**C**urrent **A**pplied **S**cience and **T**echnology e-ISSN: 2586-9396

Curr. Appl. Sci. Technol. ….., Vol... (No...), e0263395

## *Research article*

# OptISNV2: An Effective Deep Learning-Based Approach to Content-Based Image Retrieval Systems

**Rajasekharam Gulla[1]*, S. Saradha Rani[2] and Rajyalakshmi Uppada[3]**

*[1]Department of Computer Science and Engineering, Miracle Educational Society Group of Institutions, Andhra Pradesh, 535216, India*
*[2]Department of Electrical Electronics and Communication Engineering, GITAM (Deemed to be University) Visakhapatnam, Andhra Pradesh, 530045, India*
*[3]Department of Electronics and Communication Engineering, School of Engineering, Aditya University, Surampalem, Andhra Pradesh, 533437, India*

Received: 18 May 2024, Revised: 31 March 2025, Accepted: 18 April 2025, Published: 2 September 2025

## Abstract

There has been a significant increase in the production of images on the internet in recent years, necessitating the development of automated content management systems. The content-based image retrieval (CBIR) model has been developed to lessen reliance on the textual annotations-based picture retrieval model. A range of features-classifier combinations-based CBIR techniques are available for analyzing query image content and retrieving appropriate images. While these techniques improve retrieval performance in single-class scenarios, semantic similarity between images of various classes causes a considerable loss in performance in multi-class search contexts. This research proposes a novel deep learning-based technique for a content-based image retrieval system. Initially, the input/query images were taken from three publicly available datasets: Kvasir, CIFAR-10, and Corel-1k. Then, noise reduction contrast enhancement and normalization in pre-processing were performed to better understand the image features. After that, the picture's image texture, color, and shape attributes were retrieved for classification and similarity matching using the SE-ResNeXt-101 technique. Then, from the query picture and database pictures, these attribute vectors were utilized to categorize the dataset pictures using the Improved ShuffleNetV2 method; this method measures the similarities between pictures to retrieve the most alike DB pictures given as the query. The Improved Mayfly Optimization Algorithm (IMOA) was used to increase retrieval performance. The outcomes of the experiments show that the approach performed better in terms of picture retrieval than the most advanced CBIR techniques currently in use.

**Keywords:** content-based image retrieval; shufflenetv2; classification; improved mayfly optimization algorithm

*Corresponding author: E-mail: grajasekharam@gmail.com
https://doi.org/10.55003/cast.2025.263395

## 1. Introduction

There are several image databases accessible all across the world. When accessing these databases, it is crucial to have a reliable and efficient retrieval and search strategy. The conventional method of retrieving images necessitates providing a text description and keyword search for each image (Kayhan et al., 2021; Choe et al., 2022; Schall et al., 2022). Since the quantity and variety of image contents have greatly expanded, these are incredibly challenging and unclear stages. As a result, there is a lot of interest in content-based image retrieval (CBIR). In the past, many researchers worked on extracting pictures from massive databases by examining the contents of the pictures (Hu et al., 2021; Majhi et al., 2021). Indeed, since the 1990s, the multimedia community has actively pursued research in CBIR. The user must enter the query image's description to use this method. Features are extracted from Query Image (QI) by the CBIR systems. Next, a search is conducted using the derived features to locate relevant images. The feature vector computation for the QI makes use of these features. In actuality, CBIR uses a vector to represent every image in the database (Trappey et al., 2021; Zhao et al., 2021; Sivakumar et al., 2022). Following the QI entry, the CBIR computes its feature vector. Next, a comparison is performed between the computed feature vector and the picture vectors kept in the database. The CBIR system retrieves images that have attributes in common with the QI.

A query image is an input image provided to an image retrieval system to find similar or relevant images from a database. It serves as the reference point for comparison, where its features, such as color, texture, shape, or other visual attributes, are analyzed and matched against the features of images stored in the system's database. For instance, in a content-based image retrieval (CBIR) system, the query image is processed to extract its feature vectors, which are then used to calculate similarity scores and retrieve the most visually or semantically similar images. This process is widely used in applications like medical imaging, e-commerce, and multimedia search engines. Various feature extraction techniques have been employed to retrieve image-specific essential features effectively for use in CBIR systems (Kumar et al., 2021; Mezzoudj et al., 2021). These could be edge-based characteristics, forms, or other color spaces (Lab, HSV). Three values are defined by the Lab color space: b (from blue to yellow), a (from green to red), and L (lightness from black to white). RGB colors are remapped to human-readable proportions using the HSV color scheme. The most challenging problem with such a system is the semantic discrepancy between the visual information acquired by the imaging equipment and the visual information received by the Human Vision Systems (HVS) (Hung & Phuong, 2022; Tankala et al., 2022). Determining the boundaries of images accurately takes a lot of effort and is challenging. In CBIR systems, texture-based attributes and transforms (e.g., 2D wavelet, Gabor) have assisted image retrieval across domains (Ghaleb et al., 2022; Subramanian et al., 2022; Higaki et al., 2023). Such a CBIR system's effectiveness is heavily dependent on selected features, which can lead to overfitting in cases where the number of input pictures is insufficient. Furthermore, if the input image dimensions are large, the selection process could need significant processing time.

Khan et al. (2021) suggested a CBIR technique based on a hybrid attributes descriptor using the Support Vector Machine (SVM) and Genetic Algorithm (GA) classifiers for picture retrieval in a multi-class scenario and for feature extraction, deliberately using the first three color moments— Daubechies, Haar, and Bi-Orthogonal wavelets—refining the attributes with GA before training the multi-class SVM with the one-against-all method. When comparing the obtained images from the picture repository to the query picture, the

L2 Norm was employed as a similarity measuring function. Desai et al. (2021) presented a successful deep-learning system based on SVM and Convolutional Neural Network (CNN) for quick image retrieval. The suggested architecture used CNN to extract attributes and SVM for categorization. The outcomes showed how reliable the system was. An average efficiency was obtained for the suggested CBIR system employing SVM for categorization and CNN for feature extraction.

The DTLDN-CBIRA model, capable of deep transfer learning, was presented in Karthikeyan et al. (2023) as a method for CBIR in agricultural plant disease pictures. Due to the small number of samples in the dataset, rotation and flipping were the two methods used for data augmentation. Furthermore, the DTLDN-CBIRA model extracted features using the densely connected networks (DenseNet-201) model. Simultaneously, retrieval performance was significantly impacted by the hyper-parameters of DL models; the DenseNet-201 model's hyper-parameters were tuned using the Stochastic Gradient Descent (SGD) optimizer. Ultimately, the similarity between photos was measured using the Manhattan distance metric, and the images that exhibited the highest degree of similarity were extracted from the database.

Alsmadi et al. (2020) implemented the picture retrieval procedure using attributes related to shape, text, and color. In this approach, they extracted color features using the discrete wavelet transform and the canny edge histogram. Gray Level Co-occurrence Matrix (GLCM) was then used to carry out the text feature extraction. The neutrosophic clustering algorithm and the canny edge approach were also utilized to retrieve form attributes. A simulated annealing technique and genetic algorithm were finally applied to obtain related images based on similarity. The performance evaluation was conducted using the Corel-1k dataset. Garg and Dhiman (2021) used a four phase approach to construct the CBIR framework. First, the R, G, and B channels were decomposed using a Discrete Wavelet Transformation (DWT) based method. These three R, G, and B channels were concatenated using a set of routines. Subsequently, the Particle Swarm Optimization (PSO) algorithm performed the feature selection process, choosing significant characteristics from the image. Ultimately, three classifiers carried out the classification process by determining the images from the Corel-1k dataset that were related based on similarity (Haddadi et al., 2024).

Additionally, previous methods tested their experimental analysis on tiny datasets (Chen et al., 2022). In theory, CBIR technology aids in organizing and retrieving digital image archives based on their visual content. Grouping comparable images that are one another based on their attributes is the aim of the image retrieval system. While adding features increases the information in the picture, it also increases memory and processing time (Lei et al., 2022).

The critical contributions of this research are as follows:

- The input/query images are taken from three different and publicly available datasets: Kvasir, CIFAR-10, and Corel-1k. The robustness and generalizability of the model are ensured by evaluating its performance across diverse datasets containing varied image types, categories, and complexities.

- Pre-processing steps including noise reduction, contrast enhancement, and normalization are performed to understand the image features better. The quality of the input images is enhanced by reducing noise and improving contrast, facilitating better extraction and understanding of key image features.

- A picture's image shape, color, and texture attributes are retrieved for similarity matching using the SE-ResNeXt-101 technique. The advanced capabilities of SE-ResNeXt-101 are leveraged to accurately extract critical image features, ensuring high-quality feature representation for similarity matching.

- These feature vectors are utilized to categorize the dataset images using the Improved ShuffleNetV2 method from the query and database pictures. High computational efficiency is combined with strong classification performance, enabling effective handling of large-scale datasets without compromising accuracy.

- The Improved Mayfly Optimization Algorithm (IMOA) is used to increase retrieval performance. The similarity matching process is optimized by refining search strategies and balancing exploration and exploitation, resulting in improved retrieval accuracy and reduced computational overhead.

## 2. Materials and Methods

Images were captured as input images for the proposed image retrieval system. The proposed methodology, which combines content-based picture retrieval with picture categorization, is shown in Figure 1.

