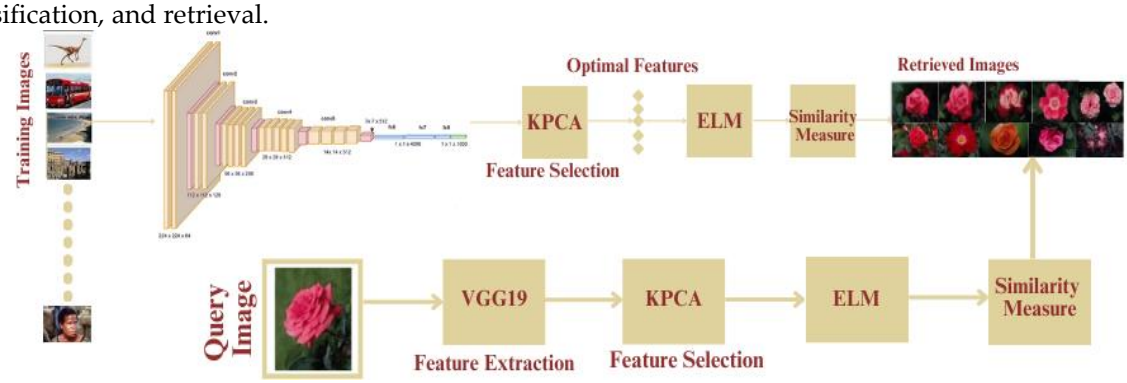

Figure 1. Flow diagram of proposed methodology

3.1. VGG19 for feature extraction

Proposed technique uses VGG19 is an advanced deep learning model designed for image classification and feature extraction. It has19 layers consisting of 16 convolutional layers and 3 fully connected layers. Selected technique using VGG19's primary role is to extract deep visual features from the input image, which are then processed through Kernel PCA for dimensionality reduction and ELM for classification [7].

The input image to VGG19 must be of a fixed size:

$$I \in \mathbb{R}^{224\times224\times3} \quad (1)$$

Where, 224×224 represents the spatial dimensions (height and width), 3 corresponds to the RGB color channels (Red, Green, Blue). If the input image is not $224 \times 224 \times 3$, it is resized to match this dimension. This term is known as Resizing. Each pixel value (originally in the range [0, 255]) is scaled between 0 and 1. This is normalization which is defined as,

$$I' = \frac{I}{255} \quad (2)$$

Mean subtraction is performed to reduce bias and improve model convergence. The mean pixel values of the ImageNet dataset are subtracted from each channel:

$$I'' = I' - \mu \quad (3)$$

Where the mean values are:

$$\mu = [123.68, 116.779, 103.939] \quad (4)$$

This ensures that the input is centered before feature extraction begins. VGG19 applies 16 convolutional layers, using small 3×3 filters with a stride of 1 and padding of 1. This maintains the spatial dimensions of the input. The convolution operation is given as:

$$X_{i,j}^{l} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} W_{m,n}^{l} \cdot X_{(i+m),(j+n)}^{(l-1)} + b^{l} \quad (5)$$

Where, $X_{i,j}^{l}$ Output feature map at layer l, $W_{m,n}^{l}$ is Weights of the convolutional kernel, $X_{(i+m),(j+n)}^{(l-1)}$ is Input from the previous layer, b^{l} is bias term

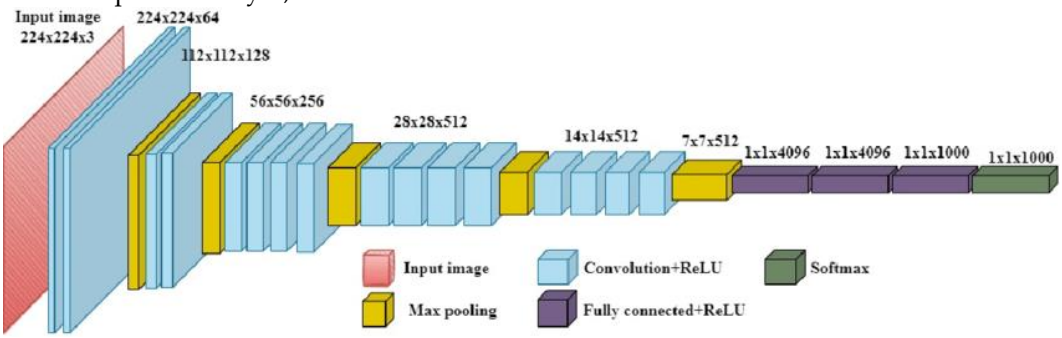

Figure 2. Visual overview of the VGG19 architecture used in the proposed Technique

As shown in Figure 2 the VGG19 used by the proposed architecture consists of multiple convolutional and fully connected layers designed to extract deep hierarchical features from the input images. ReLU (Rectified Linear Unit) is an activation function defined as $f(x) = \max(0, x)$ which introduces non-linearity while maintaining computational simplicity. After the convolution operation, after every two or four convolutional layers, max pooling is applied to reduce the spatial dimensions while retaining the most significant features. The max pooling operation is given as:

$$P(i,j) = \max\{X(2i, 2j), X(2i+1, 2j), X(2i, 2j+1), X(2i+1, 2j+1)\} \quad (6)$$

Where, $p(i,j)$ is the pooled output, the pooling window is 2×2 with a stride of 2, reducing feature map dimensions by half.

Table 1. Layers used by the proposed technique

Block	Layers	***Input → Output***
Block1	Conv1_1, Conv1_2 + maxPooling	$224 \times 224 \times 64 \rightarrow 112 \times 112 \times 64$
Block2	Conv2_1, Conv2_2 + maxPooling	$112 \times 112 \times 128 \rightarrow 56 \times 56 \times 128$
Block3	Conv3_1, Conv3_4 + maxPooling	$56 \times 56 \times 256 \rightarrow 28 \times 28 \times 256$
Block4	Conv4_1, Conv4_4 + maxPooling	$28 \times 28 \times 512 \rightarrow 14 \times 14 \times 512$
Block5	Conv5_1, Conv5_4 + maxPooling	$14 \times 14 \times 512 \rightarrow 7 \times 7 \times 512$

After the convolutional blocks, the feature map $7 \times 7 \times 512$ represented in Table 1 , is flattened into a 1D vector of length. This process is mathematically represented as:

$$F_{flattened} = 7 \times 7 \times 512 = 25088 \quad (7)$$

This vector is passed through three fully connected layers (FC):

$$z_i = W_i \cdot F_{i-1} + b_i \quad (8)$$

FC1: 4096 neurons → Output: $F_{fc1} \in \mathbb{R}^{4096}$

FC2: 4096 neurons → Output: $F_{fc2} \in \mathbb{R}^{4096}$

FC3: 1000 neurons (for ImageNet classes, but discarded in proposed system).

For proposed technique, the output of FC2 (a 4096-dimensional feature vector) is extracted and passed to Kernel PCA for dimensionality reduction. VGG19 captures multi-level features, low-level features (edges, textures) in initial layers and high-level features (object structure) in deeper layers. The uniform 3 × 3 filter and consistent architecture provide stable and detailed feature extraction. The 4096-dimensional output vector from FC2 provides a comprehensive description of image content.

3.2. Kernel principal component analysis (KPCA)

After feature extraction through VGG19, the output is a 4096-dimensional feature vector. To reduce computational complexity while retaining meaningful information, Kernel Principal Component Analysis

(KPCA) is used by the proposed technique. KPCA extends the standard PCA by using the "kernel trick" to map data into a higher-dimensional space where linear methods can identify patterns. The output of proposed VGG19 is a 4096-dimensional feature vector for each image is represented as:

$$F = [f_1, f_2, f_3, \dots., f_{4096}] \quad (9)$$

If dataset contains n images, these features can be represented as a matrix X:

$$F = [F_1, F_2, F_3, \dots., F_n] \in \mathbb{R}^{n \times 4096} \quad (10)$$

Where, n is number of images (samples), 4096 is feature dimensions from VGG19.

