Current Applied Science and Technology e-ISSN: 2586-9396

Curr. Appl. Sci. Technol. 2026, Vol. 26 (No. 4), e0263944

Research article

An Optimized Feature for Content based Multimedia Image Retrieval System Using Deep Learning Approaches

G. Sai Chaitanya Kumar[1]*, V. Srilakshmi[2], G. N. Beena Bethel[3], Narendhar Mulugu[4] and M. V. Kamal[5]

[1]Department of Artificial Intelligence, DVR & Dr. HS MIC College of Technology, Kanchikcherla, Andhra Pradesh, India
[2]Department of Computer Science and Engineering, GRIET(Autonomous), Hyderabad, Telangana, India
[3]Department of Computer Science and Engineering (Data Science), Sridevi Women's Engineering College (Autonomous), Hyderabad, Telangana, India
[4]Department of Computer Science and Engineering (AIML) at Malla Reddy Engineering College for Women (Autonomous), Hyderabad, Telangana, India
[5]Department of Computer Science and Engineering (ET), Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India

Received: 9 July 2024, Revised: 19 May 2025, Accepted: 19 May 2025, Published: 10 October 2025

Abstract

The World Wide Web and developments in computer and multimedia technologies have led to increased picture databases and collections such as digital libraries, medical imageries, and art galleries, which collectively contain millions of pictures. Developing an efficient image retrieval system that can manage these enormous volumes of pictures at once is essential. The major goal of this study was to create a reliable system that could efficiently create, manage, and react to data. An effective tool for retrieving images was found to be the content-based image retrieval (CBIR) system, which allows users to query the system to retrieve their desired image from the image collection. In addition, the variety of pictures that users can access, and the expansion of online development and transmission networks have continued to increase. In this paper, we proposed employing an Improved Mobilenetv3 method for picture retrieval. To preprocess the images, we applied noise reduction with a median filter, normalization using the min-max normalization method, and contrast enhancement using Adaptively Clipped Contrast Limited Adaptive Histogram Equalization (ACCLAHE). Then, a Modified ResNet152V2 model was employed to extract detailed features related to shape, texture, and color. After that, the Quantum Chaotic Honey Badger Algorithm (QCHBA) was utilized to select the most relevant features, improving computational efficiency and performance. Finally, the images were classified using the Improved MobileNetV3 technique, which was optimized for high accuracy and efficiency. The performance of the image retrieval framework for content-based retrieval was improved by combining these techniques. Furthermore, the precision-recall value of the outcomes was computed to assess the effectiveness of the system.

*Corresponding author: E-mail: saichaitanyakumar657@gmail.com
https://doi.org/10.55003/cast.2025.263944
Copyright © 2025 by King Mongkut's Institute of Technology Ladkrabang, Thailand. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Content-based image retrieval; feature extraction; improved mobilenetv3; contrast enhancement; modified resnet152v2; preprocessing

1. Introduction

Developing efficient image content management solutions has become a research priority due to the exponential development of digital pictures in cyberspace. Over the past ten years, the exponential growth of digital equipment and social media has contributed to a notable spike in image files (Fadaei et al., 2024; Ranjith et al., 2024; Vu et al., 2024). The research community has been motivated to investigate efficient methods without utilizing text-based descriptions for each image by obtaining relevant data from these enormous collections. The solution of content-based image retrieval (CBIR) was an outcome of this research. Using the visual contents of digital pictures, CBIR is a function of computer vision that facilitates the search of large picture datasets using digital image retrieval systems (Shetty et al., 2024; Zhang et al., 2024). Applications for CBIR have been created for various uses, including object recognition, remote sensing, surveillance systems, geographic information systems, architectural design, and medical image retrieval.

The feature repository uses various visual features, such as color, texture, form, etc., by existing CBIR approaches (Chen et al., 2024; Khunsongkiet et al., 2024). An outcome of a CBIR model that provides the semantic answer of the model against a query picture is a set of pictures that are most similar to each other based on distance. A query by example CBIR model classifies images into their respective classes utilizing weighted Euclidean distance as a similarity metric. The performance of existing CBIR methods decreases when there are several classes, but visual attributes-based CBIR models can achieve greater picture retrieval accuracy (Zhang et al., 2023).

Under some circumstances, image retrieval performance is worse due to the discrepancy between high-level and low-level attribute illustrations in various pictures. Previous techniques have also incorporated a range of learning-based approaches to close this semantic gap and enhance image retrieval performance (Hong et al., 2023; Rashad et al., 2023; Wang et al., 2023). To assess parametric computations, the optimization approach utilized color attributes and curvelet transformation in HSV space in an ideal manner. For these learning-based techniques, gap between low-level feature representations and high-level semantics in various images, are feature-dependent and thus cannot be applied to all types of images for all feature descriptors (Wang et al., 2024). Furthermore, compared to non-learning based CBIR approaches, these learning-based approaches are computationally more difficult. To overcome the issues in existing studies, a new DL-based CBIR approach to retrieve images is proposed. The key contributions of the new proposed model are as follows.

1) The model employs a combination of Median Filtering for noise reduction, Min-Max Normalization for consistent feature scaling, and Adaptively Clipped Contrast Limited Adaptive Histogram Equalization (ACCLAHE) to enhance image contrast. This preprocessing pipeline ensures that input images are clean, well-scaled, and have improved visibility of important features, leading to better feature extraction and classification accuracy.

2) The ResNet152V2 architecture is modified to improve its capability to capture intricate details related to shape, texture, and color. The modifications included are customized convolutional layers, optimized weight initialization, or the inclusion of attention mechanisms. These modifications enhance the network's ability to extract robust and discriminative features, leading to improved generalization and accuracy in classification.

3) A novel feature selection strategy based on QCHBA is integrated to retain only the most relevant features while removing redundant and unnecessary ones. This significantly improves computational efficiency, reduces model complexity, and enhances classification performance by focusing only on the most informative features.

4) The classification stage leverages an optimized version of MobileNetV3, ensuring high efficiency and accuracy in predicting image categories. The model benefits from reduced computational cost, faster inference times, and improved accuracy, making it highly suitable for real-time and resource-constrained environments.

Studies pertaining to existing CBIR techniques were conducted with a primary focus on the research and examination of interest areas/points. The primary competition in this field of study is to improve the precision rate, which is the accuracy with which the most comparable images are correctly retrieved. The key and most recent studies on CBIR are compiled. Rani et al. (2024) presented a medical image retrieval method to retrieve texture attributes from images. The input pictures were taken from Computed Tomography scans. Initially, there were some noise problems affecting the corresponding image. A fractional Hartley transform was used to eliminate the noise and lessen the distortion of the source image to lessen this and enhance the brightness of the pixels in the original picture. The specified features were then extracted using the hybrid feature extraction technique. After that, the MWBMMBO technique was used to eliminate unnecessary attributes from the enormous number of attributes and choose a subset of necessary attributes. Finally, the medical images were detected and classified by measuring the similarity between the desired characteristics. Mahalle et al. (2023) presented an efficient interactive CBIR that uses variable compressed convolutional info neural networks (VCCINN) to reliably retrieve pictures in response to the picture query. The variable info method was used to optimize the neural network's weight, and recursive density matching to handle the matching process. After removing irrelevant pictures based on user input, the interactive method retrieved only the relevant images.

Khan et al. (2021) suggested a CBIR technique based on a hybrid attributes descriptor using the SVM classifier and genetic algorithm (GA) for picture retrieval in a multi-class scenario. More precisely, they extracted attributes utilizing GA and then trained the multi-class SVM utilizing the one-against-all method. The query picture and the retrieved pictures were compared to the query picture from the picture repository using the L2 Norm similarity measurement function. Kelishadrokhi et al. (2023) suggested an ELNDP method based on texture and color attributes. An enhanced version of local neighborhood difference patterns (ELNDP) was used for the first time to achieve discriminative attributes. By using LBP and LNDP texture descriptors, the ELNDP improved color histogram attributes in HSV color space were also utilized to acquire color attributes for general attribute extraction. Salih and Abdulla (2023) presented LBP and DWT to take advantage of global and local features. As a result, a hybrid CBIR technique with two filtering layers was developed. In the first layer, as many dissimilar pictures as possible were eliminated or excluded by comparing the query picture to each picture in the dataset using the Bag of Features (BoF) method. Various pictures closer to the query picture were retrieved as a consequence. By comparing the query picture to the acquired images obtained from the first layer, the second layer sought to understand the picture patterns. The extraction of attributes based on color and texture provided a base. As texture attributes, the DWT and LBP were employed. Additionally, color attributes from three different color spaces—YCbCr, HSV, and RGB—were employed. Some methods, however, failed to consider the semantic gap—that is, the difference between the user's intent and the pictures the algorithm returned. Additionally, most other methods involved evaluation of experimental analysis on small datasets. In theory, CBIR technology aids in managing and retrieving

digital picture archives based on their visual content. An image retrieval system aims to cluster related, nearby images according to their common attributes. The number of images increases with the number of attributes, but memory and processing time also increase. To overcome these existing problems, we proposed a novel deep learning-based technique in the CBIR model.