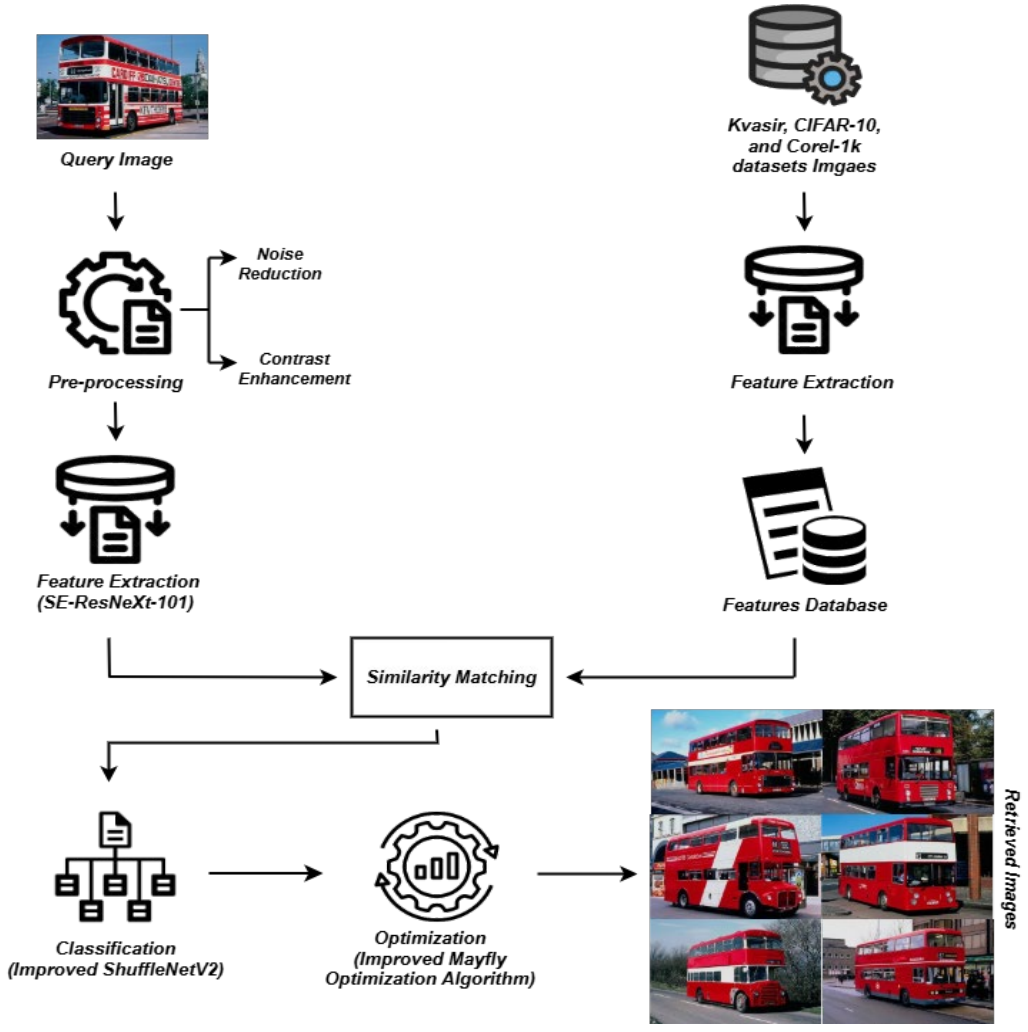



**Figure 1.** Architecture of proposed methodology

***i) Image acquisition:*** The input pictures are taken from the three publicly available datasets, such as Kvasir, CIFAR-10, and Corel-1k.

***ii). Pre-processing:*** The query image undergoes the noise reduction using a mean filter and the contrast of the query image is enhanced using CLAHE.

***iii). Feature extraction:*** All images have visual patterns, surface properties, and scenes. From these images, an image's color, texture, and shape features are extracted for similarity matching.

***iv) Image matching and similarity measure:*** This technique measures the similarities between images to retrieve the most similar DB images given as the query. The query image is retrieved from the classification output. Thus, the output retrieval rate is high compared to other models.

***v). Classification:*** These feature vectors are used to classify the database using Improved ShuffleNetV2 from the query image and database images.

***vi). Optimization:*** And improve the classification performance utilizing IMOA.

The CBIR proposed model's evaluation performance analysis provides better outcomes than the previous model.

## 2.1 Pre-processing

The pictures' contrast and noise levels were improved during the pre-processing stage. Here, the noise from the input photos is reduced using mean filtering. A conventional approach for picture de-noising in the spatial domain involves utilizing several picture smoothing patterns for image convolution processing to minimize or remove noise. The basic idea behind mean filtering is to substitute the total grey value of several nearby pixels for a pixel's solitary grey value. After mean filtering and smoothing, the picture is g(x, y). The method below can be used to find g(x, y) for a pixel point (x, y) in a given picture with f (x, y), where its neighborhood S consists of M pixels as calculated with equation (1):

$$G(x,y) = \frac{1}{M}\sum_{(i,j\in S)} f(x,y)(x,y) \notin S \tag{1}$$

To increase contrast, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) approach was utilized. CLAHE successfully enhanced low-contrast pictures and was cleverly designed to organize pictures. Beyond traditional approaches of histogram equalization, CLAHE offers two significant advantages. The picture's value distribution becomes more uniform, and the histogram's range expands as the histogram is equalized utilizing standard histogram equalization approaches.

By enhancing contrast in every zone, CLAHE enhances contrast throughout a picture using equation (2).

$$D_B = \frac{255}{8\times 8}\sum_{i=0}^{D_A} H(i) \tag{2}$$

Then, CLAHE could mitigate the issue of noise amplification by limiting enhanced contrast. This is how the final, updated histogram appears as equation (3):

$$H(i) == \begin{cases} H(i) + L, & \mathrm{H}(i) < H_{max} \\ \mathrm{H}(i)_{max}, & \mathrm{H}(i) \geq H_{max} \end{cases} \tag{3}$$

The images should explicitly reference and explain equation (3) to provide clarity on its role in mitigating noise amplification during the CLAHE process. Specifically, equation (3) defines how the histogram was adjusted, with H($i$) representing the pixel intensity distribution, $L$ indicating the redistribution factor, and $H_{max}$ setting the upper limit to restrict excessive enhancement. The text was enhanced to ensure that the equation was accurate, clearly defined, and properly integrated into the discussion, emphasizing its significance in improving contrast while controlling noise. Figure 2 shows the histogram equalization of input and enhanced images.

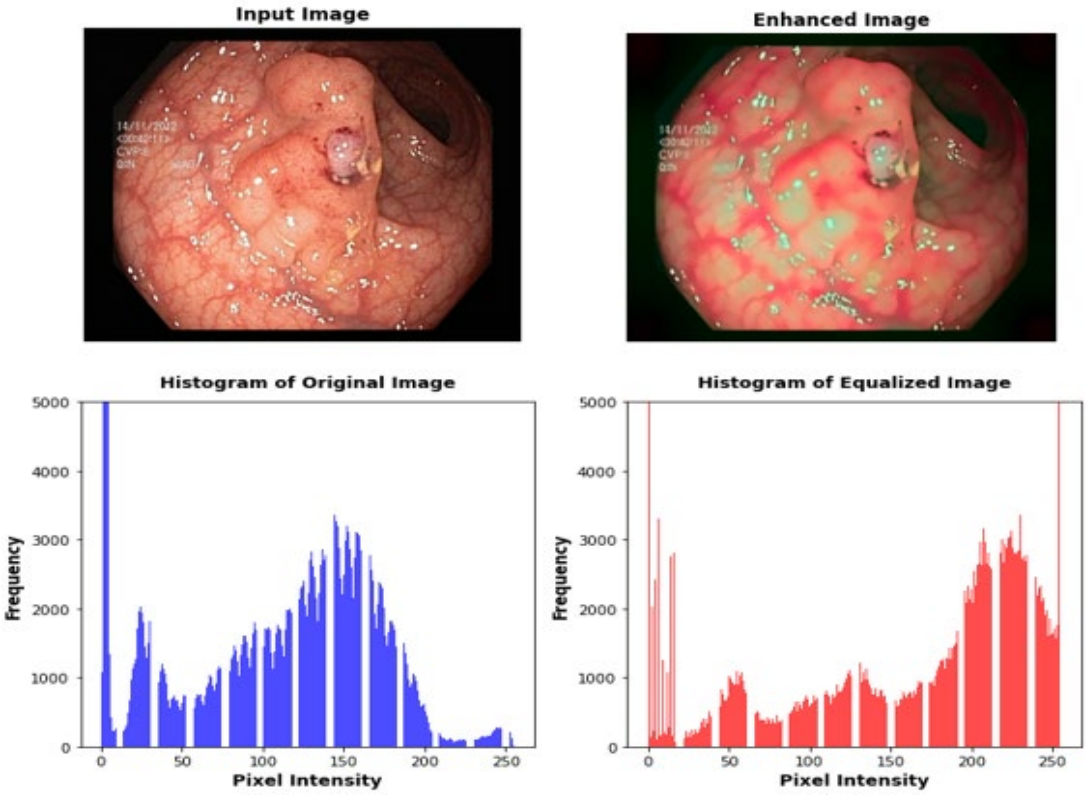



**Figure 2.** Histogram equalization of image

## 2.2 Feature extraction

A crucial phase in image retrieval is feature extraction, which involves describing the image using the fewest possible descriptors. Color is a highly potent descriptor that makes object recognition easier in real-time CBIR, and it is also one of the most commonly employed visual characteristics. Using similarity matching, one can extract a picture's color attributes. Gabor wavelets are used to acquire the texture attribute. Tree- and pyramid-structured wavelet transform features and texture features are preferred for retrieval due to their superior performance in multiresolution simultaneously autoregressive models. To extract attributes from the input data, a pre-trained SE-ResNeXt-101 algorithm was used. From the transformation, which maps the input Z2 SB0×Y0×R0 to feature mapping V2 SB×Y×R,

a computational unit called a squeeze-and-excitation block was created. The learned sets of filter kernels were represented by U = (u1, u2,..., uR) in the characters that followed, where uR was the parameter of the Rth filter. Eks was a convolutional operator. Equation (4) was therefore used to describe the output as V = (v1, v2,..., vR):

$$v_r = u_r * Z \sum_{m=1}^{R'} u_r^m * Z^m \quad (4)$$

In this case, ur = hu1r, u2r,..., urR0i, Z = hz1, z2,..., zR0i, and vr 2 SB×Y are represented as *. The 2D spatial kernel, um r, acts on the corresponding Z channels and represents a single ur channel.