Standard PCA is limited to linear transformations, meaning it only captures relationships between features if they form a straight line. However, image data from deep learning models like VGG19 often exhibits complex, non-linear patterns. Kernel PCA addresses this by using a kernel trick to map the data into a higher-dimensional feature space where non-linear relationships become linear.

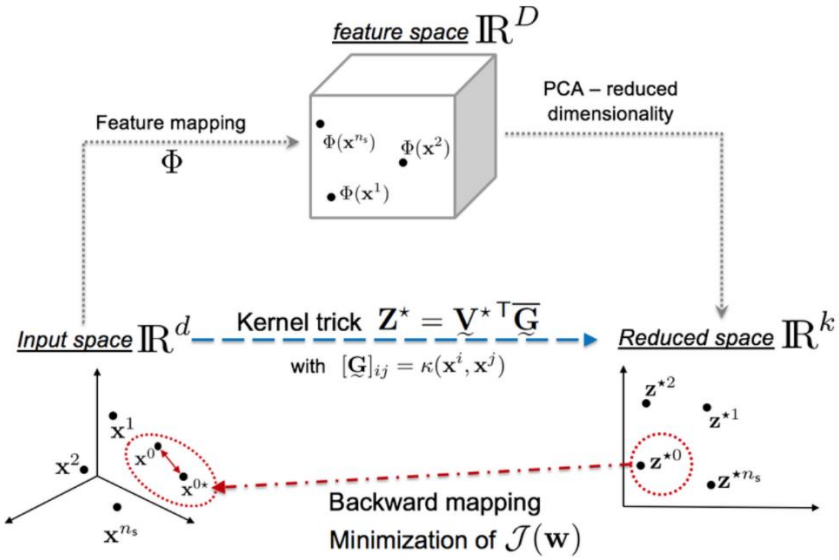

Figure 3. Illustration of the Proposed Kernel PCA (KPCA) architecture

Figure *3* illustrates the Kernel PCA process, where high-dimensional features from VGG19 are projected into a lower-dimensional space using a linear kernel to enhance feature reduction and separability. In the above given representation of KPCA, to capture non-linear structures in the data, each feature vector x_i from the VGG19 output is implicitly mapped into a higher-dimensional space using a kernel function mapped using a kernel function $\phi(x_i)$ as described below:

$$\phi: x_i \to \phi(x_i) \quad (11)$$

However, we do not need to compute ϕ directly. Instead, we use a kernel function to find the inner product between pairs of transformed vectors by using:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (12)$$

This is computationally efficient because we avoid working directly in the high-dimensional space. In the proposed methodology uses KPCA, we are using linear kernel. In Kernel PCA (Principal Component Analysis), the choice of the kernel function determines how the data is mapped into a higher-dimensional space. This is particularly useful when dealing with non-linearly separable data, as standard PCA can only capture linear relationships.

3.3. Linear kernel in kernel PCA:

The proposed technique employs a linear kernel in KPCA for computational simplicity and stability, making it suitable for high-dimensional deep feature embeddings. Dot product kernel is defined as measuring linear similarity between feature vectors and forming the basis of the linear kernel used in KPCA. It assumes that the relationships between features are linear, meaning it captures straight-line patterns in the data. The equation shows as:

$$k(x_i \cdot x_j) = x_i \cdot x_j \quad (13)$$

Where, x_i and x_j are two input feature vectors. It Computes the dot product between feature vectors, Preserves the original structure of the feature space without transforming it. Suitable when the data lies on

or near a linear subspace (e.g., when VGG19 features already follow a linear pattern). It is computationally efficient and faster. Works well if the image features have linear relationships. Linear Kernel PCA cannot capture complex, non-linear patterns in the data.

3.4. Extreme learning machine (ELM)

The reduced-dimensional output from KPCA is then fed into the Extreme Learning Machine (ELM) for rapid and accurate image classification. Proposed technique uses ELM, is a single-hidden-layer feedforward neural network (SLFN) where the input weights and biases are randomly initialized, and the output weights are analytically computed using the Moore-Penrose pseudo-inverse. The output of proposed technique of Kernel PCA is a d-dimensional reduced feature vector for each image:

$$Z = [z_1, z_2, \dots., z_n] \in \mathbb{R}^{n \times d} \quad (14)$$

Where, n is the number of samples (images), d is the Reduced feature dimensions (from KPCA, e.g., 300 or 500). This feature vector z acts as the input to the Extreme Learning Machine. A proposed ELM model consists of three key layers, Input Layer receives the d-dimensional input vector from KPCA, Hidden Layer contains L neurons with random weights and biases, Output Layer computes the output weights using a closed-form solution for classification.

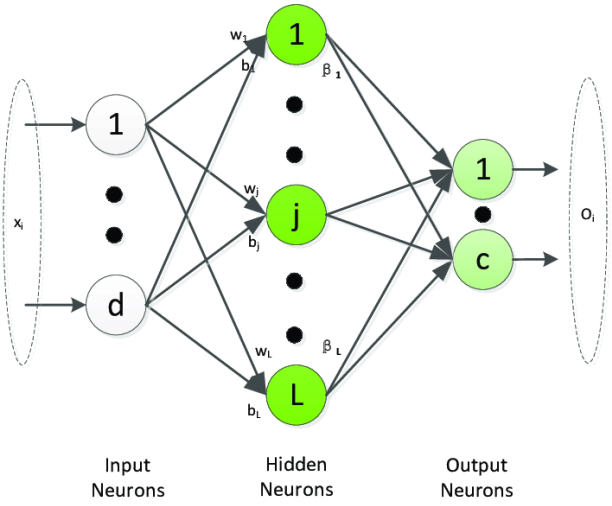

Figure 4. Structure of the Extreme Learning Machine (ELM) used by the proposed Technique

As depicted in Figure 4 of ELM used by the proposed technique, the input weights w and biases b are randomly generated and fixed during training. Let, $Z \in \mathbb{R}^{n \times d}$ is Input matrix (from KPCA), $W \in \mathbb{R}^{d \times L}$ is Random input weights, $b \in \mathbb{R}^{L}$ is Random bias vector, L *is* Number of hidden neurons. Each hidden neuron applies a non-linear activation function to the input data. For each input samplez_i, the hidden layer output H is calculated as:

$$H = g(Z \cdot W + b) \quad (15)$$

Where, H is hidden layer output matrix $(n \times L)$, $g(\cdot)$ is activation function (commonly sigmoid or ReLU).

Common activation functions used in ELM,

Sigmoid: $g(x) = \frac{1}{1+e^{-x}}$

ReLU (Rectified Linear Unit): $g(x) = \max(0, x)$

Each row in H represents the activation output for a sample across all hidden neurons. Instead of using backpropagation, ELM directly computes the output weights β using the Moore-Penrose pseudo-inverse:

$$\beta = H^{\dagger}T \quad (16)$$

Where, β is Output weights ($L \times m$, where mm is the number of classes), $H^{\dagger}$ is Moore-Penrose pseudo-inverse of H, T is Target matrix (class labels in one-hot encoding). The Moore-Penrose pseudo-inverse is calculated as:

$$H^{\dagger} = (H^T H)^{-1} H^T \quad (17)$$

This step is computationally efficient and provides a closed-form solution to minimize the error between predictions and actual outputs. During training, the ELM model only requires randomly initializing the input weights and biases, computing the hidden layer output H, finding the output weights β using the pseudo-inverse. This method is significantly faster because it does not require iterative weight updates. Once the model is trained, the reduced KPCA features of a new query image z_q are passed through the trained ELM [8]. Compute the hidden layer output for the query as:

$$H_q = g(z_q \cdot W + b) \quad (18)$$

Predict the class label by multiplying with the output weights:

$$\hat{y} = H_q \cdot \beta \quad (19)$$

Assign the class with the highest score (for multi-class problems):

$$Predicted\ Class = argmax(\hat{y}) \quad (20)$$

The output of the proposed ELM classifier is the predicted class or image category for each query image. In proposed CBIR technique, the ELM output can also be used to rank and retrieve the most relevant images from the database based on class similarity. ELM is extremely fast due to the absence of iterative weight updates. It can handle large datasets efficiently. Despite the random weights, ELM has strong generalization performance.