2. Materials and Methods

A feature extraction approach was applied to extract attributes of a picture utilizing the CBIR system. Color, texture, and shape are visual qualities that are low-level components and salient points in a picture. Additionally, each image's properties are stored in a different database called a database feature. A typical CBIR system is shown in Figure 1.

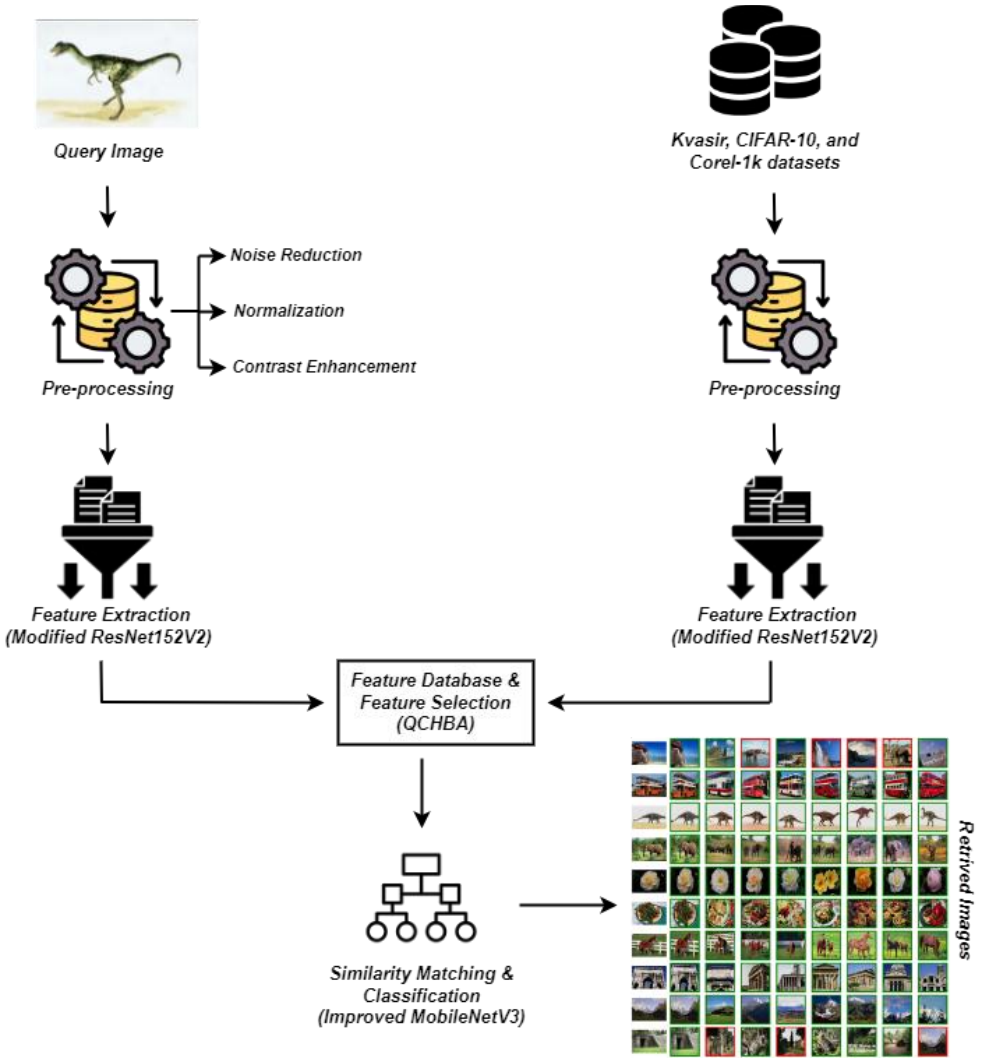

Figure 1. The performance architecture of the proposed model

We initiated preprocessing to improve image quality in the proposed content-based image retrieval system. A median filter was used for noise reduction, efficiently decreasing noise while maintaining edges. Next, pixel values were normalized using min-max normalization, which was set into a common range (0, 1) and guaranteed consistent and equivalent feature extraction. We utilized Adaptively Clipped Contrast Limited Adaptive Histogram Equalization (ACCLAHE) to significantly enhance the picture quality. ACCLAHE reveals finer details by varying the contrast in various places. We utilized a Modified ResNet152V2 model in the feature extraction step, which had been modified to capture fine data about shape, texture, and color, yielding robust and distinctive attributes. Subsequently, the quantum chaotic honey badger method was applied for feature selection, which improved computational efficiency and model performance by choosing the most pertinent features and removing redundant ones. After similarity matching, we precisely identified the images using the Improved MobileNetV3 approach, which was optimized for high accuracy and efficiency. An accurate, effective, and high-performing CBIR system was ensured by this systematic process.

2.1 Preprocessing

Preprocessing is essential when it comes to improving the quality and appropriateness of images for CBIR systems. We used a systematic preprocessing strategy in this proposed methodology to maximize image quality and enable precise feature extraction.

Noise reduction: First, noise reduction was accomplished by using a median filter, which successfully eliminated undesired noise, such as pepper and salt noise, while maintaining crucial edges and structural elements in the images. The mean filtering substitutes the multi-neighbor grey values for a pixel's single grey value. After mean filtering and smoothing, the picture is g(x, y). Equation (1) 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 illustrates M pixels:

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

Normalization: Second, we applied the min-max normalization technique to all pictures to normalize the pixel values. Pixel values are scaled to a standard range, usually (0, 1), so that the intensity levels of the pictures are represented consistently. It is essential to perform this normalization step to extract features from several images that are consistent and comparable, irrespective of their initial intensity ranges. The Min-Max approach was chosen to standardize the raw image and improve model correctness. It is shown in equation (2),

$$x = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

Normalization is aligning and enclosing thermal pictures into a comprehensive anatomic template. Normalization is required because human activities vary, making it more difficult to compare one procedure to another and turn the results into a standard attribute.

Contrast enhancement: Finally, Adaptively Clipped Contrast Limited Adaptive Histogram Equalization (ACCLAHE) was used to apply contrast enhancement. This method preserves the natural appearance of pictures while increasing contrast providing higher-quality visuals. ACCLAHE adjusts to local variations in picture contrast to properly

enhance both low-contrast and high-contrast areas. The enhancement process not only improves the visual appeal of pictures but also makes finer features more visible, which is advantageous for extracting features and categorization tasks that follow in CBIR systems. The adaptive contrast enhancement technique is called CLAHE. It is predicated on AHE, in which a pixel's contextual region is used to compute the histogram. In this way, the pixel's intensity is converted to a value that falls within the display range in proportion to the rank of the pixel intensity in the local intensity histogram. Block size (N) and clip limit (CL) are the two critical variables. These user-determined heuristic variables are primarily utilized to control image quality. The proposed approach, Adaptively Clipped Contrast Limited Adaptive Histogram Equalization (ACCLAHE), automatically uses the provided input image to estimate the clip limit (CL) value.

The preprocessing stages of median filter noise reduction, min-max normalization, and ACCLAHE contrast enhancement improved picture quality and prepared the pictures for efficient feature extraction and categorization. The steps mentioned earlier are fundamental in ensuring the high precision and dependability of the CBIR system, which in turn improved its efficiency in obtaining pertinent images through content similarity extraction. Figure 2 shows the outcome of the preprocessing stage process of the proposed methods.

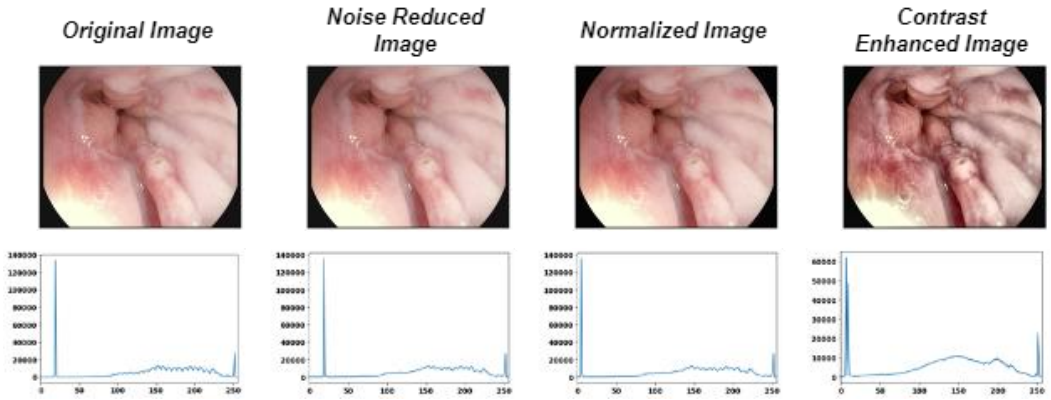

Figure 2. Outcomes of preprocessing using the proposed model

2.2 Feature extraction

Feature extraction is essential to CBIR because it converts unprocessed picture data into meaningful representations, making the retrieval of related images quickly and effectively easier. To extract features such as shape, color, and texture, a Modified ResNet152V2 model was employed. A deep convolutional neural network architecture called ResNet152V2 is known for its effectiveness and depth in vision recognition applications. ResNet152V2 was modified and optimized for CBIR, making the model better qualified to identify fine details in pictures, such as shape, texture, and color. The contours, edges, and structural components that comprise the item or scene shown in the image are identified by ResNet152V2's shape attributes extraction process. These attributes are essential to differentiate between various items or classes. Extracted texture attributes capture surface properties like roughness, smoothness, or patterns like fabrics or leaves by varying pixel intensities and patterns throughout the picture.