To simplify the terminology, bias terms were removed. One potential solution to the problem of channel dependency exploitation is to compress global geographical data into channel descriptors. Global average pooling was used to generate channel-specific data to achieve this. In formal terms, the statistic a 2 SR was generated by reducing V over its spatial dimensions B × Y. This allowed for the determination of the $r^{th}$ position of a using equation (5):

$$a_r = E_{md}(v_r) = \frac{1}{B \times Y} \sum_{j=1}^{B} \sum_{i=1}^{Y} v_r(j, i) \quad (5)$$

A second operation was carried out to fully capture channel-wise dependencies using the data collected during the squeeze operations. An adaptation of ResNet that exhibited a striking resemblance to the Inception model was the ResNeXt component. Both followed the split transform merge approach, except, in this version, the output of distinct routes was combined rather than depth combined as in the inception model.

$$e(Z) = \sum_{j=1}^{R} K_j(z) \quad (6)$$

$K_j(z)$ was any function in equation (6). Like a primary neuron, $K_j$ projects z into a (perhaps low dimensional) space, embedding and then changing it. R indicates how big the set of changes to be combined is in equation (6). R is referred to as cardinal. The dimension of cardinality determines how many more complex transformations there can be. Equation (7)'s aggregated transformation is the residual function.

$$f = Z + \sum_{j=1}^{R} K_j(Z) \quad (7)$$

Where f was the output.

We concatenated the texture and color attributes to create a hybrid feature vector (HFV) that effectively represented the input images. It is crucial to note that texture and color attributes alone cannot effectively categorize images with identical texture and color, which raises the system's false positive rate. We proposed developing a hybrid feature vector incorporating texture and color characteristics to mitigate this vulnerability. When using the picture dataset, the hybrid feature vector's overall dimension for a single semantic class was (100 × 654), whereas the single attribute vector's dimension was (1 × 654). The total dimension of all the color attributes made up the dimension of the color attribute vector. The color feature vector's overall dimension for a single picture was (1 × 264).

### 2.2.1 Image matching and similarity measure

Following the extraction of shape, color, and texture, the distance between the query picture using SE-ResNeXt-101 technique and a database image—abbreviated D(q,d)—was used to determine how similar the two images are, and it was determined by taking into account the attributes of the extracted texture, color, and form. As the two images get farther apart, they are less similar. The distance between the two pictures is zero when they are equivalent. The measurement of the separation between two locations in a multidimensional space is called the Euclidean distance. We calculated the L2 norm in the following manner to gauge the degree of similarity between the feature vector of the input query picture and the dataset pictures in equation (8).

$$d(p,q) = \sqrt{(p1-q1)^2 + (p2-q2)^2 + \cdots + (pn-qn)^2} = \sqrt{\sum_{i=1}^{n}(pi-qi)^2} \quad (8)$$

## 2.3 Classification

After extracting the feature vector from the image, the Improved ShuffleNetV2 technique was utilized to classify the images based on their category after matching them with the features database. Position-based restrictions did not apply to the feature vector placement in this method. The proposed system's accuracy in retrieval was improved by combining various optimal attributes. A lightweight network model that seeks to maximize operational speed and minimize model size while maintaining performance efficiency was the target. The primary innovation of this framework was its complete utilization of the two operations—group convolution and channel shuffle—to lower the model's parameter count and computation burden. In particular, group convolution-related poor information flow was addressed by performing an operation called "channel shuffle," which involved rearranging the feature maps' channels and creating a new one. A significant Memory Access (MAC) cost may result from excessive group convolution. The model's capacity for feature extraction improved gradually as network depth rose and detection accuracy improved. Figure 3 depicts the ShuffleNetV2 network's structure.

The ShuffleNetV2 network comprised two kinds of blocks as its fundamental building pieces. Block 1 randomly separated the input channel into two segments: the first segment conducted separable convolution to retrieve visual attributes, while the second segment kept its mapping and transmitted immediately downward. Each time Block 2 flowed through ShuffleNetV2 (23), the number of feature channels doubled, and all attribute channels had the same weight. The feature channels, which significantly influenced the categorization outcomes, were given a lot of attention now that the number of channels had doubled. However, the categorization effect quickly affected the depth-separable convolution employed in Block 2 since it was sensitive to the placement of sensitive attributes and retained excessive background data.

### 2.3.1 Model improvement

The illness dataset of edible fungal fruit bodies created for this study had a few issues, including a complicated backdrop of disease images, a high degree of background resemblance between various diseases, a wide range of disease sizes, etc. Consequently, the existing model needed to perform better in recognition, and the overall recognition was challenging. This study improved the ShuffleNetV2 model by considering its recognition accuracy and speed.

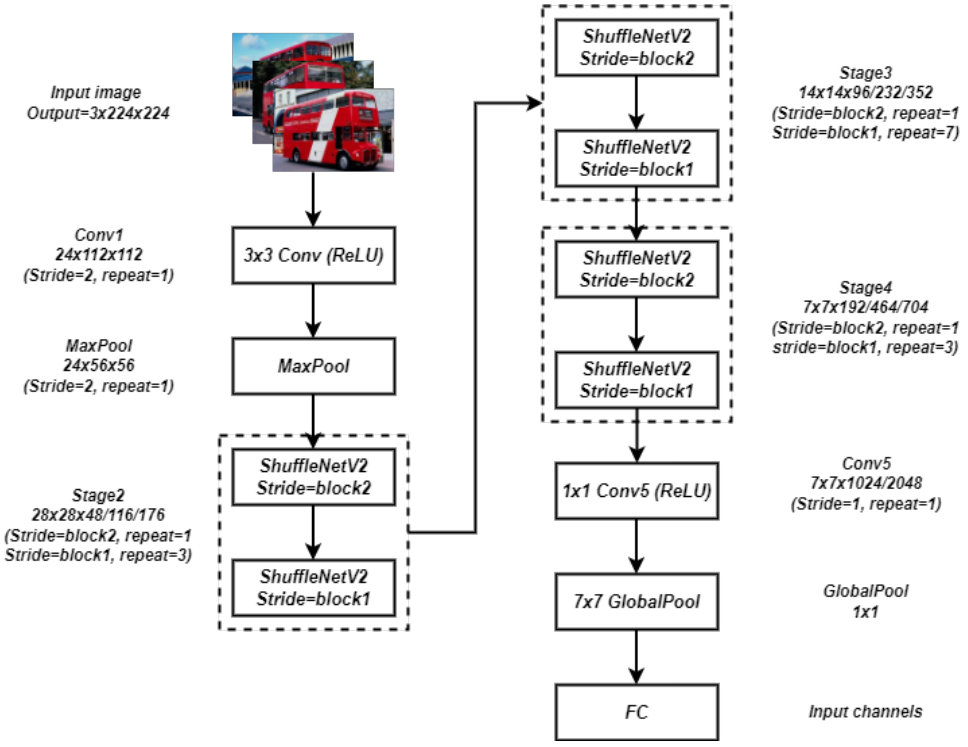



**Figure 3.** The framework of the ShuffleNetV2 network

**1) Attention mechanism**

The attention mechanism seeks to minimize the effect of areas of interest in picture segmentation by concentrating on those areas. Spatial attention and channel attention are the two categories under which deep learning CNN attention methods fall. The term "channel attention" describes identifying the relative weights of several channels, emphasizing the importance of essential channels, and suppressing less effective channels.

The channel attention mode uses the SE attention mechanism to determine weights and prioritize information by allocating weights among several channels. The SE attention module may automatically increase the feature channels in the image with rich contrast data, suppress the attribute channels irrelevant to the goal, and modify the weight based on various feature channels. One technique for channel attention is the ECA attention mechanism. The CBAM attention mechanism is an algorithm model combining spatial and channel attention processes.

**2) Simplify model framework**

Among ShuffleNetV2's fundamental elements, three convolution layer operations were carried out on the correct branch: the standard 1 × 1 convolution layer, the 3 × 3 deep convolution layer, and Batch Normalization, BN, and ReLU. It uses two 1 × 1 convolution layer operations; however, there are a few unnecessary layers. Here, only the fusing of the

DW convolution's inter-channel data using a 1 x 1 convolution layer operation was required, as the dimensionality-up and dimensionality-down operations were not required. According to this research, the ShuffleNetV2-Lite network topology that had been developed was able to efficiently lower computational complexity, improve model performance, and maintain model correctness. To improve the classification accuracy and performance, the Improved Mayfly Optimization Algorithm (IMOA) was utilized.

## 2.4 Optimization

The Improved Mayfly Optimization Algorithm (IMOA) was used to increase the accuracy of anomaly categorization performance. The Mayfly algorithm is a novel class of intelligent optimization algorithms with high research value and strong optimization ability. In addition, it draws inspiration from the mayfly's social behavior, particularly from their mating ritual. After emerging from an egg, a mayfly is considered an adult, and the most appropriate mayfly will live. Furthermore, every mayfly's location within the search space indicates a possible fix. The algorithm's basic idea is as follows: first, two groups of mayflies, symbolizing the male and female populations, were created randomly. Every mayfly is randomly positioned as a potential solution x = (x1,....., xd ) depicted as a d-dimensional vector in the problem space, and its effectiveness is assessed by a preset objective function f (x). Moreover, the mayfly's velocity, given as v = (v1,....., vd), represents the change in its position. Ultimately, every mayfly continuously modifies its course based on its current personal best location (best) and the group's best location (best) achieved by all mayflies combined.

### 2.4.1 Movement of male mayfly

Male mayflies congregate in groups and modify their locations according to their own and their neighbors' experiences. Let us assume that $x_i^t$ represents the mayfly current location in the search space at time step t as equation (9).

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{9}$$

It is likely that male mayflies cannot move quickly and will move continuously since they always dance a few meters above the water. The male mayfly's velocity is thus updated to equation (10):

$$v_{ij}^{t+1} = v_{ij}^t + a1e^{-\beta r_p^2}\left(pbest_{ij} - x_{ij}^t\right) + a2e^{-\beta r_g^2}\left(gbest_{ij} - x_{ij}^t\right) \tag{10}$$

The positive attraction constants are $t.a_1$ and $a_2$. The best position in Mayfly history is $pbest_i$, while $i.gbest$ represents the best location at a given time step. Its fixed visibility coefficient, β, controls the mayfly's visibility. The Cartesian distance between $pbest_i$ and the present position is denoted by $r_p$. The Cartesian distance from the present point to $gbest$ is denoted by $r_g$. Here's how the distance is computed as equation (11):

$$\|x_i - X_i\| = \sqrt{\sum_{j=1}^{n}\left(x_{ij} - X_{ij}\right)^2} \tag{11}$$

$X_i$ is equivalent to $pbest_i$ or $gbest$, and $x_{ij}$ is the $j^{th}$ element of mayfly i.