3.5. Similarity measure

Once the feature vectors are classified through the Extreme Learning Machine (ELM), the system proceeds to the similarity evaluation stage. At this point, the output vector from ELM serves as the query representation. This query vector is compared against the feature vectors of the images in the database using multiple similarity measures to determine which images are most alike. This step is crucial in the Content-Based Image Retrieval (CBIR) pipeline, as it directly affects retrieval accuracy and relevance.

3.6. Cosine similarity

Instead of distance, it measures the cosine of the angle between two vectors. It is particularly useful when the magnitude of vectors may vary but their orientation (direction) remains similar.

$$Cosine(x, y) = \frac{x \cdot y}{\|x\|\|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \quad (21)$$

Cosine similarity is widely employed when the vector length is not insightful, and direction alone is significant, particularly in high-dimensional information. The distance metrics used are crucial in that they provide a means of ensuring that the retrieval system does not become biased toward any given feature dimension. By normalizing and reasonably assessing the level of dissimilarity with respect to all feature elements, these metrics permit an equitable comparison between the query image and the database entries. Most similar images those with the lowest distance scores or highest similarity values are chosen and ranked accordingly. This approach improves the system's capability to provide semantically and visually consistent results, thus improving retrieval accuracy. Additionally, the proposed framework is versatile to deal with varying feature representations, either obtained from deep learning models, transformed statistical descriptors. Due to its computational efficiency, the system stays scalable and efficient even when used on large image collections.

4. Results and Discussion

In the Results section, the performance of the proposed content-based image retrieval (CBIR) model is comprehensively tested on several benchmark datasets such as Corel-1K, Oxford Flower, Corel-1500, and OTScene. The combination of deep visual features learned with VGG19, Kernel PCA for nonlinear dimensionality reduction, and Extreme Learning Machine (ELM) classification has dramatically improved the retrieval performance. Recall, F1-score, and mean average precision (mAP) measures were utilized to provide a well-balanced comparison of retrieval accuracy against these heterogeneous datasets. While our model scores higher in performance metrics, it also offers better efficiency due to ELM's fast learning and KPCA's compression. The performance indicates a uniform enhancement in the retrieval efficiency, especially in the Corel datasets that include a broad variety of image categories, as well as OTScene, which provides more challenging outdoor and environmental images. These datasets diversity represents the resilience and ability to generalize of the suggested framework. The integration of deep feature extraction, dimensionality reduction, and ELM classification surpasses traditional methods, as can be seen through

greater mean Average precision, recall, F1-scores and Confusion Matrix across all datasets. Through tabular and graphical observations, it is clear that every part VGG19, Kernel PCA, and ELM plays an important role in contributing to the overall system performance, thereby providing useful insights for future development in intelligent image retrieval systems.

4.1. Datasets

The datasets considered in this research i.e., Corel-1K, Oxford Flower, Corel-1500, and OTScene, are each representing a variety of images, which are crucial in measuring the robustness and efficacy of the suggested content-based image retrieval (CBIR) system. Corel-1K depicted in Figure 5(a) is comprised of 1,000 images in 10 classes, e.g., animals, buildings, flowers and people. This dataset is commonly used for evaluating basic image retrieval systems due to its relatively smaller scale and controlled category distribution. The Oxford Flowers dataset consists of 8,189 images spanning 102 flower categories, providing a fine-grained benchmark for evaluating retrieval performance in visually similar and biologically related classes. Some sample images are represented by Figure 5(c). The Corel-1500 dataset demonstrates from Figure 5(b) further expands upon Corel-5K, with 1,500 images across 50 categories, which allows for more complex and large-scale evaluation, particularly useful in assessing the scalability of the retrieval system. Lastly, the few images of dataset in Figure 5(d) is OTScene, distinct from the Corel datasets, focuses on outdoor scenes, containing 1,500 images grouped into 15 categories, including urban and natural environments such as beaches, forests, and cities. This dataset is particularly useful for testing how well the retrieval system generalizes to more complex and varied scene types. These datasets collectively represent a wide range of image categories and complexities, making them ideal for testing the proposed CBIR framework's ability to handle diverse image retrieval tasks. Each dataset challenges the system in unique ways, testing not only the accuracy but also the robustness and scalability of the approach.

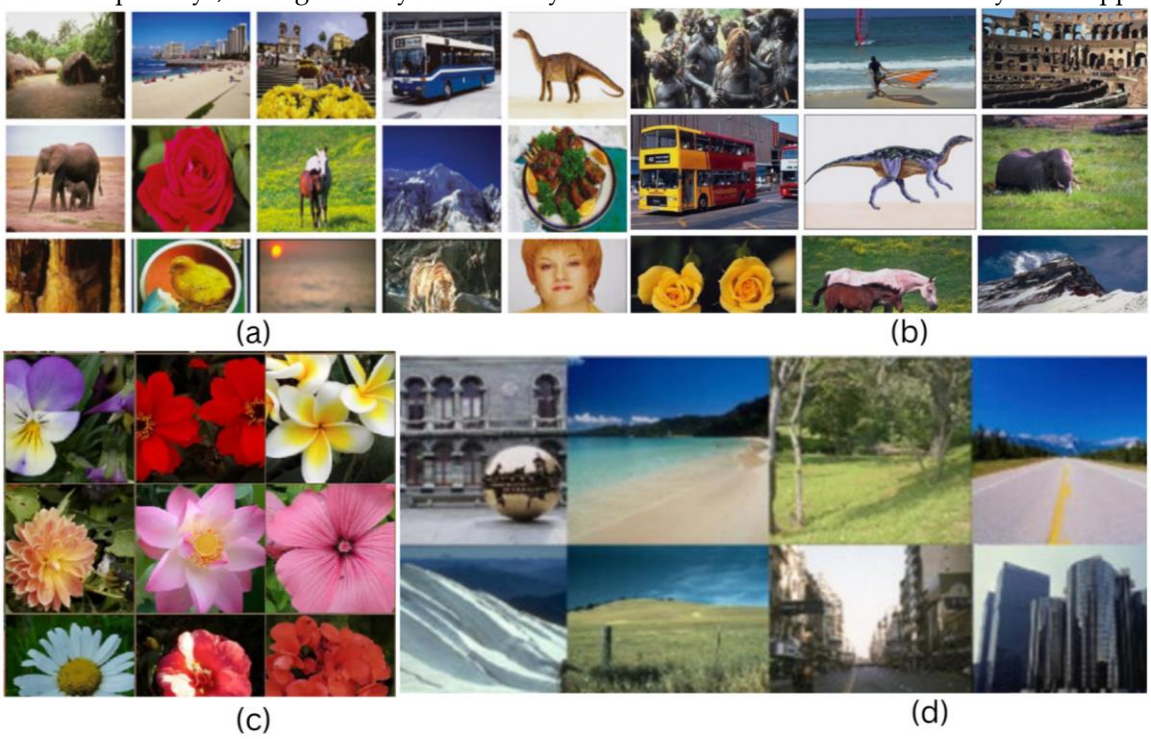

Figure 5. Sample Images: (a) Corel-1k, (b) Corel-1500, (c) Oxford Flower, (d) OTScene

4.2. Performance evaluation parameters

To systematically assess the retrieval performance of the proposed CBIR framework, several key evaluation metrics are employed. These metrics not only quantify the accuracy of retrieval results but also highlight the model's capability to distinguish between relevant and irrelevant images. Below is a detailed overview of the metrics used:

1. Precision

Precision measures the correctness of positive retrievals, it reflects how many of the retrieved images are actually relevant. High precision means fewer irrelevant images are retrieved.

$$Precision = \frac{I_r}{I_t} \quad (22)$$

Where I_r represents no. of correctly retrieved images, I_t represents no. of total retrieved images

This metric is crucial in CBIR systems where users expect mostly relevant images in the top results.