To extract the features from the query image, the Modified ResNet152v2 technique was used. The CNN design, Residual Network (ResNet), can have hundreds or even thousands of convolutional layers. The ResNet152v2's 152 convolutional layers can individually learn a level's properties. The extraction network uses pre-trained initial weights for input training. This method expedites the training and obtains high accuracy. Every model design begins with the original model and proceeds through several stages: reshape, flatten, dropout layer, first dense layer, second dense layer, and activation function for picture retrieval. The original ResNet's post-activation was replaced with a pre-activation to improve the feature extraction outcomes. Following feature extraction layers, an output layer with a softmax activation function, a fully linked layer with 64 filters, and a new average pooling layer were added. By including a dropout layer with 0.5 ratios, the network's over-fitting issue was avoided. The original ResNet152V2 model was pre-trained on the CBIR dataset to retrieve attributes in the present study. This indicates that the model continued employing appropriately verified feature extraction weights and hyper-parameters. By freezing the feature extraction layers, information loss was prevented during further training. The primary goal of freezing the weights of the pre-trained CNNs was to take advantage of their feature extraction capabilities.

As may be observed, equation (3) is the fundamental equation for a residual block.

$$y_1 = f(x_1, w_1) + h(x_1), x_2 = f(y_1) \tag{3}$$

The residual function is represented as f, and the ith residual unit is marked as $x_1, x_2, \dots, i$. The weight of a certain residual unit is indicated as $w_1, w_2, \dots, i$. Since $x_2 = y_1$, in equations (4)-(7)

$$x_2 = x_1 + f(x_1, w_1) \tag{4}$$

$$x_3 = x_2 + f(x_2, w_2) = x_1 + f(x_1, w_1) + f(x_2, w_2) \tag{5}$$

$$x_4 = x_3 + f(x_3, w_3) = x_1 + f(x_1, w_1) + f(x_2, w_2) + f(x_3, w_3) \tag{6}$$

$$x_i = x_1 + \sum k = 1^{i-1} f(x_k, w_k) \tag{7}$$

Equation (8) illustrates how equation (3) also affects the back-propagation.

$$\frac{\partial \phi}{\partial x_1} = \frac{\partial \phi}{\partial x_i} \cdot \frac{\partial x_i}{\partial x_1} = \frac{\partial \phi}{\partial x_i}\left(1 + \frac{\partial}{\partial x_i} \sum k = 1^{i-1} f(x_k, w_k)\right) \tag{8}$$

Equations (3) and (8) demonstrate that all units could move data in both forward and backward directions rapidly when the loss function was represented by the symbol ϕ. In contrast to AlexNet, DenseNet, ShuffleNet, and U-Net and among other deep learning models. The following lists show the benefits of the enhanced resNet152v2 model over other deep learning models. With 152 layers, ResNet152V2 is a highly complex neural network. It is appropriate for challenging image feature extraction tasks because of its depth, which enables it to catch complex structures and features in pictures. ResNet152V2's deep design allows it to extract intricate hierarchical features from images, resulting in outstanding picture classification accuracy. It works effectively on complex datasets and is frequently employed in applications where precision is essential. Residual connections are introduced in ResNet152V2 to aid with the vanishing gradient issue during

training. This makes optimizing extremely deep networks easy. ResNet152V2 pre-trained versions on huge datasets are easily accessible. Transfer learning, which starts a new task with a pre-trained model, may preserve computational power and training time while producing good results on the target position. ResNet152V2 has shown good standardization capabilities. Figure 3 demonstrates the outcome of feature extraction using the proposed methods.

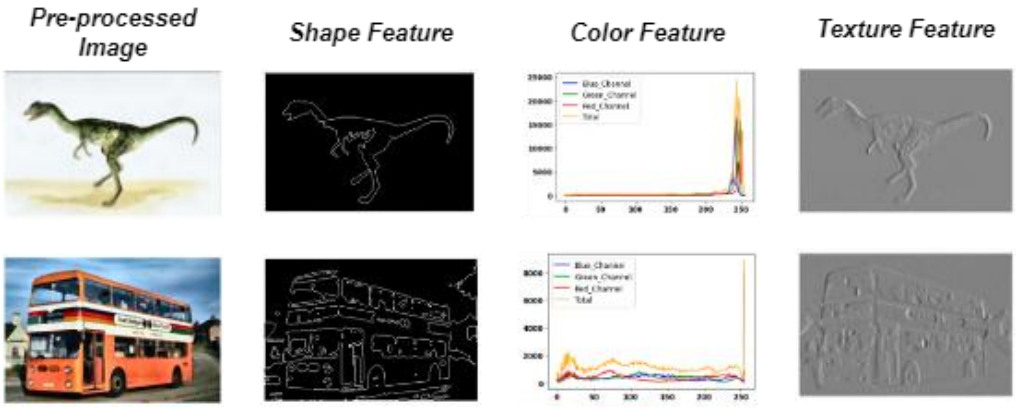

Figure 3. Outcomes of feature extraction using the proposed model

2.3 Feature selection

In content-based image retrieval (CBIR), feature selection is essential since it increases overall model performance and computational efficiency. To optimize processing and assure accurate retrieval outcomes, it is crucial to identify the most pertinent features and remove unnecessary or redundant ones in CBIR systems, where the amount of picture data can be substantial. The Quantum Chaotic Honey Badger Algorithm (QCHBA) provides a complex method to accomplish this. QCHBA uses chaotic optimization and quantum computing concepts to go through feature spaces quickly and effectively. In contrast to conventional techniques, which can have trouble with high-dimensional data or nonlinear feature correlations, QCHBA constantly modifies its search approach to find the most distinguishing attributes. In this way, it maximizes the feature set's relevance to the picture retrieval position. QCHBA assesses feature subsets in real-world applications according to how well they contribute to CBIR similarity metrics or classification. It prioritizes features that improve model performance and eliminates redundant or unnecessary characteristics in an iterative process of feature selection refinement. By concentrating resources on the most illuminating parts of the data, this selection approach not only increases the accuracy of picture classification but also lowers computational complexity.

2.3.1 Honey badger algorithm (HBA)

In this part, the features of the HBA are explained. The way honey badgers forage impacted the design of the HBA. The honey badger locates its food primarily via smell, although it also employs digging as a backup strategy. The honey badger uses honey-guide birds to find the hives and then enter. The honey badger's sensitivity to scent determines its movement; if the scent is strong, it will travel faster, and vice versa.

The following are the primary phases of the HBA and the associated equations:

Initialization process: The first possible solution is identified at this stage by utilizing the upper (HU) and lower (HL) borders of the issue space. Therefore, according to equation (9), the first answers are stochastic sets that can be produced by the following technique.

$$H_i = HL + r_1(1, D) \times (HU - HL), i = 1,2, \dots, N \tag{9}$$

Where N is the number of answer providers (honey badgers), H is the total amount of possible solutions, and D is the dimension of the solution.

Updating positions: At this moment, the candidates' coordinates are updated. This could entail, for example, using a method that employs the honey or digging phase.

Digging phase: During this stage, the capability of the predator's scent and the distinction between the prey and the honey badger affect the possible search subjects' movements. In a polarized circle, the honey badger excavates. Its motion is described by the following equation (10):

$$H_{new} = P + Fg \times \beta \times \ln \times \mathrm{P} + Fg \times r_3 \times (P - H_t) \times (\cos 2\pi r_4) \times (\cos 2\pi r_5) \tag{10}$$

Where the capacity of an insect to gather food is measured by β. The smallest possible value of β is 6. The random variables r_3, r_4, and r_5 were chosen from a uniform distribution with a range of 0 to 1. The intensity is $\ln$. The following equation (11) yields the Fg, an indication of the search direction:

$$Fg = \begin{cases} 1 & if\ r_6 \leq 0.5 \\ -1 & if\ else \end{cases} \tag{11}$$

Honey phase: Honey badgers utilize the honey phase to move about the honey lead bird when searching for beehives. The honey phase was discovered using the subsequent equation (12):

$$H_{new} = P + Fg \times r_7 \times \sigma \times (P - H_i) \tag{12}$$

Modeling intensity In: The honey badger's behavior is determined by its perception of insect scent; hence the following equation (13) represents each candidate's scent intensity In_i of the prey.

$$In_i = r_2 \times \frac{(H_i - H_{i+1})^2}{4\pi(P - H_i)} \tag{13}$$

where P is the position of the prey and $r2$ is a random amount in the interval (0, 1).