The group's most skilled mayflies must keep up their special up-and-down dance. Therefore, the best mayfly has to alter their speed all the time. Here is how the velocity is updated as equation (12):

$$v_{ij}^{t+1} = v_{ij}^{t} + d.r \tag{12}$$

Where r is a random number within the interval (−1, 1) and d is the nuptial dance coefficient.

### 2.4.2 Movement of female mayfly

Female mayflies do not have a collecting phenomenon like male mayflies; instead, they must fly to male groups to reproduce. It is assumed that a mayfly $i$'s location at time step t is represented by $y_i^t$, and that it updates its position by speeding up. Consequently, the computation looks as equation (13):

$$y_i^{t+1} = y_i^t + v_i^{t+1} \tag{13}$$

As a result, the velocity is determined as following equation (14):

$$v_{ij}^{t+1} = \begin{cases} v_{ij}^t + a2e^{-\beta r_{mf}^2\left(x_{ij}^t - y_{ij}^t\right)}, & if\ f(y_i) > f(x_i) \\ v_{ij}^t + fl.r, & if\ f(y_i) < f(x_i) \end{cases} \tag{14}$$

A random number between -1 and 1 is denoted by r. The Cartesian distance between male and female mayflies is denoted by $r_{mf}$. The calculation of $r_{mf}$ is done utilizing equation (14).

### 2.4.3 Mayfly mating

Male and female mayflies use the fitness function to choose pairings for mating. Additionally, the female and male mayflies with the highest fitness values mate, and so on. In addition, mating producing two offspring was calculated using the following equations (15)-(16):

$$offspring1 = L \cdot male + (1 - L) \cdot female \tag{15}$$
$$offspring2 = L \cdot female + (1 - L) \cdot male \tag{16}$$

where v0 is the beginning velocity of the progeny, and L is a random value within a given range.

### 2.4.4 Mayfly variation

To prevent premature convergence, a randomly generated integer with a normal distribution is added to the selected progeny mayfly during mutation, helping to avoid getting trapped in a local optimum instead of reaching the global optimum. The following is the mayfly progeny mutation equation (17):

$$offspring_n = offspring_n + \sigma.N(0,1) \tag{17}$$

In this case, σ represents the normal distribution's standard deviation. N(0, 1) is a conventional normal distribution with zero as the average value and one as the variance. The number of mutated individuals is 0.05 times that of male mayflies, resulting in a rounded total. With increasing iterations, the random flying coefficient and wedding dance coefficient will likewise decrease. In this context, equations (18) and (19) describe the exponential decay of key behavioural coefficients over successive iterations. Specifically, equation (18) governs the attenuation of the random flying coefficient:

$$d_t = d_0 . d_{damp}^1 \tag{18}$$

$$fl_t = fl_0 . fl_{damp}^1 \tag{19}$$

The attenuation variable of random flight and the wedding dance coefficients are denoted by $d_{damp}$ and $fl_{damp}$, respectively. Ultimately, the BiLSTM parameters are allocated the optimal values that were achieved. As a result, there is a 0.2 dropout rate, 100 hidden neurons, and a 0.0005 learning rate.

## 3. Results and Discussion

Two experiments were carried out in this work to assess and test the proposed similarity metrics. The color and texture features that were retrieved using conventional statistical techniques were applied in the first experiment. Like most retrieval models, this experiment consists of two stages: first, it computes the correlation of values between the query picture and every image in the database. Next, it ranks the images and determines how precise the retrieval method is. Five pictures were chosen at random from every class to symbolize the questions.

### 3.1 Hyper-parameter settings

This part presents the development findings of the system. Table 1 shows the settings for hyper-parameters.

**Table 1.** Settings for hyper-parameters

| Hyper-Parameter | Setting |
|---|---|
| Epochs | 850 |
| Batch size | 15 |
| Dropout rate | 0.5 to 0.1 |
| Activation output layer | Sigmoid |
| Activation Function | Relu |
| Loss | Binary_Crossentropy |

### 3.2 Experimental settings

The system was developed in a variety of situations. Table 2 demonstrates the environment setting for the system.

**Table 2.** Environment setup of the proposed model

| Resource | Details |
|---|---|
| RAM | 8 GB |
| CPU | Core i5 Gen6 |
| Software | Python |
| GPU | 4GB |

## 3.3 Dataset description

Three image datasets—Kvasir, CIFAR-10, and Corel-1k—were used in this investigation. In the first scenario, which involved the retrieval of the texture and color features, Kvasir's initial version—which contains 4,000 images—was employed. Eight classes were present in images, which displayed GI tract endoscopic procedures, pathological results, and anatomical landmarks. The CIFAR-10 dataset included 60,000 pictures divided into ten semantic classes with 6000 pictures per class. The proposed approach was tested using a general-purpose WANG dataset, including 1000 Corel-1k pictures with ten distinct subject groups. The WANG dataset is a widely used benchmark dataset in the field of image retrieval and classification. It consists of 1,000 images divided evenly into 10 semantic categories, with 100 images per category. These categories include diverse subjects such as African people and villages, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains, and food. This dataset was originally derived from the Corel image gallery and is often referred to as the Corel-1K or WANG dataset. These were offered in JPEG format of 384 × 256 and 256 × 384. This dataset had 100 photos in each of the ten categories—Africa, buildings, beaches, horses, buses, dinosaurs, elephants, trees, food, and mountains. Other CBIR systems were also tested with the help of this data collection. It is frequently employed because the dataset class data is robust in size and availability. Every dataset was additionally separated into test and training datasets. In the training data set, every training class had 80% of the photos, whereas the test data contained 20% of the pictures, as shown in Table 3.

**Table 3.** Dataset images count

| Dataset | Total Images | Classes | Training Images | Testing Images |
|---|---|---|---|---|
| Kvasir | 4000 | 8 | 3200 | 800 |
| CIFAR-10 | 60,000 | 10 | 48,000 | 12,000 |
| Corel-1k | 1000 | 10 | 800 | 200 |

## 3.4 Evaluation metrics

Recall and precision, two famous performance evaluation metrics, were employed in this work. These metrics are still utilized to assess various documents and CBIR techniques. The following formulas provide the generic recall and precision - equations (20)-(21):

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Recall = \frac{TP}{TP+FN} \tag{21}$$

True positives (TP) are the pictures in the query picture class that the CBIR algorithm correctly identified. In contrast, false positives (FP) are the pictures the system returns that do not match the query picture class. False negatives (FN) stand for images that belong to the query picture class but are not given back by the system. Since recall or precision alone is insufficient to evaluate a CBIR system's effectiveness, we measured the proposed system's performance using both recall and accuracy.

### 3.5 Evaluation of the Kvasir dataset

In this investigation, the proposed CBIR method was tested using a small number of pictures as queries. Next, comparable photos were taken from the Kvasir dataset for a particular query image. Figure 4 displays the query image and the matching, comparable images obtained from the database.

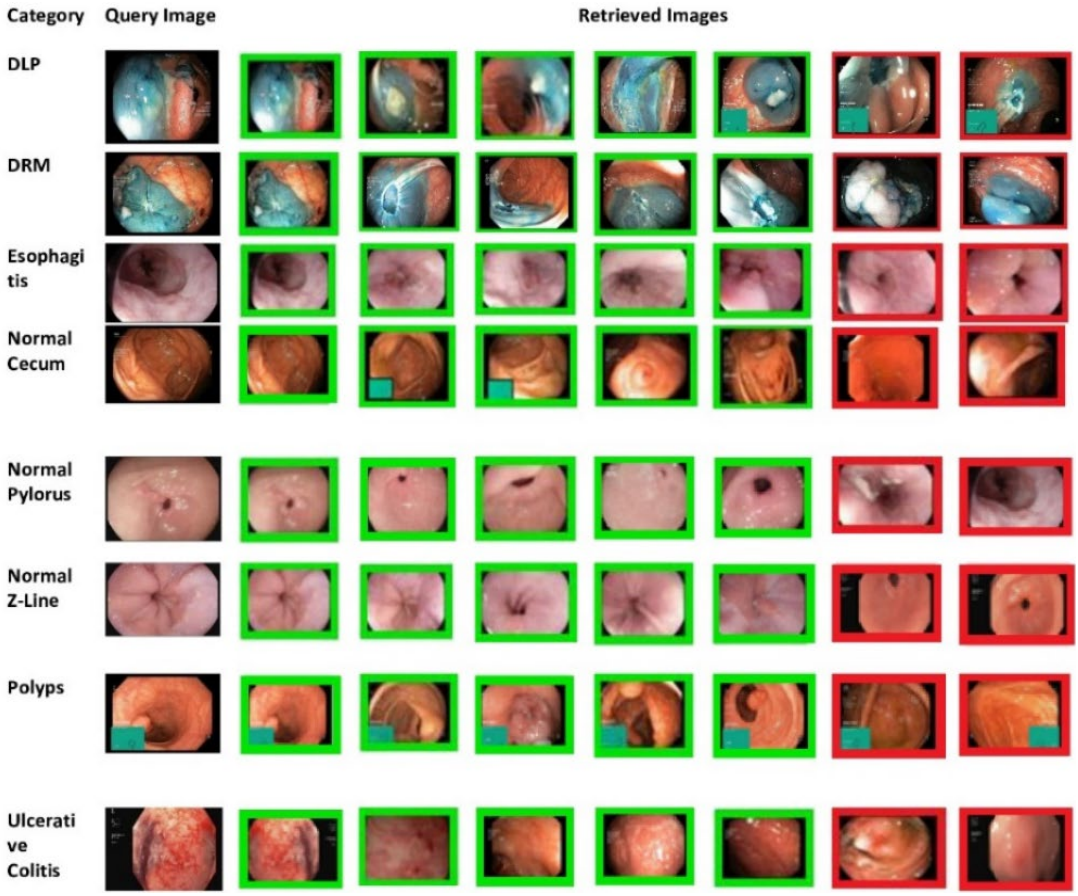



**Figure 4.** The Kvasir dataset's representative retrieval outcomes for a single query image are displayed. Samples from the opponent class are marked in red, while images returned from the same class as the query picture are marked with a green bounding box.