2. Recall

Recall, also referred to as sensitivity or true positive rate, and evaluates the system's ability to retrieve all relevant images from the database. A higher recall indicates fewer relevant images are missed.

$$Recall = \frac{I_r}{I_{dt}} \quad (23)$$

Where, I_r represents no. of correctly retrieved images, I_{dt} represents no of images in that class

While high recall ensures most relevant items are retrieved, it may also include some irrelevant ones if precision is low.

3. Average precision and mAP

Average Precision is the average of precision values computed at the ranks where relevant items are retrieved for a single query**.**

$$AP = \frac{1}{R}\sum_{k=1}^{n} P(k).rel(k) \quad (24)$$

Where, $P(k)$ is precision at rank k, $rel(k) = 1$ if the item at rank k is relevant, otherwise 0, R is the total number of relevant items for query

Mean Average Precision is the mean of Average Precision values across all queries. It gives an overall performance score of the retrieval system.

$$mAP = \frac{1}{Q}\sum_{q=1}^{Q} AP(q) \quad (25)$$

Where Q is the total number of queries, $AP(q)$ is the average precision for query q

4. F1-score

The F1-score provides a harmonic mean of precision and recall, balancing the trade-off between the two. It is particularly useful when there is an uneven class distribution or when both false positives and false negatives are equally important.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (26)$$

This single metric gives a more comprehensive view of retrieval effectiveness compared to using precision or recall alone.

5. Confusion Matrix

A confusion matrix is a table used to evaluate a classification model by showing the number of correct and incorrect predictions for each class. It helps identify where the model is performing well and where it is making mistakes.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

All of these measures have distinct roles in assessing the CBIR system. Precision and recall emphasize relevance of the retrieved items, and the F1-score compromises between the two to give an unbiased assessment. The confusion matrix gives a quantitative and visual summary of correct and wrong classification based on various classes. The combination of these gives a multi-dimensional assessment of the performance of the system on different data sets.

4.3. Performance analysis of the proposed technique on the Corel-1k

In the CBIR domain, the Corel-1K dataset is a widely adopted benchmark consisting of 1,000 images across 10 distinct semantic categories, each containing 100 images. The dataset includes images with resolutions of either 256×384 or 384×256 pixels. In this study, the performance of the proposed CBIR pipeline was evaluated using standard retrieval metrics: mean Average Precision (mAP), Recall, F1-score and Confusion Matrix. These metrics were computed to ensure a fair and robust comparison. The

experimental outcomes were analyzed at various retrieval levels to measure the effectiveness of the system in retrieving semantically relevant images. The proposed image retrieval framework, combining VGG19 for deep feature extraction, linear Kernel PCA for dimensionality reduction, and ELM for fast classification, achieves superior performance on the Corel-1K dataset due to the harmonized interaction between feature quality, transformation stability, and classifier generalization. VGG19 provides deep hierarchical representations that effectively capture object-level semantics, while linear KPCA preserves class-wise feature separability without introducing artificial curvature. The retrieval system was assessed using a linear kernel on retrieval performance. This structural integrity enables ELM to efficiently map features to correct categories with minimal training overhead. Performance detail of the proposed technique is presented in Table 2.

Table 2. Performance detail of the proposed technique on Corel-1K Dataset

Category	Precision	Recall	F1-Score
African People	89.00	17.80	29.66
Beach	81.00	16.20	27.00
Buildings	94.00	18.80	31.33
Buses	100.0	20.00	1.904
Dinosaurs	100.0	20.00	1.904
Elephants	100.0	20.00	1.904
Flowers	91.00	18.20	30.33
Foods	89.00	17.80	29.66
Horses	100.0	20.00	1.904
Mountains	91.00	18.20	30.33
mAP, Avg. Recall, Avg. F1-Score	93.00	18.70	18.59

The high performance on Corel-1K reflects the model's ability to leverage well-structured visual categories. Since the dataset is relatively small with distinct class boundaries, the model achieves strong results with minimal computational overhead and fast inference time. The linear variant strikes the right balance between transformation simplicity and discriminative preservation, allowing the proposed system to perform reliably across all Corel-1K categories. This makes it a more scalable and dependable choice for CBIR tasks involving medium-scale, semantically balanced datasets. In various categories of the Corel-1K dataset, the proposed pipeline shows enhanced precision and recall, particularly for visually distinct image such as beach. Figure 6 shows the top-20 retrieved images for a query from the “Beach” category. These results were ranked using Cosine similarity, where smaller distance values indicate a higher degree of content similarity with the query image.

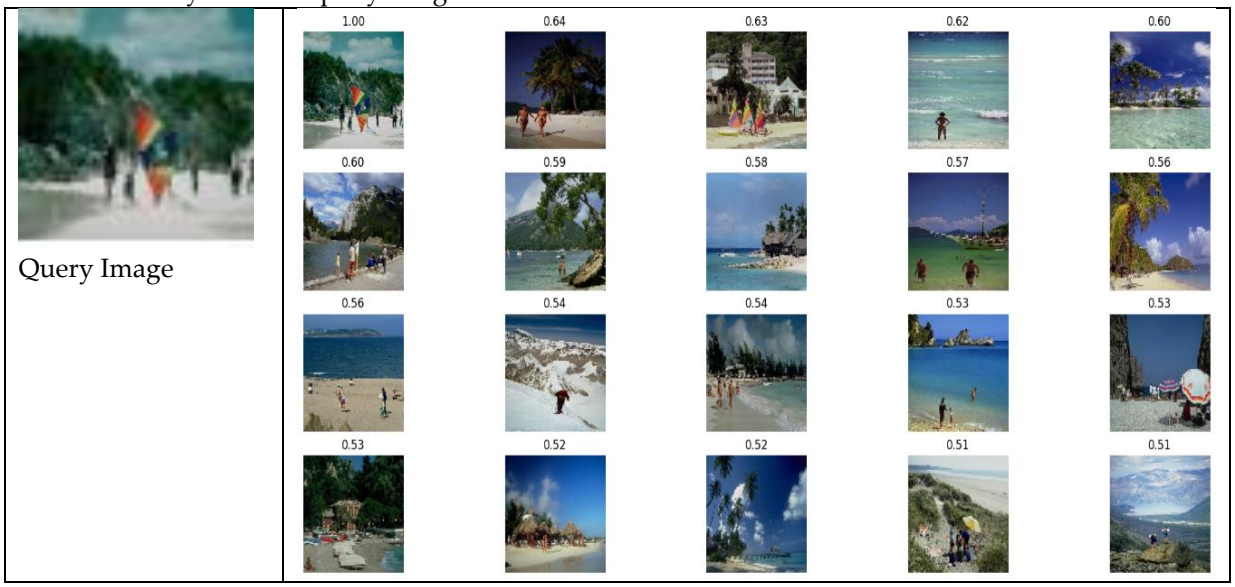

Figure 6. Top 20 retrieved images according to the visual contents of the query image on the Corel-1K dataset.

The suggested approach demonstrates enhanced retrieval performance on various evaluation criteria in comparison to other current CBIR methods. But even better statistical verification like confidence intervals or significance testing would further confirm these performance assertions. Unlike traditional approaches that suffer from low recall or imbalanced precision, the proposed framework maintains strong and stable performance. Its ability to effectively extract deep features, reduce dimensionality, and accurately classify images ensures better generalization and robustness across varied image categories. The combination of deep features, nonlinear dimensionality reduction, and a fast, generalizing classifier results in a highly efficient and scalable CBIR system presented from Table 3. This makes the proposed method particularly suitable for large-scale image retrieval tasks.