Modelling the density parameter ($sigma$): It is hypothesized that $beta$ is represented in each cycle, as illustrated below equation (14):

$$\sigma = C \times exp\left(\frac{-IT}{IT_{max}}\right) \tag{14}$$

Where IT and ITmax stand for the number of iterations that are now occurring and the total, respectively. It was proposed that the value of the constant C be 2.

Escaping from local solutions: To prevent getting bogged down in local answers, to indicate the direction of the search.

2.3.2 Two-dimensional Hénon map

A discrete-time dynamical model, the Hénon map is also called the Hénon-Pomeau attractor/map. Equation (15) represents the Hénon map's mathematical formula:

$$\begin{cases} x_{i+1} = 1 - 1.4 \cdot x_i^2 + y_i \\ y_{i+1} = 0.3 \cdot x_i \end{cases} \tag{15}$$

2.3.3 Quantum chaotic honey badger algorithm (QCHBA)

This method enhances the algorithm's essential evaluation by combining the HBA with the 2D Hénon map. Additionally, applying the quantum-based optimization method improved finding a balance between exploitation and exploration. The proposed QCHAB technique typically began by utilizing the training set comprising 80% of the provided image to identify the pertinent attributes. Afterwards, the quantum-based optimization method was used to produce the answer. The validation of every solution was then calculated, and the best solution was found after that. The following stage involved revising the solution in light of the advantages of QBO and the 2D chaotic maps. Reducing the testing set's attributes to reflect 20% of the dataset was the next step. Next, several performance measures were used to assess the effectiveness of the chosen features.

1) Initial solutions

This stage's primary goal was to use the QBO technique to create the population or set of N solutions. These solutions were expressed as follows: D Q-bits (where D is the number of attributes).

$$H_i = [q_{i1}|q_{i2}| \dots |q_{iD}] = [\theta_{i1}|\theta_{i2}| \dots |\theta_{iD}], i = 1,2, \dots, N \tag{16}$$

The symbol Hi in equation (16) represents the super-positions of the probability of the attributes that are chosen and correspond to ones and those that are not selected correspond to zeros.

2) Updating solution

Utilizing the following formula, the QCHAB obtained the binary form of answer Hi,=1,2,…,N at this stage.

$$BH_{i,j} = \begin{cases} 1 & if\ rand < |\beta|^2 \\ 0 & otherwise \end{cases} \tag{17}$$

The random number rand∈(0,1) was used in equation (17).

After that, the following formula was used to calculate the fitness value for each Hi.

$$Fit_i = \rho \times \gamma + (1 - \rho) \times \left(\frac{|BH_{i,j}|}{D}\right) \tag{18}$$

$|BHi,|$ in equation (18), represents the amount of attributes that match ones. The error categorization that results from using the KNN classifier is denoted by γ, while the variable

utilized to balance the aims (i.e., feature selection and error categorization) is denoted by $\rho \in (0,1)$.

The optimal solution for each Hi was then found by computing the fitness value for each Hi. Next, the existing solutions were updated using the modified HBA that was based on the 2D Hénon map. The two alterations in the controlled equations and the proposed changes are summarized below.

The basic HBA optimizer's performance was enhanced by modifying the parameters of C and β in equations (19)-(20), respectively, using the 2D Hénon map. The revised values of C and β were from the equations as follows:

$$C(t) = 4 * y_{i+1} \tag{19}$$

$$\beta(t) = 7 * H_{i+1} \tag{20}$$

The CHBA used the values 4 and 7 to offer wide variability. The initialization of the Hénon map is 0 (x(1) = 0; y(1) = 0) when it is implemented.

Next, the density variable and the digging phase were modelled as in equations (21)-(22),

$$H_{new} = P + Fg \times \beta(t) \times In \times P + Fg \times r_3 \times (P - H_i) \times (\cos 2\pi r_4) \times (1 - \cos 2\pi r_5) \tag{21}$$

$$\sigma = C(t) \times exp\left(\frac{-Iter}{Iter_{max}}\right) \tag{22}$$

Updates to the solutions were made continuously until the stop criteria were satisfied. After that, the output of this step was the best answer, Hb.

3) Evaluate quality of Hb

In this stage, the testing set's irrelevant features were eliminated using the optimal solution Hb. To forecast the testing set's target, the learnt KNN classifier was fed this minor testing set as input. Next, a set of performance criteria was used to calculate the expected target's performance. A notable development was incorporating feature selection via QCHBA in the CBIR systems. Ultimately, this provided more accurate and dependable picture retrieval results in various applications, from imaging to multimedia content management. It also used computational resources and improved the system's capacity to handle large-scale image collections. Figure 4 illustrates the outcome of feature selection using QCHBA in the CBIR system.

2.4 Classification

Classification is an essential component of content-based image retrieval (CBIR), which efficiently groups and classifies images according to their visual content. An example of the latest technique for attaining precise and effective classification in CBIR systems is the application of the Improved MobileNetV3 technology. Improved MobileNetV3 is ideal for real-time applications such as CBIR since it is specifically built to balance computing efficiency and accuracy. Advances in neural network design, including effective building blocks and optimized procedures like depth-wise separable convolutions, are included in its architecture. In cases involving significant image datasets, MobileNetV3's ability to

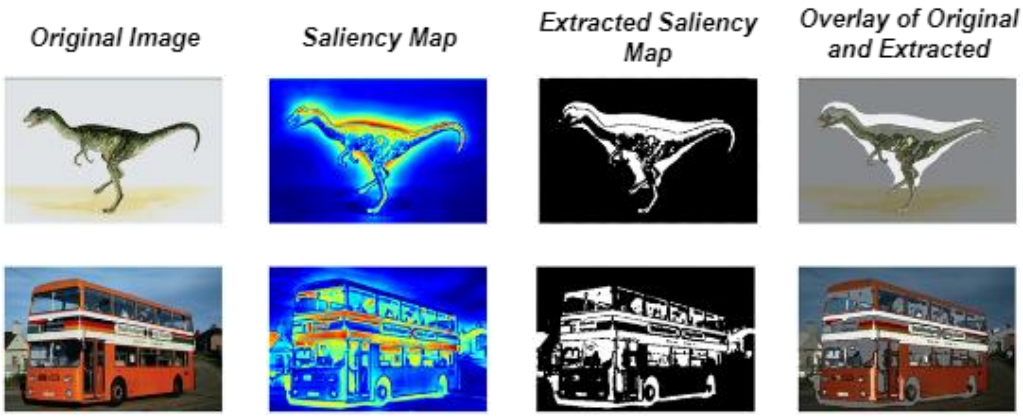

Figure 4. Outcomes of feature selection using the proposed model

retain excellent classification performance while minimizing computational resources is crucial. These attributes make this possible. Matching these attributes to a query picture is the next phase in the CBIR process, which comes after extracting complete feature representations of pictures (such as shape, texture, and colour) using Modified ResNet152V2. Following the computation of similarities, the obtained images are labelled or categorized according to the visual content classes using the Improved MobileNetV3 classifier. This categorization stage ensures that retrieved pictures are semantically categorized based on established classes or kinds, in addition to being visually relevant.

The NetAdapt technique is used by MobileNetV3 to determine the ideal amount of channels and convolution kernels. It changes the MobileNetV2 back-end output and introduces the SE channel attention structure. Relu6 is replaced by a new activation function called h-swish (x) in Mobilenetv3. In the SE module, it simulates sigmoid using Relu6 (x + 3)/6. There are three sections to the Mobilenetv3 network model structure: The initial segment has a single convolutional layer that utilizes a 3 × 3 convolution to extract information. The second section has several convolutional layers, with large and small versions (numbering 13 for small and 15 for large) owing to varying levels and parameters. The parameters and computation are reduced in the third section, and the Average Pooling is advanced. Two 1 × 1 convolutional layers are used in place of the complete connection, and the category is finally output.

2.4.1 Dilated convolution

The notion that down-sampling lowers image resolution and loss of information is the basis of the dilated Convolution, which attempts to address the issue in pictures. When the variable quantity stays the same, the Convolution receptive field rises. With dilated Convolution, the convolution kernel processes the information and adds a new variable, "dilation rate", to the convolutional layer. This variable determines the spacing of every value. By including holes, the receptive field is increased while the computation and variables are the same. There is no need to down-sample because it enables the original 3x3 convolution kernel to have a 5x5 or bigger relevant field. Stacking numerous dilated convolution kernels with different dilation rates results in multi-scale information from different relevant fields. The receptive field can be increased with dilated Convolution while maintaining the pixels' relative spatial positions and resolution.