Precision, another name for specificity, is a capacity measure used by image retrieval systems, which exactly look like query images. Simultaneously, sensitivity, also referred to as recall rate, assesses the ability of CBIR systems to retrieve images that are exact replicas of query images (QIs). The results are elaborated by computing recall and precision measures based on the number of query pictures (from the test dataset) and retrieval pictures (from the Kvasir dataset). Table 4 shows the results of the average precision and average recall of the proposed approach in the Kvasir dataset.

Figure 5 shows a graphical illustration of the recall and precision of several query images. The provided CBIR system demonstrated positive precision and recall outcomes, as can be seen from the graph. The enhanced performance of the CBIR approach is demonstrated by the experimental results obtained from the extracted features utilizing the SE-ResNeXt-101 technique, the Improved ShuffleNetV2, and the Improved Mayfly Optimization Algorithm. The comparison of findings from all of the techniques reported eight categories, including DLP, DRM, esophagitis, normal caecum, normal pylorus, normal Z-Line, polyps, and ulcerative colitis, is the reason for selecting this method—an analysis comparing average precision (AP). Table 5 and Table 6 also provide the average recall (AR) for the system being given compared to the other systems.

**Table 4.** Findings from the Kvasir dataset regarding the proposed technique's average precision and recall

| Semantic Class | Precision | Recall |
|---|---|---|
| Dyed Resection Margins (DRM) | 0.952 | 0.972 |
| Esophagitis | 0.965 | 0.946 |
| Dyed Lifted Polyps (DLP) | 0.96 | 1.00 |
| Normal Z-Line | 0.964 | 0.974 |
| Normal Caecum | 1.00 | 0.964 |
| Normal Pylorus | 0.972 | 1.00 |
| Polyps | 0.973 | 1.00 |
| Ulcerative Colitis | 1.00 | 1.00 |
| Average | 0.9734 | 0.97925 |

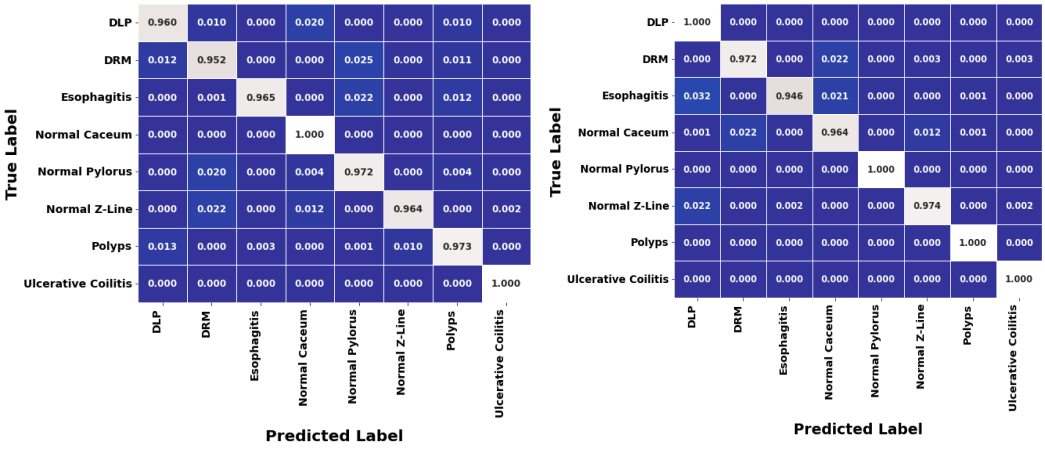



**Figure 5.** Confusion matrix for all the categories of the Kvasir dataset

**Table 5.** Comparison of average precision values with different approaches utilizing the Kvasir dataset

| Semantic Class | RFRM | Bray | Canberra | City Block | Cosine | Proposed Model |
|---|---|---|---|---|---|---|
| DLP | 0.90 | 0.54 | 0.48 | 0.54 | 0.45 | 0.96 |
| DRM | 0.90 | 0.48 | 0.49 | 0.49 | 0.48 | 0.952 |
| Esophagitis | 0.85 | 0.47 | 0.49 | 0.49 | 0.47 | 0.965 |
| Normal Caceum | 0.85 | 0.68 | 0.63 | 0.67 | 0.75 | 1.00 |
| Normal Pylorus | 0.85 | 0.76 | 0.76 | 0.76 | 0.74 | 0.972 |
| Normal Z-Line | 0.90 | 0.57 | 0.56 | 0.57 | 0.56 | 0.964 |
| Polyps | 0.85 | 0.53 | 0.57 | 0.54 | 0.46 | 0.973 |
| Ulcerative Colitis | 0.80 | 0.51 | 0.51 | 0.51 | 0.48 | 1.00 |
| Average | 0.862 | 0.567 | 0.561 | 0.571 | 0.541 | 0.9734 |

**Table 6.** Comparison of average recall values with different approaches utilizing the Kvasir dataset

| Semantic Class | RFRM | Bray | Canberra | City Block | Cosine | Proposed Model |
|---|---|---|---|---|---|---|
| DLP | 0.191 | 0.138 | 0.137 | 0.139 | 0.136 | 1.00 |
| DRM | 0.190 | 0.142 | 0.142 | 0.141 | 0.142 | 0.972 |
| Esophagitis | 0.180 | 0.138 | 0.139 | 0.138 | 0.140 | 0.946 |
| Normal Caceum | 0.191 | 0.142 | 0.142 | 0.141 | 0.142 | 0.964 |
| Normal Pylorus | 0.200 | 0.138 | 0.141 | 0.142 | 0.138 | 1.00 |
| Normal Z-Line | 0.180 | 0.140 | 0.138 | 0.139 | 0.140 | 0.974 |
| Polyps | 0.200 | 0.140 | 0.141 | 0.140 | 0.140 | 1.00 |
| Ulcerative Colitis | 0.190 | 0.140 | 0.140 | 0.140 | 0.140 | 1.00 |
| Average | 0.191 | 0.138 | 0.140 | 0.140 | 0.140 | 0.97925 |

The aforementioned comparative results, as supported by the evidence, show that the proposed method can produce higher recall and precision rates. Regarding recall rates and precision, it performs better than other cutting-edge work. Specifically, the AP and AR rates obtained were 0.9734 and 0.97925. This was influenced by the authors' CBIR systems being created using a limited number of feature sets, which limits retrieval effectiveness.

## 3.6 Evaluation of CIFAR-10 dataset

This study used a few pictures as test queries for the proposed approach. Next are related pictures taken from the CIFAR-10 dataset for a particular query image. Figure 6 displays the query image and the matching, comparable images obtained from the database. Recall and precision metrics were calculated using the number of query pictures and obtaining pictures from the CIFAR-10 dataset to provide detailed results. Table 7 shows the results of the average precision and average recall of the proposed approach in the CIFAR-10 dataset.


**Figure 6.** The CIFAR-10 dataset's representative retrieval outcomes for a single query image are displayed. Samples from the opponent class are marked in red, while images returned from the same class as the query picture are marked with a green bounding box.

**Table 7.** Findings from the CIFAR-10 dataset regarding the proposed technique's average precision and average recall

| Semantic Class | Precision | Recall |
|---|---|---|
| Airplane | 0.972 | 0.954 |
| Bird | 1.00 | 0.972 |
| Deer | 0.965 | 0.965 |
| Frog | 0.984 | 1.00 |
| Ship | 1.00 | 1.00 |
| Buttercup | 0.987 | 0.974 |
| Automobile | 0.961 | 1.00 |
| Cat | 1.00 | 0.981 |
| Dog | 1.00 | 1.00 |
| Horse | 0.968 | 1.00 |
| Truck | 1.00 | 0.957 |
| Average | 0.9837 | 0.9803 |

A graphical representation of the recall and precision of many query pictures is presented in Figure 7. The graph shows that the offered system produced positive results regarding precision and recall. The experimental outcomes produced from the extracted attributes using the Improved ShuffleNetV2, the Improved Mayfly Optimization Algorithm, and the SE-ResNeXt-101 method illustrate the improved performance. The rationale behind this approach is to compare results from all the methods given in eight categories: automobile, airplane, cat, dog, horse, truck, bird, deer, frog, and ship. The AR for the system being offered about the other methods is also provided in Table 8 and Table 9.