Table 3. Performance Comparison of the proposed technique with state-of-the-art CBIR techniques on Corel-1k dataset

Method	Precision	Recall	F1-Score
Raja *et al.* [9]	79.41	15.88	26.46
Kanarparthi *et al.* [10]	85.06	55.51	44.97
Hussain *et al.* [11]	93.04	18.61	31.02
Natarajan *et al.* [12]	74.21	14.84	24.73
Proposed Technique	**93.00**	**18.70**	**18.59**

The confusion matrix of the proposed technique for the Corel-1K dataset shown from Figure 7, provides a detailed overview of the model's classification performance across all image categories. Diagonal elements of the matrix represent the true positives, indicating the number of images correctly classified into their respective classes. Off-diagonal entries reflect misclassifications, highlighting the false positives and false negatives, which help in identifying specific classes where the model confuses one category with another.

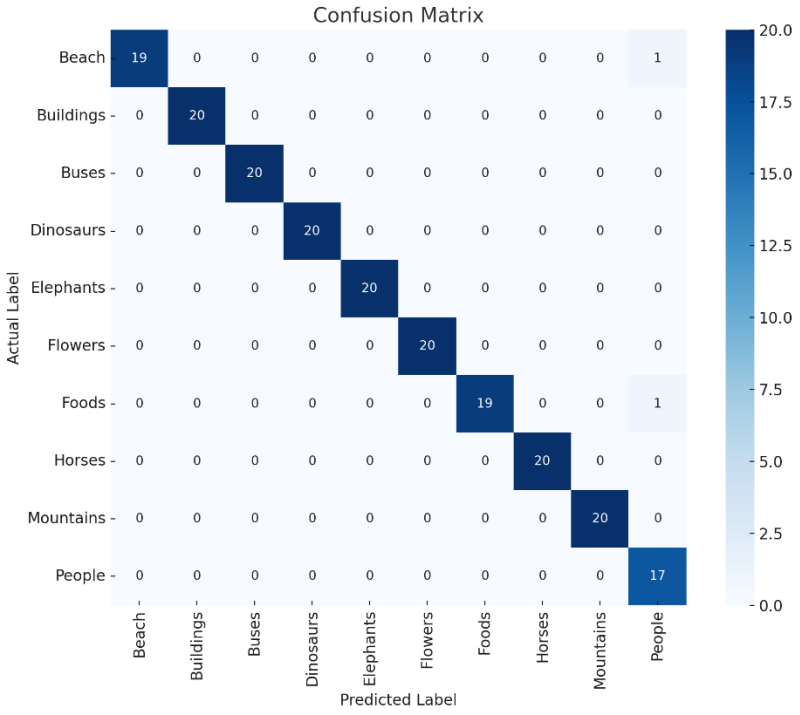

Figure 7. Confusion Matrix for Corel-1k Dataset

4.4. Performance analysis of the proposed technique on the Corel-1500 Dataset

The Corel-1500 dataset is a challenging benchmark for CBIR systems, consisting of 1,500 images distributed evenly across 15 semantic categories, with 100 images per class. Each of these images is of size either 256×384 pixels or 384×256 pixels. For this research, the performance of the suggested content-based

image retrieval pipeline was assessed by using mean Average Precision, Recall, and F1-score as presented in Table 4 . The evaluation measures were calculated by utilizing traditional formulas to assure consistency and replicability throughout retrieval experiments. This experimental configuration sought to explore the impact of kernel transformation on retrieval accuracy within the context of the introduced VGG19 + KPCA + ELM architecture. The data contains visually close but semantically different classes like bus, sunset, and woman that are helped by the clear separation provided by linear KPCA. This helps ELM preserve high retrieval precision even in overlapping visual environments.

Table 4. Performance detail of the proposed technique on Corel-1500 dataset

Category	Precision	Recall	F1-Score
Beach	70.00	14.00	23.33
Buildings	81.00	16.20	27.00
Bus	95.00	19.00	31.66
Cave	74.00	14.80	24.66
Dino	100.0	20.00	33.33
Eat_Feasts	83.00	16.60	27.66
Elephants	95.00	19.00	31.66
Flowers	100.0	20.00	33.33
Horse	100.0	20.00	33.33
Mountains	86.00	17.20	28.66
People	71.00	14.20	23.66
Postcard	100.0	20.00	33.33
Sunset	87.00	17.40	29.00
Tiger	89.00	17.80	29.66
Woman	100.0	20.00	29.33
mAP, Avg. Recall, Avg. F1-score	88.73	16.54	29.30

The slightly lower performance on Corel-1500 is due to increased class diversity, yet the model maintains good precision. Importantly, training time remains reasonable due to the reduced feature space achieved via KPCA, demonstrating scalability without excessive resource use. These observations confirm that for large and multi-class datasets like Corel-1500, linear KPCA provides a more stable and interpretable projection that complements the ELM classifier. The proposed pipeline thus ensures both efficiency and scalability, outperforming non-linear alternatives in scenarios where visual ambiguity and class overlap are prevalent.

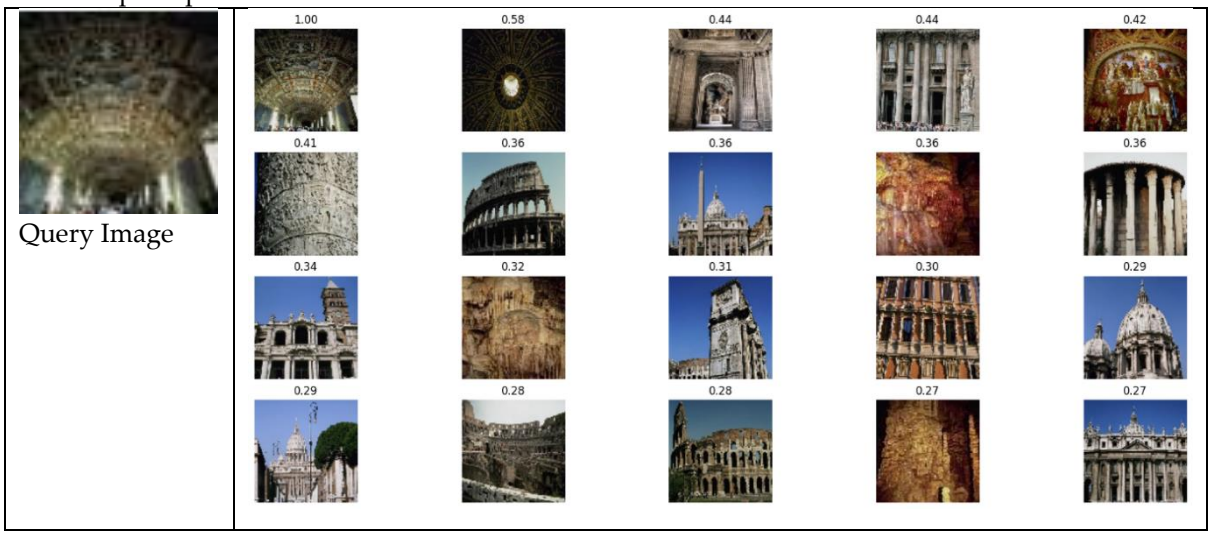

Figure 8. Top 20 retrieved images according to the visual contents of the query image on the Corel-1500 dataset.

Figure 8 showcases retrieval results for a query image from the "Buildings" category, where the system retrieved top-20 similar images ranked by cosine similarity. Comparative analysis in Table 5 confirms that the proposed method surpasses existing CBIR systems using traditional handcrafted features or standalone classifiers. It demonstrates the advantage of combining deep learning with non-linear dimensionality reduction and fast learning models.