2.4.2 Bias loss function

As the variance of the data points declines, the scale of the dynamic scaling cross-entropy loss known as bias loss also decreases. The Bias Loss function facilitates learning by concentrating on samples with distinctive properties. It lessens the issues that arise during optimization from random prediction. Definition of bias loss is expressed as equations (23)-(24):

$$L_{bias} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} z(v_i) y_{ij} \log f_j\,(X_i;\theta) \tag{23}$$

$$z(v_i) = exp(v_i * \alpha) - \beta \tag{24}$$

Allowing the feature space to be $X \in R^{C\times h\times w}$. h, w represents the input data's height and width. This is the label space, $Y = \{1,,,,k\}$. The amount of classes is k. $D = (x_i, y_i)_{i=1}^{N}$ is a dataset in a typical scenario. α and β are variables that can be adjusted in the formula. v represents the convolution layer output's scaling variance.

2.4.3 ECA module

An effective channel attention method is the ECA module. This module improves performance significantly and only requires a few variables. A sigmoid activation function, a 1*1 convolution, and an average pooling layer are all included in the ECA phase. The ECA phase, which is based on the SE module, avoids dimension reduction and successfully realizes cross-channel interaction by substituting one-dimensional Convolution for the Multi-Layer Perceptron (MLP) module. The ECA module uses 1D Convolution with a kernel size of k to realize information interaction between channels, i.e.

$$\omega = \sigma(C1D_k(y)) \tag{25}$$

In equation (25), 1 D denotes a 1 D convolution, where σ is a Sigmoid function. The size of the kernel is k. Without dimension reduction, y is the aggregated feature. Between k and C, typically, a power of two is used for channel dimension C.

$$C = \phi(k) = 2^{\gamma * k - b} \tag{26}$$

where b is a constant and C is the channel dimension in equation (26).

$$k = \psi(c) = \left|\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{27}$$

where the nearest odd amount of t is indicated by $|t|odd$ in equation (27). High-dimensional channels map to a larger range through nonlinear mapping, while low-dimensional channels map to a smaller range.

By introducing dilated Convolution, more information could be extracted by Convolution and the receptive field was expanded. The ECA module was added to the model in place of the SE module, significantly reducing parameter computation. Incorporated shallow attributes into deep layers, developed cross-layer connections across the mobile phase, and efficiently used local and in-depth features. Convolution was able to extract more information because dilated Convolution expanded the receptive field. Initially,

an image with dimensions of 448*448*3 was input. It employed a 3*3 dilated convolution to acquire attributes. The retrieved image's amount of channels rose to 16, and its dimensions were 16*224*224. The original picture size was reduced to half. Dilated Convolution decreased loss value and increased model correctness. The Improved Mobilenetv3 reduced the number of parameters in the model and increased accuracy by substituting the ECA module for the SE module in the mobile module. A sigmoid activation function, a 1*1 convolution, and an average pooling layer were all included in the ECA module. The enhanced mobile module was created using the ReLU/H_Swish activation function, Batch norm, and 1*1 Convolution. The activation function decreased the calculation quantity, while the Batch norm increased the pace at which the network convergences. The feature weight was determined via the ECA phase to create a weighted feature set, and the original attribute and attribute weight were multiplied. After going through the Batch norm, the input was sent to a 1*1 convolution to reduce the channel dimension, and the output was sent to the following mobile module.

We used cross-layer connections across the mobile phase to combine the attributes of various layers and minimize feature loss during the transfer process. Two cross-layer connections were created when the number of channels and the 24*112*112 mobile phase were output: ReLU, Batch norm, and 1*1 Convolution were used in the first cross-layer connection. By stacking fifteen mobile phases, a deep network was created, and additional attribute information was retrieved. After dimensionality reduction using 960*1*1 convolutions, a feature set with 960*7*7 dimensions was produced. The attributes retrieved from the Convolution were linearly merged after a fully linked layer with 1280 groups. Lastly, the results that were found were the output. Because Improved MobileNetV3's precision allowed for precise categorization, it improved user satisfaction and system usability, which in turn increased the overall effectiveness of CBIR systems. Furthermore, because of its effectiveness, pictures can be processed and retrieved quickly, which makes it suitable for applications like multimedia content management, security surveillance, and medical diagnosis that call for prompt replies.

The Improved MobileNetV3 likely incorporates architectural enhancements to optimize performance for the specific dataset and classification task. These improvements may include fine-tuning depth-wise separable convolutions, modifying activation functions such as Hard-Swish for better efficiency, adjusting the number of layers or channels for improved feature extraction, or integrating an attention mechanism to focus on critical image regions. Additionally, optimization techniques such as quantization-aware training or knowledge distillation may be applied to enhance accuracy while maintaining low computational cost. Providing these details would clarify how the proposed improvements contribute to superior classification performance.

2.4.4 Image matching and similarity measure

The similarity between the query picture and a dataset picture is determined by their distance from each other, which is represented by the same symbol $d(p,q)$ and is evaluated based on the color, texture, and form features that were extracted. As the two images get farther apart, there is less similarity between them. The distance between two images 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 equation (28) to gauge the degree of similarity between the feature vector of the input query picture and the dataset pictures.

$$d(p,q) = \sqrt{(p1-q1)^2 + (p2-q2)^2 + \cdots + (pn-qn)^2} = \sqrt{\sum_{i=1}^{n}(pi-qi)^2} \quad (28)$$

3. Results and Discussion

The datasets utilized, the results of several experiments, and the comparisons between the Improved MobileNetV3 and conventional image retrieval methods were explained. It should be mentioned that Python was used to do every experiment.

3.1 Experimental setup

The system was created under a range of circumstances. Table 1 shows the system's environment configuration.

Table 1. Environment setup of the proposed model

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

3.2 Dataset description

The Kvasir, CIFAR-10, and Corel image datasets were used. The initial instance of the color and texture attribute-based retrieval approach employed Kvasir's first version, which contained 4,000 pictures. The CIFAR-10 dataset included 60,000 pictures divided into 10 classes with 6000 images per class. The CIFAR-10 dataset can be accessed by everyone. The proposed approach was tested using a general-purpose WANG dataset, including 1000 Corel pictures with ten distinct subject groups. It is offered in JPEG format in two different sizes: 384 × 256 and 256 × 384. This dataset has 100 photos in each of the ten categories—Africa, Beaches, Buses, Elephants, Horses, Dinosaurs, Trees, Buildings, Food, and Mountains, among others. Other CBIR systems were also tested using the data collection process. Because of the dataset's robust size and availability of class information, it is frequently used. Additionally, each dataset was divided into training and test datasets. While only 20% of the images were in the test data, every training class has 80% of the total pictures in the training data set. Dataset image count is shown in Table 2.

Table 2. Dataset image 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

3.3 Evaluation metrics

In this work, we used two well-known metrics for performance evaluation: recall and precision. These measures have been used, and continue to be used, to evaluate different documents and CBIR methods. The general formula for recall and precision is given by the following equations (29)-(30):

$$Precision = \frac{TP}{TP+FP} \tag{29}$$

$$Recall = \frac{TP}{TP+FN} \tag{30}$$

True positives (TP) are the pictures the CBIR algorithm adequately identified in the query image class. On the other hand, pictures returned by the model that do not correspond to the query picture class were known as false positives (FP). Images that were part of the query image class but not returned by the system were called false negatives (FN). We used both recollection and accuracy to assess the validation of the proposed model because recall or precision by themselves is insufficient to determine the efficacy of a CBIR model.

3.4 Evaluation of the Kvasir dataset

The retrieval outcomes of query images on the Kvasir dataset are shown in Figure 5. In this Figure, the red border samples represent wrong retrievals, and the green border images are the correct retrievals for the query image.

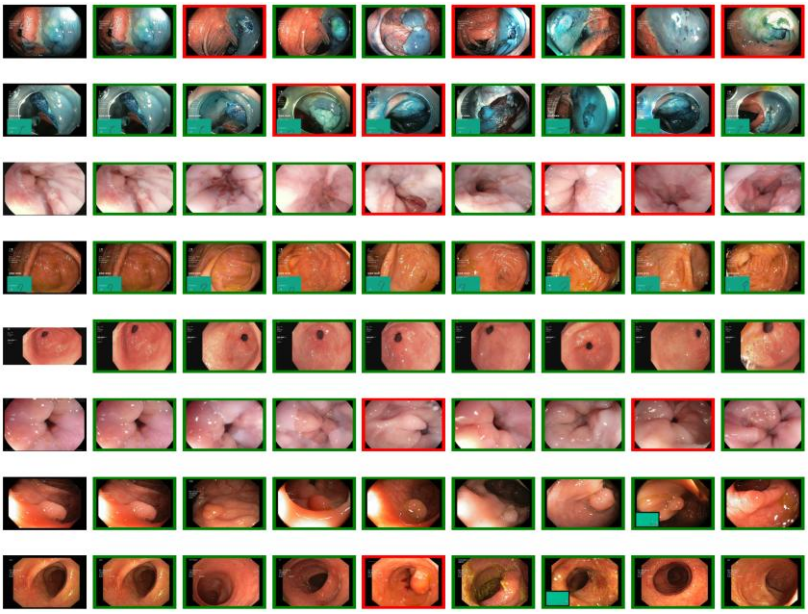

Figure 5. The representative retrieval results from the Kvasir dataset for a query picture are shown. Red marks denote images from the different class, while a green bounding box appears for images returned from the same class as the query picture.