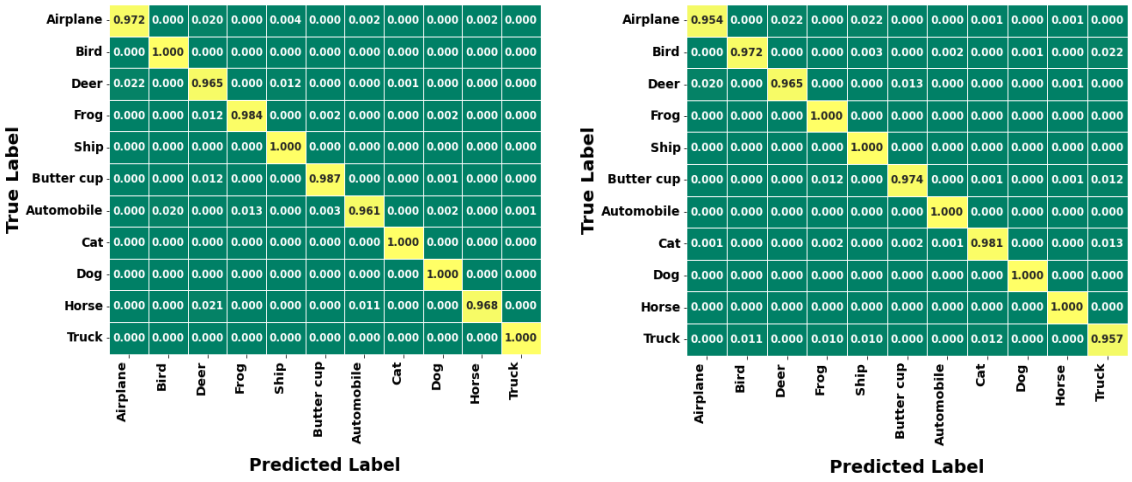



**Figure 7.** Confusion matrix for all the categories of the CIFAR-10 dataset

**Table 8.** Comparison of average precision values with different approaches utilizing the CIFAR-10 dataset

| Semantic Class | GA | CTCHIRS | MCM | EDPS | Proposed Model |
|---|---|---|---|---|---|
| Automobile | 0.705 | 0.681 | 0.432 | 0.395 | 0.961 |
| Airplane | 0.762 | 0.732 | 0.521 | 0.568 | 0.972 |
| Cat | 0.921 | 0.910 | 0.861 | 0.826 | 1.00 |
| Dog | 0.854 | 0.756 | 0.435 | 0.482 | 1.00 |
| Horse | 0.893 | 0.842 | 0.692 | 0.736 | 0.968 |
| Truck | 0.826 | 0.764 | 0.642 | 0.689 | 1.00 |
| Bird | 0.816 | 0.789 | 0.365 | 0.464 | 1.00 |
| Deer | 0.918 | 0.872 | 0.821 | 0.782 | 0.965 |
| Frog | 0.768 | 0.743 | 0.692 | 0.719 | 0.984 |
| Ship | 0.754 | 0.682 | 0.425 | 0.568 | 1.00 |
| Average | 0.821 | 0.777 | 0.589 | 0.622 | 0.9837 |

**Table 9.** Comparison of average recall values with different approaches utilizing the CIFAR-10 dataset

| Semantic Class | GA | CTCHIRS | MCM | EDPS | Proposed Model |
|---|---|---|---|---|---|
| Automobile | 0.235 | 0.224 | 0.138 | 0.127 | 1.00 |
| Airplane | 0.178 | 0.167 | 0.129 | 0.146 | 0.954 |
| Cat | 0.137 | 0.136 | 0.115 | 0.118 | 0.981 |
| Dog | 0.169 | 0.163 | 0.151 | 0.136 | 1.00 |
| Horse | 0.153 | 0.148 | 0.128 | 0.125 | 1.00 |
| Truck | 0.162 | 0.153 | 0.141 | 0.147 | 0.957 |
| Bird | 0.210 | 0.164 | 0.153 | 0.158 | 0.972 |
| Deer | 0.115 | 0.110 | 0.084 | 0.112 | 0.965 |
| Frog | 0.134 | 0.128 | 0.107 | 0.121 | 1.00 |
| Ship | 0.258 | 0.231 | 0.162 | 0.182 | 1.00 |
| Average | 0.175 | 0.165 | 0.130 | 0.137 | 0.9803 |

Regarding recall rates and precision, it performed better than other state-of-the-art components. Specifically, the AP and AR rates obtained were 0.9837 and 0.9803. This was likely influenced by the authors' CBIR systems being created using a limited number of feature sets, which limits retrieval effectiveness. Conversely, the approach proposed in this work retrieved an effective and wide range of properties.

## 3.7 Evaluation of Corel-1k dataset

In the present research, just a few pictures were used as test queries for the proposed method. Next, related pictures were taken from the Corel-1k dataset for a specific query picture. Figure 8 displays the query picture and the matching, comparable images obtained from the dataset. The findings of the calculations on recall and precision evaluations based on the number of query pictures and the pictures retrieved from the Corel-1k dataset are detailed. The outcome of Corel-1k datasets using proposed model based on average accuracy and recall values is shown in Table 10. Figure 9 displays a graphical illustration of the recall and precision of various QI for the proposed approach. The provided CBIR system demonstrated positive precision and recall outcomes, as can be seen from the graph. The number of existing CBIR systems was used to assess the effectiveness of the proposed system. The comparison of the outcomes of all the ways given via the Corel-1k database of ten categories—buildings, mountains, African tribe people, food, beaches, horses, buses, flowers, elephants, and dinosaurs—is the motivator for choosing this approach. Table 11 also compares the AP and AR for the system being presented and the other systems that were studied. The comparison of average recall values using the proposed model in Corel-1k dataset is demonstrated in Table 12. Regarding accuracy and recall rates, it performed better than other state-of-the-art work. Specifically, the AP and AR rates achieved were 0.9734 and 0.9745, respectively. This was likely influenced by the fact that the authors created systems with a limited number of feature sets, which limit retrieval effectiveness.


**Figure 8.** The Corel-1k dataset's representative retrieval outcomes for a single query image are displayed. Samples from the opponent class are marked in red, while images returned from the same class as the query picture are marked with a green bounding box.

**Table 10.** Findings from the Corel-1k dataset regarding the proposed technique's average accuracy and recall

| Semantic Class | Precision | Recall |
|---|---|---|
| Beach | 0.957 | 0.942 |
| buses | 0.938 | 0.968 |
| Flower | 0.972 | 0.951 |
| Food | 1.00 | 1.00 |
| Elephants | 0.967 | 1.00 |
| Africa | 1.00 | 0.954 |
| buildings | 0.976 | 1.00 |
| dinosaurs | 0.948 | 1.00 |
| Horse | 1.00 | 0.975 |
| Mountain | 1.00 | 0.976 |
| Average | 0.9734 | 0.9745 |

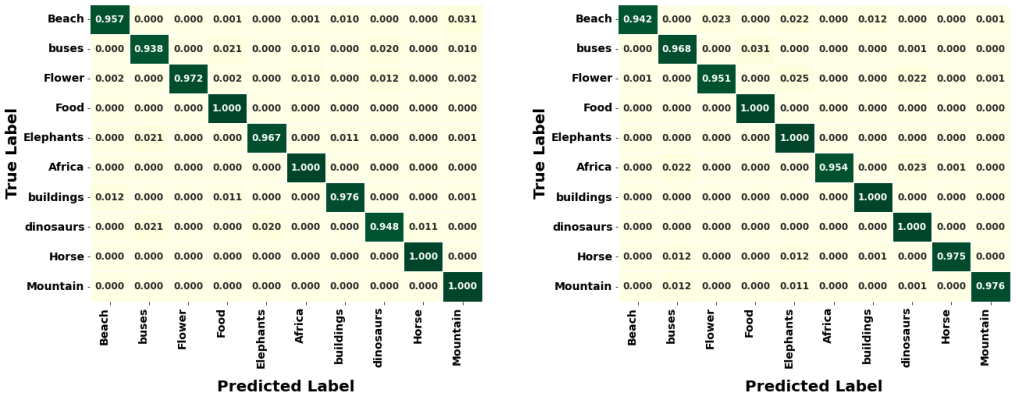



**Figure 9.** Confusion matrix for all the categories of the Corel-1k dataset

**Table 11.** Comparison of average precision values with approaches utilizing the Corel-1k dataset

| Semantic Class | SVM | K-means | BiCBIR | Tetrolet Transform+Edge Histogram | Proposed Method |
|---|---|---|---|---|---|
| Beach | 0.9656 | 0.76 | 0.750 | 0.600 | 0.957 |
| Buses | 0.833 | 0.79 | 1.000 | 1.000 | 0.938 |
| Elephants | 0.775 | 0.7 | 0.900 | 0.900 | 0.967 |
| Horse | 0.863 | 0.7 | 1.000 | 1.000 | 1.00 |
| Food | 0.861 | 0.56 | 0.950 | 1.000 | 1.00 |
| Africa | 0.812 | 0.72 | 0.950 | 0.950 | 1.00 |
| Buildings | 0.782 | 0.55 | 0.850 | 0.550 | 0.976 |
| Dinosaurs | 0.822 | 1 | 1.000 | 1.000 | 0.948 |
| Flower | 0.863 | 0.87 | 1.000 | 1.000 | 0.972 |
| Mountain | 0.902 | 0.58 | 0.800 | 0.750 | 1.00 |
| Average | 0.847 | 0.723 | 0.920 | 0.875 | 0.9734 |

**Table 12.** Comparison of average recall values with approaches utilizing the Corel-1k dataset

| Semantic Class | SVM | K-means | BiCBIR | Tetrolet Transform+Edge Histogram | Proposed Method |
|---|---|---|---|---|---|
| Beach | 0.87 | 0.152 | 0.150 | 0.120 | 0.942 |
| Buses | 0.80 | 0.158 | 0.200 | 0.200 | 0.968 |
| Elephants | 0.83 | 0.14 | 0.180 | 0.180 | 1.00 |
| Horse | 0.82 | 0.14 | 0.200 | 0.200 | 0.975 |
| Food | 0.87 | 0.112 | 0.190 | 0.200 | 1.00 |
| Africa | 0.82 | 0.144 | 0.190 | 0.190 | 0.954 |
| Buildings | 0.90 | 0.11 | 0.170 | 0.110 | 1.00 |
| Dinosaurs | 0.79 | 0.2 | 0.200 | 0.200 | 1.00 |
| Flower | 0.82 | 0.174 | 0.200 | 0.200 | 0.951 |
| Mountain | 0.93 | 0.116 | 0.160 | 0.150 | 0.976 |
| Average | 0.845 | 0.1446 | 0.184 | 0.175 | 0.9745 |

### 3.8 Overall comparison

Table 13 provides a thorough comparison of different image classification techniques, encompassing established methods from the literature and a devised method, which is OptISNV2. Average precision and average recall, two essential variables across several datasets, formed the basis of the evaluation.