Table 5. Performance Comparison of the proposed technique with state-of-the-art CBIR techniques on Corel-1500 dataset

Methods	Precision	Recall	F1-Score
Ali *et al.* [13]	72.60	14.52	24.20
Zeng *et al.* [14]	63.95	12.79	21.31
Ali *et al.* [15]	74.95	14.99	25.35
Proposed Technique	**88.73**	**16.54**	**29.30**

The confusion matrix of the proposed technique for the Corel-1500 dataset shows from Figure 9, demonstrate the model's capacity to differentiate among a greater number of image classes. It offers a good overview of accurate predictions on the diagonal and misclassifications on the off-diagonal terms. With 15 classes with different variations, this confusion matrix is useful for assessing class-specific accuracy, revealing confusion between visually equivalent categories, and informing improvements in feature representation and classification approach.

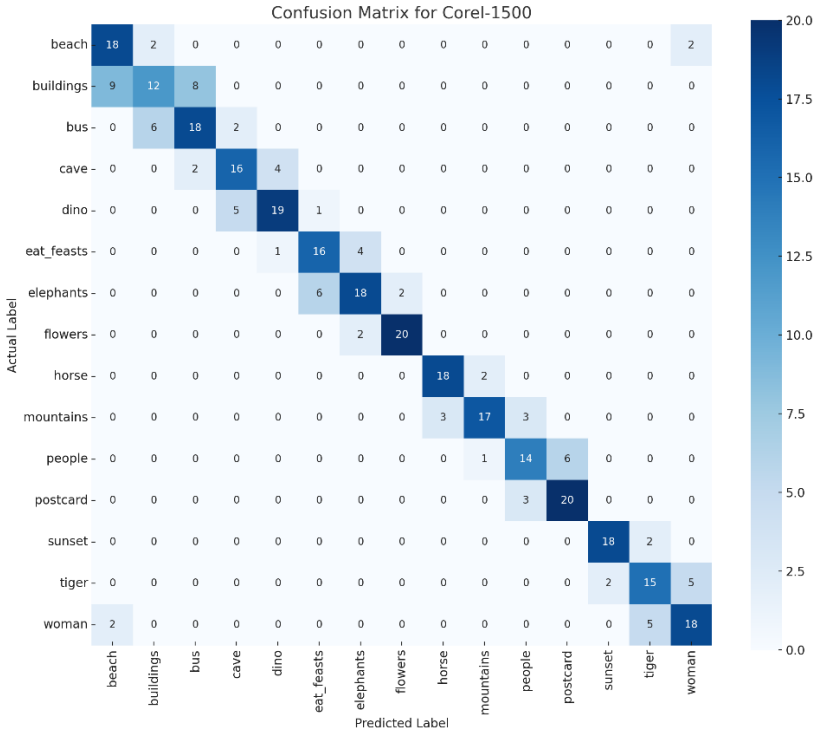

Figure 9. Confusion Matrix for Corel-1500

4.5. Performance analysis of the proposed technique on the OT-Scene Dataset

OT Scene dataset offers a challenging benchmark composed of diverse natural and urban scenes across eight semantic classes. Each class includes a varying number of images, amounting to a total of 538 samples. The images represent complex environmental contexts such as "open country," "coast," "forest," "highway," "inside city," "mountain," "street," and "tall buildings." The retrieval performance was rigorously evaluated using mean average precision, recall, F1-score and Confusion Matrix. The proposed pipeline, combining VGG19, linear KPCA, and ELM, demonstrates from Table 6 the superior performance by effectively preserving semantic structure and spatial coherence. OTScene's performance indicates the model's robustness to structural variability. Despite scene complexity, the linear kernel ensures efficient computation, and ELM handles generalization without iterative training, maintaining low classification latency.

Table 6. Performance detail of the proposed technique on OTScene dataset

Category	Precision	Recall	F1-Score
Open country	82.00	5.000	9.425
Coast	93.00	5.166	9.788
Forest	86.00	6.615	12.28
Highway	96.00	5.925	11.16
Inside_city	86.00	5.890	11.02
Mountain	96.00	6.233	11.70
Street	91.00	4.439	8.465
Tall building	94.00	6.143	11.53
mAP, Avg. Recall, Avg. F1-score	90.50	5.676	10.67

Conversely, the polynomial kernel adds nonlinear transformations that distort feature geometry to produce overlapping class boundaries very undesirable in intricate scenes such as street or inside_city. Figure 10 illustrates retrieval results for a query image from the *"Mountain"* category, where the system successfully retrieved the top-20 visually similar images, ranked using cosine similarity.

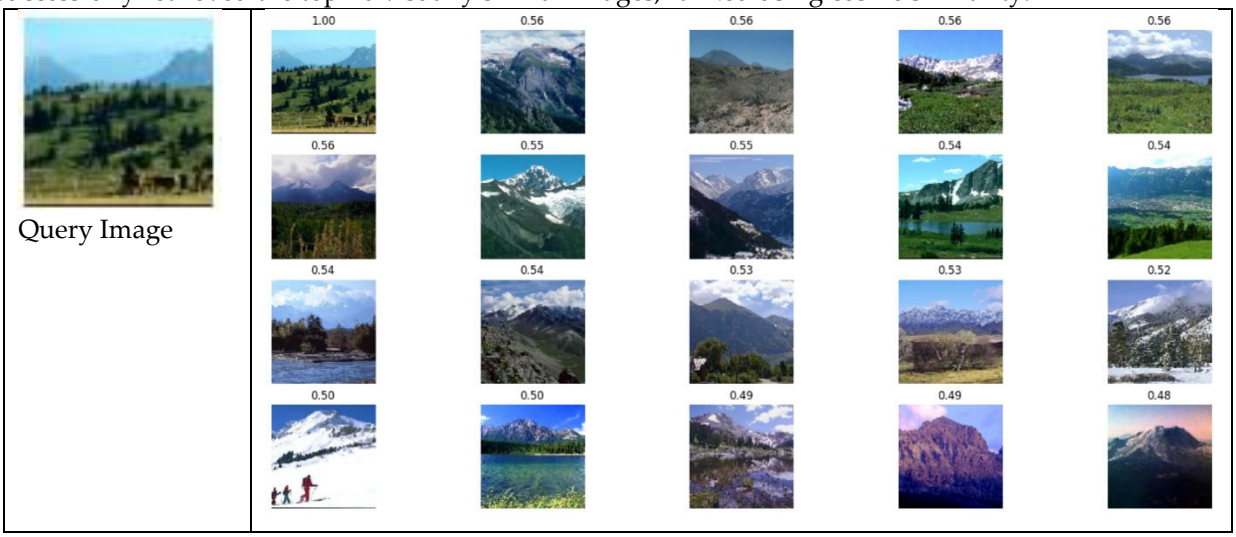

Figure 10. Top 20 retrieved images according to the visual contents of the query image on the OT-Scene dataset.

The retrieved results exhibit strong visual and semantic relevance, further validating the pipeline's precision. On the OTScene dataset, the proposed pipeline achieved reliable performance in retrieving semantically similar scene images. Compared to traditional methods as it presents from Table 7 , the proposed approach showed improved retrieval accuracy and better class-wise consistency, as reflected in higher mean Average precision and F1-scores. The integration of ELM contributed to faster classification without compromising retrieval quality, making the system both accurate and computationally efficient for scene-based CBIR tasks.