With an emphasis on average precision and recall metrics, Table 3 compares many methods for categorizing different classes in the Kvasir dataset. The Canberra, Bray, RFRM, and the proposed approach are compared. The Table displays each method's average recall and precision values for each class. The proposed strategy outperformed the other techniques significantly for the "DLP" class, with the highest average precision (1.00) and average recall (0.992). In the same way, the proposed method demonstrated near-perfect precision (0.989) and perfect recall (1.00) in the "DRM" class. Again, the proposed method performed best in the case of "Esophagitis," with precision (0.993) and recall (0.997).

With accuracy scores of 0.987 and 0.986 and perfect recall scores (1.00 and 0.989, respectively), the "Normal Caceum" and "Normal Pylorus" classes, the proposed technique performed excellently. Using the proposed method, recall and precision for the "Normal Z-Line" class both reached 1.00. In "Polyps," the proposed method received a precision score of 1.00 and a recall score of 0.991. Lastly, the proposed method produced 0.989 precision and 1.00 recall for "Ulcerative Colitis."

Table 3. Comparing average precision and average recall value with different approaches utilizing the Kvasir dataset

Class	Average Precision				Average Recall			
	RFRM	**Bray**	**Canberra**	**Proposed**	**RFRM**	**Bray**	**Canberra**	**Proposed**
DLP	0.90	0.54	0.48	**1.00**	0.191	0.138	0.137	0.992
DRM	0.90	0.48	0.49	0.989	0.190	0.142	0.142	**1.00**
Esophagitis	0.85	0.47	0.49	0.993	0.180	0.138	0.139	0.997
Normal Caceum	0.85	0.68	0.63	0.987	0.191	0.142	0.142	**1.00**
Normal Pylorus	0.85	0.76	0.76	0.986	0.200	0.138	0.141	0.989
Normal Z-Line	0.90	0.57	0.56	**1.00**	0.180	0.140	0.138	**1.00**
Polyps	0.85	0.53	0.57	**1.00**	0.200	0.140	0.141	0.991
Ulcerative Colitis	0.80	0.51	0.51	0.989	0.190	0.140	0.140	**1.00**
Average	0.862	0.567	0.561	0.993	0.191	0.138	0.140	0.9961

The proposed method outperformed the others on average in every class, showing significantly higher average recall and precision values (0.993 and 0.9961, respectively). The Bray and Canberra approaches displayed comparable and noticeably lower performance metrics, with Bray averaging 0.567 in precision and 0.138 in recall and Canberra averaging 0.561 in precision and 0.140 in recall. Figures 6 and 7 show the performance comparison of the proposed method with different methods on the Kvasir dataset.

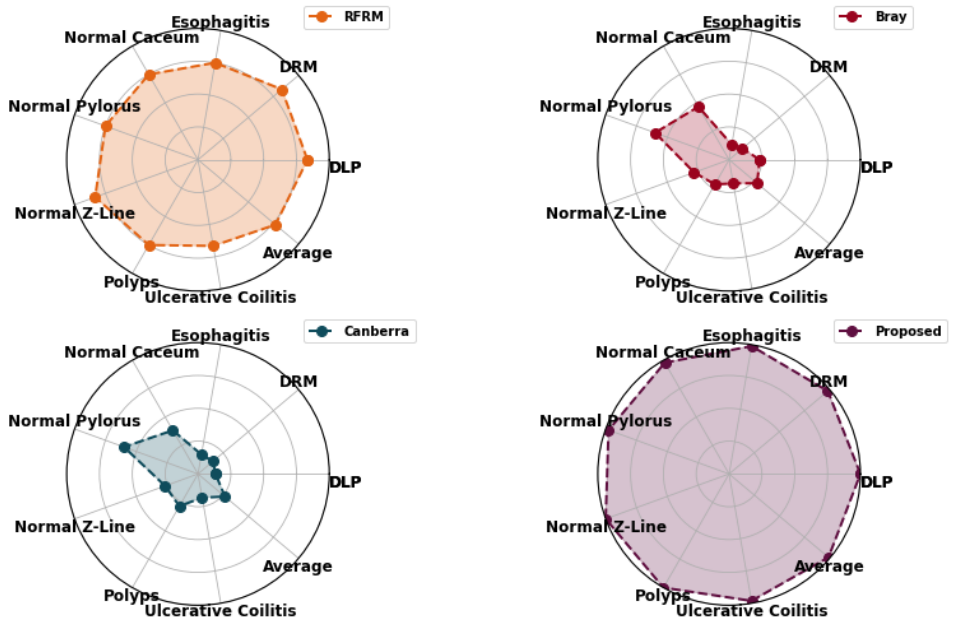

Figure 6. Comparison of the performance of the average precision of the proposed model with different methods on the Kvasir dataset

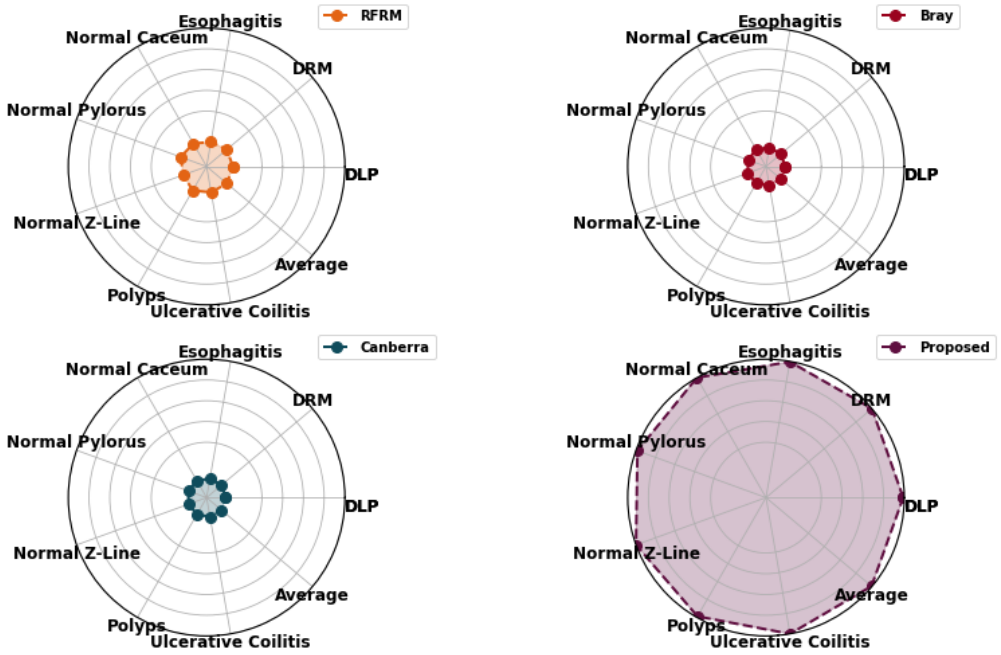

Figure 7. Comparison of the performance of the average recall of the proposed model with different methods on the Kvasir dataset

3.5 Evaluation of the CIFAR-10 dataset

The retrieval outcomes of query images on the CIFAR-10 dataset are shown in Figure 8. In this Figure, the red border samples are wrong retrievals, and the green border images are correct retrievals for the query image.

Table 4 compares various methods used on the CIFAR-10 dataset, emphasizing the average precision and average recall measures. The approaches GA, MCM, EDPS, and the proposed method are examined. The Table shows the average recall and precision values for every technique for each class. The proposed strategy significantly outperformed the previous methods, achieving the highest average precision (0.986) and perfect recall (1.00) for the "Automobile" class. The proposed method similarly led with precision and almost perfect recall (1.00 and 0.986, respectively) in the "Airplane" class. The proposed technique performed exceptionally well in the "Cat" class, with a precision of 0.984 and a recall of 0.989. Once again, the proposed approach achieved near-perfect recall (1.00) and precision (0.992) in the "Dog" class. Additionally, for the "Horse" class, the proposed method received flawless recall and precision ratings (1.00 and 0.994, respectively). The proposed method was implemented with 1.00 precision and 0.991 recall for the "Truck" class.

Figure 8. The representative retrieval results from the CIFAR-10 dataset for a query picture are shown. Red marks denote images from the different class, while a green bounding box appears for images returned from the same class as the query picture.