**Table 13.** Overall comparison of proposed and existing methods in literature

| Reference | Method | Dataset | Average Precision | Average Recall |
|---|---|---|---|---|
| Khan et al. (2021) | GA and SVM | CIFAR-10 | 0.916 | 0.183 |
| | | Kvasir | 0.913 | 0.218 |
| Desai et al. (2021) | CNN and SVM | Corel-1k | 0.84 | - |
| Karthikeyan & Raja (2023) | DTLDN-CBIRA | Apple Plant disease | 0.735 | 0.502 |
| | | Grape disease | 0.975 | 0.819 |
| Alsmadi (2020) | DWT | Corel-1k | 0.9015 | 0.1803 |
| Garg & Dhiman (2021) | SVM | Corel-1k | 0.847 | 0.845 |
| Proposed Method | OptISNV2 | Kvasir | 0.9734 | 0.97925 |
| | | CIFAR-10 | 0.9837 | 0.9803 |
| | | Corel-1k | 0.9734 | 0.9745 |

On the CIFAR-10 and Kvasir datasets, the Genetic Algorithms (GA) and Support Vector Machines (SVM) combination presented in Khan et al. (2021) performed competitively with significant precision and recall levels. On the CIFAR-10 dataset, the GA and SVM approach produced average precision and recall values of 0.916 and 0.183, respectively. The average precision and recall for the Kvasir dataset were 0.913 and 0.218, respectively. On the Corel-1k dataset, SVM and Convolutional Neural Networks (CNN) in Desai et al. (2021) show an appropriate average precision. However, there was no information available regarding average recall. Significant capabilities are demonstrated by the DTLDN-CBIRA technique in Karthikeyan and Raja (2023), which performed exceptionally well on a grape disease dataset in terms of recall and precision. On the apple plant disease dataset, the approach yielded an average precision of 0.735 and an average recall of 0.502 using DTLDN-CBIRA. With an average recall of 0.819 and an average precision of 0.975 for the grape disease dataset, the results were noticeably better. On the Corel-1k dataset, the Discrete Wavelet Transform (DWT) in Alsmadi (2020) achieved excellent precision despite having a slightly low recall value. Using the DWT, the approach's average precision and average recall were 0.9015 and 0.1803, respectively. Using the Corel-1k dataset, the application of SVM in Garg & Dhiman (2021) shows balanced precision and recall levels. The SVM-based approach produced an impressive average recall of 0.845 and an average precision of 0.847. OptISNV2, our proposed method, demonstrates an incredible average precision of 0.9734 and an outstanding average recall of 0.97925 on the Kvasir dataset, outperforming existing methods. With an average precision of 0.9837 and an excellent average recall of 0.9803 for CIFAR-10,

OptISNV2 continued to perform exceptionally well. On the Corel-1k dataset, OptISNV2 continued to perform well, obtaining an average precision of 0.9734 and an average recall of 0.9745. Based on the experiments on multiple datasets, the proposed method, OptISNV2, consistently performed better than the referenced methods.

### 3.9 Discussion

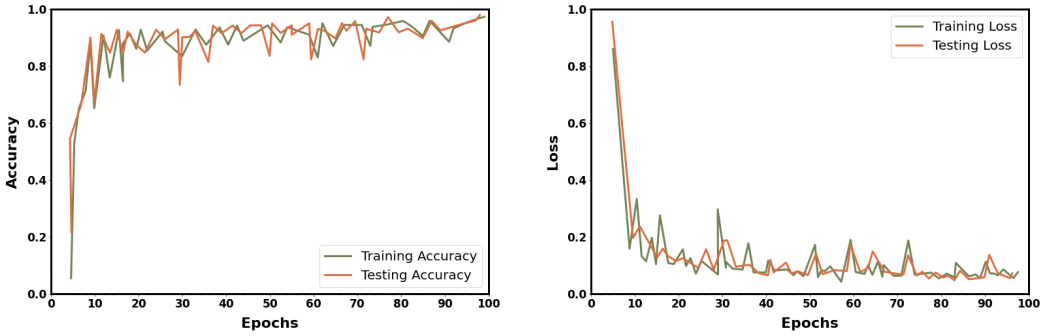
The proposed model, OptISNV2, demonstrates remarkable performance in image classification across multiple datasets, including Kvasir, CIFAR-10, and Corel-1k. Its effectiveness is evident from its consistently high average precision and average recall scores, which surpass those of existing methods. The combination of advanced feature extraction, optimized classification, and intelligent search strategies contributes to the model's superior results. One of the key strengths of OptISNV2 lies in its ability to extract and utilize diverse image features, including shape, texture, and color attributes, through the SE-ResNeXt-101 architecture. This deep learning model enhances feature representation, enabling more precise similarity matching between query and database images. By leveraging the Improved ShuffleNetV2 (ISNV2) for classification, the model benefits from high computational efficiency and strong classification accuracy, making it suitable for large-scale datasets. The integration of lightweight yet powerful architecture allows OptISNV2 to maintain efficiency without compromising performance.

Furthermore, OptISNV2 incorporates the Improved Mayfly Optimization Algorithm (IMOA) to optimize the image retrieval process. This evolutionary-based optimization technique effectively balances exploration and exploitation during similarity matching, leading to better retrieval accuracy and reduced computational overhead. IMOA refines the search process, ensuring that the most relevant images are retrieved with minimal error, which significantly enhances the model's robustness. The performance analysis presented in Table 13 highlights OptISNV2's superiority over traditional techniques such as GA-SVM, CNN-SVM, DWT, and DTLDN-CBIRA. While approaches like GA-SVM and DWT achieve high precision in specific datasets, they suffer from low recall, indicating that these methods might miss a significant number of relevant images. DTLDN-CBIRA, on the other hand, shows impressive results in plant disease classification but lacks generalizability across diverse datasets. OptISNV2, however, maintains a balanced and superior performance across all three datasets, reinforcing its ability to handle varied image complexities and categories.

For instance, in the Kvasir dataset, OptISNV2 achieved an outstanding average precision of 0.9734 and an average recall of 0.97925, significantly outperforming GA-SVM, which had a recall as low as 0.218. Similarly, on CIFAR-10, the proposed method attained 0.9837 precision and 0.9803 recall, making it the most efficient model for this dataset. The Corel-1k dataset also saw OptISNV2 lead the performance metrics, achieving 0.9734 precision and 0.9745 recall, outperforming previous SVM-based and CNN-based approaches. Overall, the results demonstrate that OptISNV2 is a highly robust and generalizable image classification model. Its ability to maintain high precision and recall across multiple datasets underscores its effectiveness in handling diverse image types and complexities. By integrating advanced feature extraction, efficient classification, and optimized retrieval strategies, OptISNV2 sets a new benchmark for image classification and retrieval tasks.

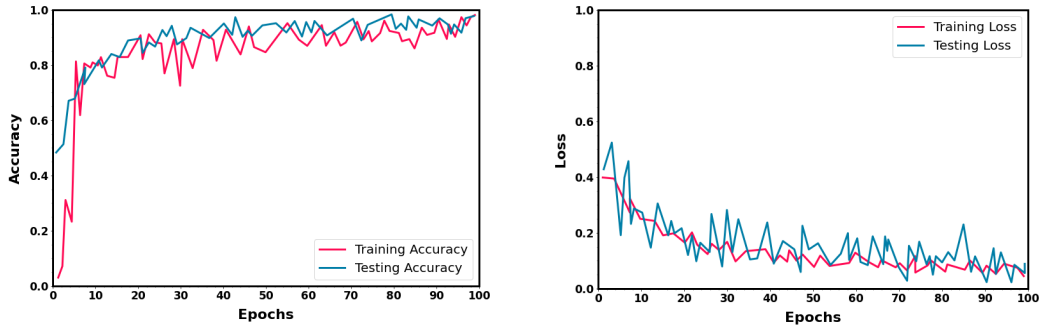
### 3.10 Evaluation of testing and training

Assessing training and testing is essential in developing and accessing DL models. It uses two sets of data—the training set and the testing set—to assess a model's functionality and generalization potential. The assessment of the performance of the proposed approach's training and accuracy and loss testing is shown in Figures 10, 11, and 12.



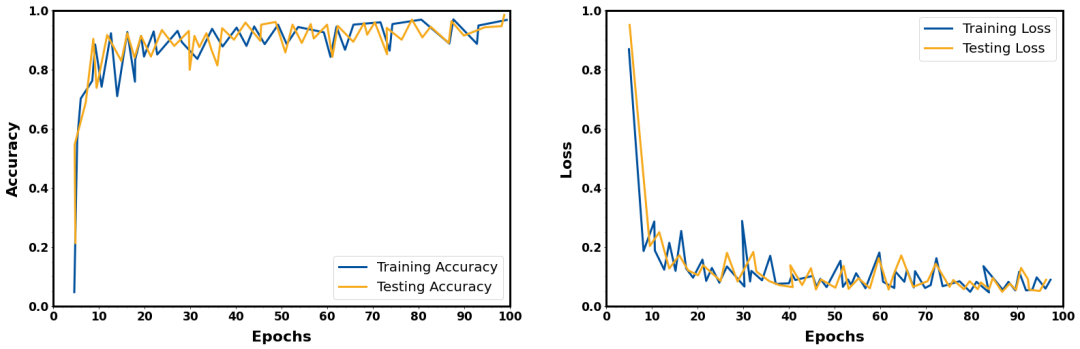

**(a)** **(b)**

**Figure 10.** Performance of training and testing of Kvasir dataset, (a) accuracy, (b) loss




**(a)** **(b)**

**Figure 11.** Performance of training and testing of CIFAR-10 dataset, (a) accuracy, (b) loss




(a) (b)

**Figure 12.** Performance of training and testing of Corel-1k dataset, (a) accuracy, (b) loss

The systematic approach to testing and training evaluation involves assessing a model's performance on the data it was trained on and the data that was discovered. Real-world predictions that are accurate indicate that the model has picked up essential patterns and is capable of broad generalization.

### 3.11 Computational time complexity

Another concept discussed is that of computation time. Simplifying computations is the technique's aim. In Table 14, the computation times of the proposed method are contrasted with those of previously used methods. With minimal computing work, prediction accuracy is raised. Computation time using the most recent methods and the proposed approach are shown in Figure 13. The proposed approach achieved lower computing time and higher prediction accuracy than other approaches.