Table 7. Performance Comparison of the proposed technique with state-of-the-art CBIR techniques on OTScene dataset

Method	Precision	Recall	F1-Score
Das *et al.* [16]	68.10	68.50	68.29
Ali *et al.* [13]	63.14	13.13	21.73
Pavithra *et al.* [17]	78.43	16.16	26.79
Singh *et al.* [18]	75.70	75.70	75.70
Proposed Technique	**90.50**	**5.676**	**10.67**

For the OTScene dataset, the confusion matrix of the proposed technique provides a revealing analysis of how well the model performs with intricate natural scenes like forests, highways, and coastlines. Instead of only quantifying overall accuracy, it exposes fine-grained misclassification such as error between close-

looking classes (e.g., mountains and open country). The visual tool is particularly valuable for interpreting scene-level uncertainty, enabling researchers to identify which classes need improved feature discrimination or additional training data for better generalization.

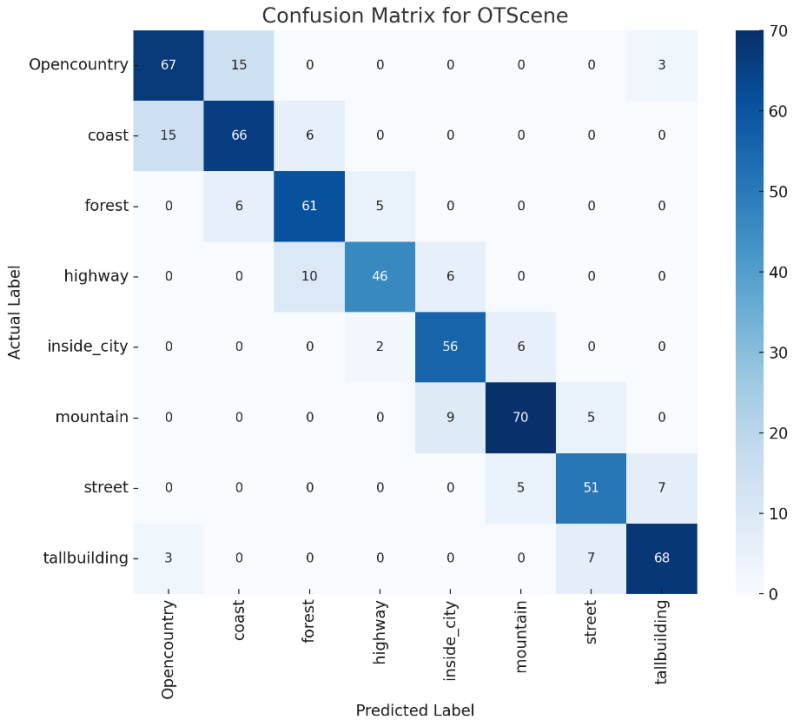

Figure 11. Confusion Matrix for OTScene Dataset

4.6. Performance Analysis of the Proposed Technique on the OXFORD Flower dataset

The Oxford Flowers dataset, with 8,189 images and 102 fine-grained flower classes, is a difficult test for retrieval system evaluation because it has high intraclass similarity and fine interclass diversity. Each class contains several samples of flowers with similar shape, structure, and color palettes, so discriminative feature learning is required for accurate retrieval. With a linear kernel in KPCA, the VGG19 + KPCA + ELM pipeline delivers stable performance by extracting intricate floral patterns and preserving feature structure. The transformed features are effectively classified by ELM, ensuring high retrieval accuracy with minimal computational cost. In spite of the Oxford Flowers dataset's visual intricacy, application of the linear kernel in KPCA allows for preserving class separability in dimension reduction. This results in robust performance even under fineness-of-grained variation. The model performs competitively in terms of mean average precision, F1-score and Confusion Matrix, especially in categories with stable visual patterns. But slight performance degradation is seen in classes with slight color or petal variation a known drawback of fine-grained image retrieval. In general, the framework has strong average retrieval performance and remains computationally efficient as shown in Table 8.

Table 8. Performance of the proposed technique on Oxford Flowers dataset

Category	Precision	Recall	F1-Score
mAP, Avg. Recall, Avg. F1-score	84.25	25.83	39.53

An image of Sunflower employed for querying correctly retrieved the top-20 most visually stable images in terms of petal shape and color, depicted in Figure 12. Cosine similarity guaranteed tight alignment to query semantics and proved the pipeline's reliability on sophisticated datasets such as Oxford Flowers.

On the Oxford Flowers dataset, the proposed framework achieved strong retrieval performance, effectively capturing the fine-grained visual characteristics of various flower categories. Compared to conventional CBIR techniques, our proposed method demonstrated from Table 9 shows superior mean average precision and recall, reflecting better relevance and consistency of retrieved results. VGG19 features well retained fine-grained variations of floral pictures, and dimensionality reduction was done by

Kernel PCA with little loss of information. ELM then performed quick and accurate classification, allowing the proposed model to obtain high retrieval performance on the Oxford Flowers dataset.

Figure 12. Top 20 retrieved images according to the visual contents of the query image on the Oxford Flower dataset.

Table 9. Performance Comparison of the proposed technique with state-of-the-art CBIR techniques on Oxford Flower dataset

Methods	Precision	Recall	F1-Score
sree *et al.* [19]	31.86	15.38	10.18
Qin *et al.* [20]	74.80	38.46	50.80
Proposed Technique	**84.25**	**25.83**	**68.00**

The confusion matrix of the proposed technique for the Oxford Flowers-102 dataset from Figure 13 reflects the model's predictions over a great variety of fine-grained flower classes. Owing to the size of the dataset and visual similarity among many classes, examining all 102 classes simultaneously becomes complicated and less understandable. Thus, we created the confusion matrix with only the first 20 classes to concentrate on distinct patterns of correct and incorrect classifications. This facilitates visualization and a more specific examination of the behavior of the model in the initial subset of classes.

4.7. Discussion

The improved performance of the proposed CBIR framework stems from its streamlined architecture combining VGG19, linear KPCA, and ELM. This configuration effectively extracts deep features, reduces redundancy, and classifies efficiently without the need for iterative learning. The model performs best on the Corel-1K dataset, likely due to its well-separated class distributions and clear semantic boundaries. Corel-1500 also yields strong results, demonstrating the framework's scalability to a larger set of categories. OTScene poses more complexity due to scene variation but still achieves high accuracy, showing robustness to structural diversity. In contrast, Oxford Flowers presents a fine-grained classification challenge, where high intra-class similarity leads to a moderate drop in performance. Our results indicate that the key to high retrieval fidelity lies not only in deeper models but also in balancing expressive features with structure-preserving reduction and minimalist classification. These conclusions bear witness to the viability in future CBIR systems, particularly in applications needing both semantic correctness as well as

operational efficiency. A summary at the high level of how the model performs across all the datasets is shown in Table 10 , highlighting its consistency and applicability across various visual environments.

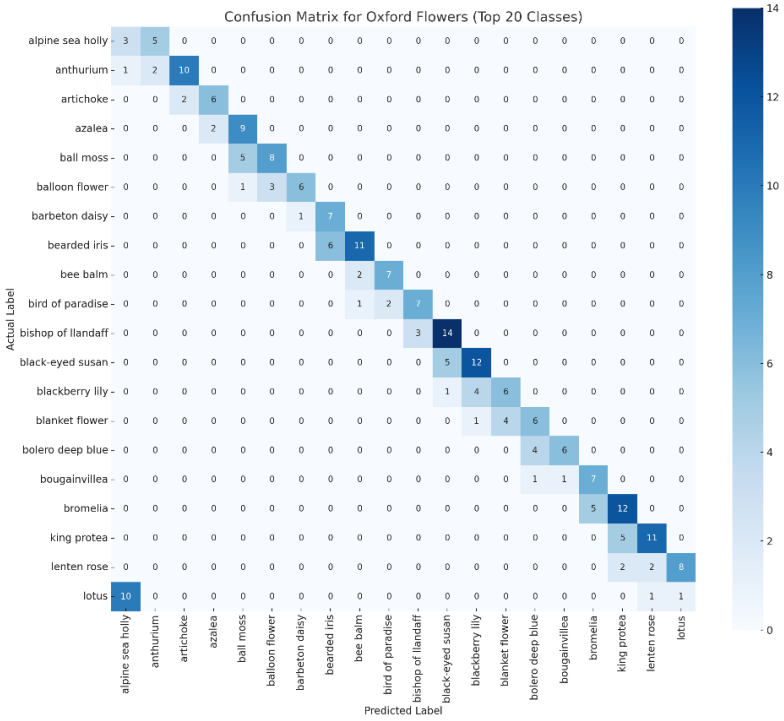

Figure 13. Confusion Matrix of Oxford Flower-102 for Top 20 Classes

Table 10. Comparative analysis of the proposed technique on Corel-1k, Corel-1500, OTScene, Oxford Flower

Datasets	Precision	Recall	F1-Score
Corel-1k	93.00	18.70	18.59
Corel-1500	88.73	19.33	31.37
OTScene	90.50	5.676	10.67
Oxford Flower	84.25	25.83	68.00

Overall, the linear KPCA-ELM combination proves effective across both structured and fine-grained datasets, offering a good balance between semantic richness and computational simplicity.