Table 4. Comparing average precision and average recall value with different approaches utilizing the CIFAR-10 dataset

Class	Average Precision				Average Recall			
	GA	MCM	EDPS	Proposed	GA	MCM	EDPS	Proposed
Automobile	0.705	0.432	0.395	0.986	0.235	0.138	0.127	**1.00**
Airplane	0.762	0.521	0.568	**1.00**	0.178	0.129	0.146	0.986
Cat	0.921	0.861	0.826	0.984	0.137	0.115	0.118	0.989
Dog	0.854	0.435	0.482	0.992	0.169	0.151	0.136	**1.00**
Horse	0.893	0.692	0.736	**1.00**	0.153	0.128	0.125	0.994
Truck	0.826	0.642	0.689	**1.00**	0.162	0.141	0.147	0.991
Bird	0.816	0.365	0.464	0.989	0.210	0.153	0.158	**1.00**
Deer	0.918	0.821	0.782	**1.00**	0.115	0.084	0.112	0.984
Frog	0.768	0.692	0.719	0.983	0.134	0.107	0.121	0.983
Ship	0.754	0.425	0.568	0.987	0.258	0.162	0.182	**1.00**
Average	0.821	0.589	0.622	0.9921	0.175	0.130	0.137	0.9927

The proposed method performed exceptionally well in the "Bird" class, with a precision of 0.989 and excellent recall (1.00). Using the proposed approach, the "Deer" class demonstrated excellent recall and precision (1.00 and 0.984, respectively). The proposed method maintained significant recall (0.983) and precision (0.983) for the "Frog" class. Ultimately, the proposed method produced excellent recall (1.0) and high precision (0.987) in the "Ship" class. With an average precision of 0.9921 and an average recall of 0.9927, the proposed approach outperformed the other approaches significantly on average across all classes.

The GA approach outperformed with an average recall of 0.175 and precision of 0.821, while the MCM and EDPS methods performed worse, averaging 0.589 in precision and 0.130 in recall and 0.622 in precision and 0.137 in recall, respectively. Figures 9 and 10 show the performance comparison of the proposed methods with different methods on the CIFAR-10 dataset.

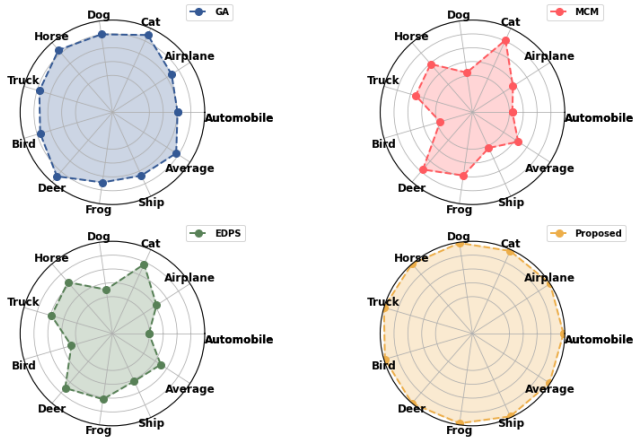

Figure 9. Comparison of the performance of the average precision of the proposed model with different methods on the CIFAR-10 dataset

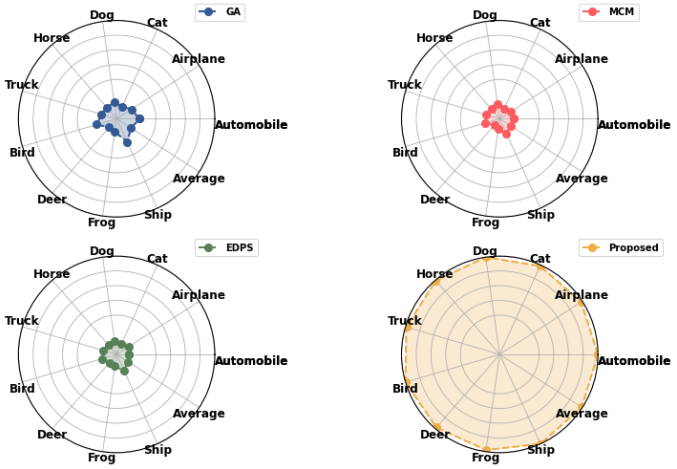

Figure 10. Comparison of the performance of the average recall of the proposed model with different methods on the CIFAR-10 dataset

3.6 Evaluation of Corel-1k dataset

The retrieval outcomes of query images on the Corel-1k dataset are shown in Figure 11. In this Figure, the red border samples indicated wrong retrievals, and the green border images show correct retrievals for the query image. Table 5 shows the analysis of several methods—SVM, K-means, BiCBIR, and a proposed method assessed using the Corel dataset. The average recall and precision metrics for the various classes are the main focus of the comparison. The proposed method significantly outperformed the other approaches, achieving the maximum recall (0.994) and precision (1.00) for the "Beach" class. BiCBIR received a perfect precision score of 1.00 in the "Buses" class, whereas the proposed technique received a close score of 0.992 and an ideal recall score of 1.00. The proposed method was performed with perfect precision (1.00) and great recall (0.985) in the "Elephants" class. The proposed method received excellent recall and precision scores (1.00) in the "Horse" and "Flower" groups. The proposed method had the best precision (0.986) and perfect recall (1.00) for the "Food" class. With the proposed approach, the "Africa" and "Buildings" classes likewise showed perfect recall (1.00), along with excellent precision scores of 1.00 and 0.987, respectively.

SVM, K-means, and BiCBIR all performed well in the "Dinosaurs" class, but the proposed method had almost perfect precision (0.982) and recall (0.986). With the proposed approach, the "Mountain" class showed strong recall (0.984) and precision (0.984). The proposed method outperformed the others in terms of overall performance, with an average recall of 0.993 and a precision of 0.9931.

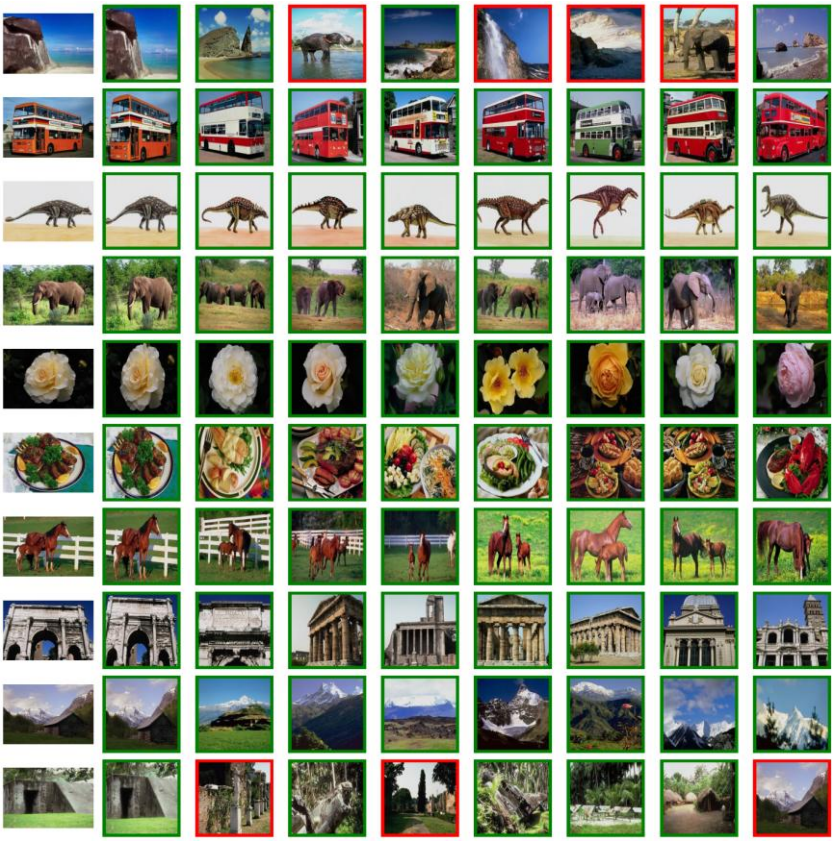

Figure 11. The representative retrieval results from the Corel-1k dataset for a query picture are shown. Red marks denote images from the different class, while a green bounding box appears for images returned from the same class as the query picture.

Table 5. Comparing average precision and average recall value with approaches utilizing the Corel-1k dataset

Class	Average Precision				Average Recall			
	SVM	K-means	BiCBIR	Proposed	SVM	K-means	BiCBIR	Proposed
Beach	0.9656	0.76	0.750	**1.00**	0.87	0.152	0.150	0.994
Buses	0.833	0.79	1.000	0.992	0.80	0.158	0.200	**1.00**
Elephants	0.775	0.7	0.900	**1.00**	0.83	0.14	0.180	0.985
Horse	0.863	0.7	1.000	**1.00**	0.82	0.14	0.200	0.981
Food	0.861	0.56	0.950	0.986	0.87	0.112	0.190	**1.00**
Africa	0.812	0.72	0.950	**1.00**	0.82	0.144	0.190	**1.00**
Buildings	0.782	0.55	0.850	0.987	0.90	0.11	0.170	**1.00**
Dinosaurs	0.822	1	1.000	0.982	0.79	0.2	0.200	0.986
Flower	0.863	0.87	1.000	**1.00**	0.82	0.174	0.200	**1.00**
Mountain	0.902	0.58	0.800	0.984	0.93	0.116	0.160	0.984
Average	0.847	0.723	0.920	0.9931	0.845	0.1446	0.184	0.993

K-means and BiCBIR performed worse, with average recall of 0.1446 and 0.184 and precision of 0.723 and 0.920, respectively. Figures 12 and 13 show the performance comparison of the proposed with different methods on the Corel-1k dataset. The comparison between supervised (SVM) and unsupervised (K-means) learning methods in Figure 12 highlights the effectiveness of different learning paradigms in image classification. SVM, a supervised approach, leverages labeled data to create decision boundaries, leading to more precise classifications, whereas K-means, an unsupervised clustering method, groups similar images without prior labels, making it less accurate but useful for exploratory analysis. By including both methods, the author demonstrates the advantage of supervised learning for achieving higher accuracy while also showcasing the limitations of unsupervised techniques, reinforcing the effectiveness of the proposed model.