**Table 14.** Comparison of overall time complexity for proposed and existing methods

| Reference | Method | Computational Time (ms) |
|---|---|---|
| Khan et al. (2021) | GA and SVM | 0.37 |
| Desai et al. (2021) | CNN and SVM | 0.29 |
| Karthikeyan & Raja (2023) | DTLDN-CBIRA | 0.39 |
| Alsmadi (2020) | DWT | 0.32 |
| Garg & Dhiman (2021) | SVM | 0.27 |
| Proposed Model | OptISNV2 | 0.19 |

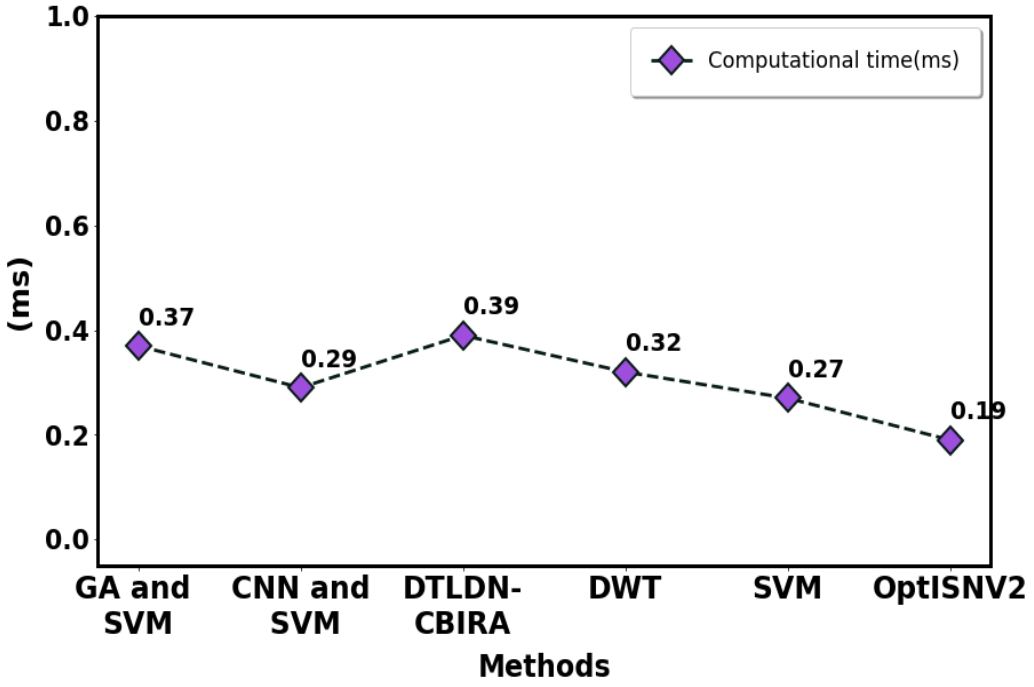



**Figure 13.** Overall comparison of computation time for proposed and existing methods

## 4. Conclusions

To find the relevant pictures and minimize the search space, we offer a content-based picture retrieval system that uses a DL architecture for the first classification of query picture. Our approach lowers the computational overheads and raises the system's precision. This research uses normalization, contrast enhancement, and noise reduction during pre-processing to improve comprehension of the image's features. Then, using the SE-ResNeXt-101 technique, the picture's texture, color, and form are retrieved for similarity matching. Next, these feature vectors are applied to the query and database images to categorize the dataset images using the Improved ShuffleNetV2 approach. Additionally, this method computes picture similarity to identify the most similar database images the query provides. Retrieval performance is improved by employing the Improved Mayfly Optimization Algorithm (IMOA). The proposed model OptISNV2 achieved 0.9734 precision and 0.97925 recall for the Kvasir dataset, 0.9837 precision and 0.9803 recall for the CIFAR-10 dataset, and 0.9734 precision and 0.9745 recall for the Corel-1k dataset. Compared to the previous research in CBIR, the proposed model achieved higher precision and recall values with less computation time. Further research should be done to create high-performing, more straightforward similarity or distance measurements based on any other mathematical idea.

## 5. Authors' Contributions

**Rajasekharam Gulla, S. Saradha Rani:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Writing – Original Draft, Writing - Review & Editing, Visualization.
**Rajyalakshmi Uppada:** Conceptualization, Writing - Review & Editing, Original Draft, Writing - Review & Editing, Visualization.

## 6. Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**ORCID**

Rajasekharam Gulla https://orcid.org/0009-0005-4473-7762

## References

Alsmadi, M. K. (2020). Content-based image retrieval using color, shape, and texture descriptors and features. *Arabian Journal for Science and Engineering*, 45(4), 3317-3330. https://doi.org/10.1007/s13369-020-04384-y

Chen, Z., Yang, J., Chen, L., & Jiao, H. (2022). Garbage classification system based on improved ShuffleNet v2. Resources, *Conservation and Recycling*, 178, Article 106090. https://doi.org/10.1016/j.resconrec.2021.106090

Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M. J., Jeong, J., Lee, Y., Jin, K., Park, R., Kim, J., Jeon, H., Kim, N., Yi, J., Yu, D., & Kim, B. (2022). Content-based image retrieval using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology*, 302(1), 187-197. https://doi.org/10.1148/radiol.2021204164

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, Article 170. https://doi.org/10.1007/s42979-021-00529-4

Garg, M., & Dhiman, G. (2021). A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants. *Neural Computing and Applications*, 33, 1311-1328. https://doi.org/10.1007/s00521-020-05017-z

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2023, November). Content-based image retrieval using fused convolutional neural networks. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 260-270). Springer International Publishing. https://doi.org/10.1007/978-3-031-20601-6_24

Haddadi, Y. R., Mansouri, B., & Khodja, F. Z. I. (2024). A novel medical image enhancement algorithm based on CLAHE and pelican optimization. *Multimedia Tools and Applications*, 83(42), 90069-90088. https://doi.org/10.1007/s11042-024-19070-6

Higaki, A., Kawaguchi, N., Kurokawa, T., Okabe, H., Kazatani, T., Kido, S., Aono, T., Matsuda, K., Tanaka, Y., Hosokawa, S., Kosaki, T., Kawamura, G., Shigematsu, T., Kawada, Y., Hiasa, G., Yamada, T., & Okayama, H. (2023). Content-based image retrieval for the diagnosis of myocardial perfusion imaging using a deep convolutional autoencoder. *Journal of Nuclear Cardiology*, 30(2), 540-549. https://doi.org/10.1007/s12350-022-03030-4

Hu, H., Zheng, W., Zhang, X., Zhang, X., Liu, J., Hu, W., & Si, J. (2021). Content-based gastric image retrieval using convolutional neural networks. *International Journal of Imaging Systems and Technology*, 31(1), 439-449. https://doi.org/10.1002/ima.22470

Hung, B. T., & Phuong, P. H. (2022). Content based image retrieval based on deep learning approach. In *Computer vision and robotics: Proceedings of CVR 2021* (pp. 319-328). Springer Singapore. https://doi.org/10.1007/978-981-16-8225-4_25

Karthikeyan, M., & Raja, D. (2023). Deep transfer learning enabled DenseNet model for content based image retrieval in agricultural plant disease images. *Multimedia Tools and Applications*, 82, 36067-36090. https://doi.org/10.1007/s11042-023-14992-z

Kayhan, N., & Fekri-Ershad, S. (2021). Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns. *Multimedia Tools and Applications*, 80, 32763-32790. https://doi.org/10.1007/s11042-021-11217-z

Khan, U. A., Javed, A., & Ashraf, R. (2021). An effective hybrid framework for content-based image retrieval (CBIR). *Multimedia Tools and Applications*, 80(17), 26911-26937. https://doi.org/10.1007/s11042-021-10530-x

Kumar, S., Pradhan, J., & Pal, A. K. (2021). Adaptive tetrolet based color, texture and shape feature extraction for content based image retrieval application. *Multimedia Tools and Applications*, 80(19), 29017-29049. https://doi.org/10.1007/s11042-021-10835-x

Lei, G., Chang, X., Tianhang, Y., & Tuerxun, W. (2022). An improved mayfly optimization algorithm based on median position and its application in the optimization of PID parameters of hydro-turbine governor. *IEEE Access*, 10, 36335-36349. https://doi.org/10.1109/ACCESS.2022.3160714

Majhi, M., Pal, A. K., Islam, S. K. H., & Khan, M. K. (2021). Secure content-based image retrieval using modified Euclidean distance for encrypted features. *Transactions on Emerging Telecommunications Technologies*, 32(2), Article e4013. https://doi.org/10.1002/ett.4013

Mezzoudj, S., Behloul, A., Seghir, R., & Saadna, Y. (2021). A parallel content-based image retrieval system using spark and tachyon frameworks. *Journal of King Saud University-Computer and Information Sciences*, 33(2), 141-149. https://doi.org/10.1016/j.jksuci.2019.01.003

Schall, K., Barthel, K. U., Hezel, N., & Jung, K. (2022). GPR1200: a benchmark for general-purpose content-based image retrieval. In *International Conference on Multimedia Modeling* (pp. 205-216). Springer International Publishing. https://doi.org/10.1007/978-3-030-98358-1_17

Sivakumar, M., Kumar, N. M. S., & Karthikeyan, N. (2022). An efficient deep learning-based content-based image retrieval framework. *Computer Systems Science and Engineering*, 43(2), 683-700. https://doi.org/10.32604/csse.2022.021459

Subramanian, M., Lingamuthu, V., Venkatesan, C., & Perumal, S. (2022). Content-based image retrieval using colour, gray, advanced texture, shape features, and random forest classifier with optimized particle swarm optimization. *International Journal of Biomedical Imaging*, 2022, Article 3211793. https://doi.org/10.1155/2022/3211793

Tankala, Y., Paul, J. K., & Manikandan, V. M. (2022). A content-based image retrieval scheme with object detection and quantised colour histogram. *International Journal of Computational Science and Engineering*, 25(4), 367-374. https://doi.org/10.1504/IJCSE.2022.124558

Trappey, A. J., Trappey, C. V., & Shih, S. (2021). An intelligent content-based image retrieval methodology using transfer learning for digital IP protection. *Advanced Engineering Informatics*, 48, Article 101291. https://doi.org/10.1016/j.aei.2021.101291

Zhao, M., Liu, J., Zhang, Z., & Fan, J. (2021). A scalable sub-graph regularization for efficient content-based image retrieval with long-term relevance feedback enhancement. *Knowledge-based Systems*, 212, Article 106505. https://doi.org/10.1016/j.knosys.2020.106505