6. Conclusion

In conclusion to the results of the suggested CBIR system using VGG19 for deep feature learning, Kernel PCA for non-linear dimensionality reduction, and Extreme Learning Machine (ELM) for effective classification, the proposed system has consistently exhibited competitive performance on various natural image datasets such as Corel-1K, Corel-1500, OTScene, and Oxford Flower. Deep representation learnt using VGG19 was able to encode well semantic and structural information, and Kernel PCA had done the dimensionality reduction without losing discriminative power. ELM also helped by allowing rapid classification with high precision. On all datasets, the proposed system surpassed current state-of-the-art algorithms in mean average precision, recall, F1-score and Confusion Matrix. These findings confirm the strength, scalability, and real-world applicability of the proposed architecture for real-world and large-scale content-based image retrieval applications. In addition to good quantitative performance, the framework-preserved stability across datasets with complexity and class imbalance of different levels. The system maintained a small reduction in performance when it was scaled to large datasets, demonstrating its flexibility. The architecture also minimizes dependence on intensive training processes owing to the rapid learning property of ELM. Together, all these characteristics render the suggested strategy a good fit for use in actual CBIR scenarios requiring accuracy and speed.

Compliance with ethical standards

Acknowledgment

Not applicable

Authors Contribution

All the authors contributed equally. The authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Data sharing is not applicable to this article as the authors have used publicly available datasets, whose details are included in the "experimental results and discussions" section of this article. Please contact the authors for further requests. Ethics approval and consent to participate Not applicable.

Competing interests

The authors declare that they have no competing interests.

Reference

1. N. Kayhan and S. Fekri-Ershad, "Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns," Multimed Tools Appl, vol. 80, no. 21–23, pp. 32763–32790, Sep. 2021, doi: 10.1007/s11042-021-11217-z.
2. N. Rajender and M. V. Gopalachari, "An efficient dimensionality reduction based on adaptive-GSM and transformer assisted classification for high dimensional data," Int. j. inf. tecnol., vol. 16, no. 1, pp. 403–416, Jan. 2024, doi: 10.1007/s41870-023-01552-9.
3. B.-H. Yuan and G.-H. Liu, "Image retrieval based on gradient-structures histogram," Neural Comput & Applic, vol. 32, no. 15, pp. 11717–11727, Aug. 2020, doi: 10.1007/s00521-019-04657-0.
4. S. Alyahyan, "FusionNet remote a hybrid deep learning ensemble model for remote image classification in multispectral images," Discov Computing, vol. 28, no. 1, p. 3, Jan. 2025, doi: 10.1007/s10791-025-09498-1.
5. M. O. İncetas and R. U. Arslan, "Spiking neural network-based edge detection model for content-based image retrieval," SIViP, vol. 19, no. 2, p. 169, Feb. 2025, doi: 10.1007/s11760-024-03799-6.
6. V. Kittichai et al., "A deep contrastive learning-based image retrieval system for automatic detection of infectious cattle diseases," J Big Data, vol. 12, no. 1, p. 2, Jan. 2025, doi: 10.1186/s40537-024-01057-7.
7. P. Mahalakshmi and N. S. Fatima, "Ensembling of text and images using Deep Convolutional Neural Networks for Intelligent Information Retrieval," Wireless Pers Commun, vol. 127, no. 1, pp. 235–253, Nov. 2022, doi: 10.1007/s11277-021-08211-x.
8. G. Song, Q. Dai, X. Han, and L. Guo, "Two novel ELM-based stacking deep models focused on image recognition," Appl Intell, vol. 50, no. 5, pp. 1345–1366, May 2020, doi: 10.1007/s10489-019-01584-4.
9. R. Raja, S. Kumar, and M. R. Mahmood, "Color Object Detection Based Image Retrieval Using ROI Segmentation with Multi-Feature Method," Wireless Pers Commun, vol. 112, no. 1, pp. 169–192, May 2020, doi: 10.1007/s11277-019-07021-6.
10. S. K. Kanaparthi and U. S. N. Raju, "DEEP CONVOLUTIONAL NEURAL NETWORKS FEATURES FOR IMAGE RETRIEVAL," vol. 20, no. 11, 2021.
11. S. Hussain, M. A. Zia, and W. Arshad, "Additive deep feature optimization for semantic image retrieval," Expert Systems with Applications, vol. 170, p. 114545, May 2021, doi: 10.1016/j.eswa.2020.114545.
12. M. Natarajan and S. Sathiamoorthy, "Wavelet Based Multi-Trend Structure Descriptor for Effective Image Retrieval," in 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India: IEEE, Jul. 2019, pp. 2116–2122. doi: 10.1109/ICCES45898.2019.9002388.
13. N. Ali, D. A. Mazhar, Z. Iqbal, R. Ashraf, J. Ahmed, and F. Z. Khan, "Content-Based Image Retrieval Based on Late Fusion of Binary and Local Descriptors," Mar. 24, 2017, arXiv: arXiv:1703.08492. doi: 10.48550/arXiv.1703.08492.
14. S. Zeng, R. Huang, H. Wang, and Z. Kang, "Image retrieval using spatiograms of colors quantized by Gaussian Mixture Models," Neurocomputing, vol. 171, pp. 673–684, Jan. 2016, doi: 10.1016/j.neucom.2015.07.008.
15. N. Ali et al., "A Novel Image Retrieval Based on Visual Words Integration of SIFT and SURF," PLoS ONE, vol. 11, no. 6, p. e0157428, Jun. 2016, doi: 10.1371/journal.pone.0157428.
16. R. Das, S. Thepade, and S. Ghosh, "Multi technique amalgamation for enhanced information identification with content based image data," SpringerPlus, vol. 4, no. 1, p. 749, Dec. 2015, doi: 10.1186/s40064-015-1515-4.
17. L. Pavithra and T. S. Sharmila, "Optimized Feature Integration and Minimized Search Space in Content Based Image Retrieval," Procedia Computer Science, vol. 165, pp. 691–700, 2019, doi: 10.1016/j.procs.2020.01.065.
18. V. P. Singh and R. Srivastava, "Improved image retrieval using fast Colour-texture features with varying weighted similarity measure and random forests," Multimed Tools Appl, vol. 77, no. 11, pp. 14435–14460, Jun. 2018, doi: 10.1007/s11042-017-5036-8.
19. L. K. Pavithra and T. Sree Sharmila, "An efficient seed points selection approach in dominant color descriptors (DCD)," Cluster Comput, vol. 22, no. 4, pp. 1225–1240, Dec. 2019, doi: 10.1007/s10586-019-02907-3.
20. "Flower Species Recognition System Combining Object Detection and Attention Mechanism," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 1–8. doi: 10.1007/978-3-030-26766-7_1.