3.7 Overall comparison of proposed model with existing models and discussion

Table 6 presents a comprehensive comparison of the proposed model with existing methods across three different datasets: Kvasir, CIFAR-10, and Corel-1K. The evaluation is based on two key performance metrics: Average Precision and Average Recall. These metrics are crucial in assessing the accuracy and effectiveness of classification models, as precision measures how many of the selected items are relevant, and recall measures how many relevant items are selected. The following is a detailed explanation of the results from each method and dataset. Rani et al. (2024) employed a Multi-class SVM method on the TCIA-CT dataset, achieving a high average precision of 0.99 and a moderate recall of 0.83. The high precision indicates that the model classifies relevant instances with high accuracy. However, the recall value of 0.83 suggests that while the model was good at identifying relevant instances, it missed a significant portion of them, thus limiting its ability to identify all true positives.

Mahalle et al. (2023) used a VCCINN method for the Caltech-101 dataset, achieving very high recall (0.99) and good precision (0.985). This demonstrates the model's ability to identify almost all relevant instances (high recall), while maintaining a strong classification performance (precision). This indicates that VCCINN is particularly effective in terms of identifying relevant instances, although it still maintains a solid precision score. Khan et al. (2021) used a Genetic Algorithm (GA) combined with SVM on both the CIFAR-10 and Kvasir datasets. For CIFAR-10, the model achieved 0.916 precision and low recall of 0.183, while for Kvasir, the precision was slightly lower at 0.913, with recall also low at 0.218. While the precision was relatively good, the low recall indicates that the method was not very effective in capturing all relevant instances. This suggests that the model may struggle to generalize to unseen or complex patterns, leading to lower recall.

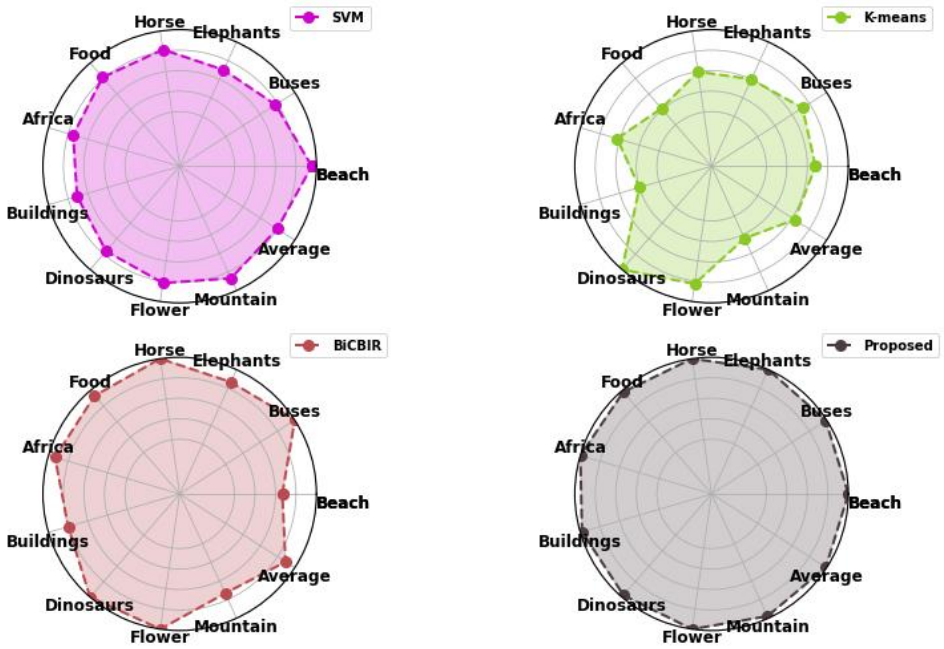

Figure 12. Comparison of the performance of the average precision of the proposed model with different methods on the Corel-1k dataset

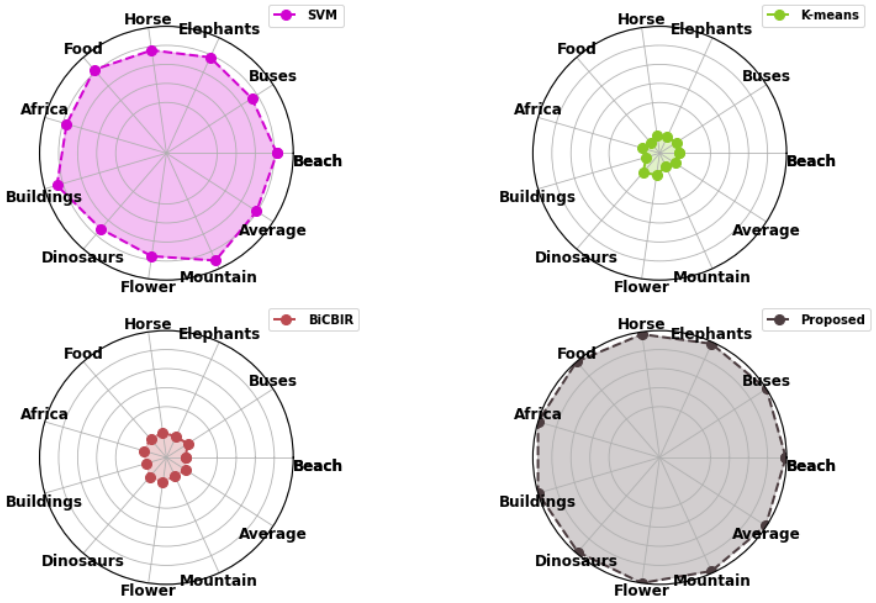

Figure 13. Comparison of the performance of the average recall of the proposed model with different methods on the Corel-1k dataset

Table 6. Overall comparison of proposed and existing methods in literature

Reference	Method	Dataset	Average Precision	Average Recall
(Rani et al., 2024)	Multi-class SVM	TCIA-CT	0.99	0.83
(Mahalle et al., 2023)	VCCINN	Caltech-101	0.985	0.99
(Khan et al., 2021)	GA + SVM	CIFAR-10	0.916	0.183
		Kvasir	0.913	0.218
(Kelishadrokhi et al., 2023)	ELNDP	Corel 1K	0.8271	0.5397
		Colored Brodatz	0.9475	0.9771
(Salih et al., 2023)	LBP + DWT	Corel 1K	0.867	-
Proposed Model	Improved MobileNetV3	Kvasir	0.993	0.9961
		CIFAR-10	0.9921	0.9927
		Corel-1k	0.9931	0.993

Kelishadrokhi et al. (2023) utilized ELNDP for the Corel-1K and Colored Brodatz datasets. On the Corel-1K dataset, they achieved precision of 0.8271 and recall of 0.5397, which indicates moderate performance. The precision was good, but the lower recall suggested that the method missed a substantial number of relevant instances. However, for the Colored Brodatz dataset, the model achieved excellent results with precision of 0.9475 and recall of 0.9771, indicating its capability to both classify correctly and identify most relevant instances effectively. Salih & Abdulla (2023) employed LBP + DWT on the Corel-1K dataset and achieved an average precision of 0.867. However, the recall value was not provided, so it is unclear how well the model identified relevant instances. Despite the lack of recall data, the precision value suggests that the model did a decent job at identifying correct instances, though its performance in terms of recall remains uncertain. The proposed model uses an Improved MobileNetV3 for image classification on the Kvasir, CIFAR-10, and Corel-1K datasets. The results are exceptional across all three datasets:

1) Kvasir Dataset: The model achieved 0.993 precision and 0.9961 recall, indicating nearly perfect performance in both identifying relevant instances and classifying them correctly.

2) CIFAR-10 Dataset: The model performed similarly well, with 0.9921 precision and 0.9927 recall, showcasing its robustness across diverse datasets.

3) Corel-1K Dataset: The model achieved 0.9931 precision and 0.993 recall, again demonstrating excellent classification accuracy and the ability to capture most relevant instances.

The proposed model outperformed existing methods in terms of both precision and recall, indicating its ability to deliver superior classification results with minimal missed instances (high recall) and excellent classification accuracy (high precision). The consistent performance across multiple datasets underscored the robustness and efficiency of the Improved MobileNetV3 architecture. The proposed model consistently outperformed the existing methods in both precision and recall across the Kvasir, CIFAR-10, and Corel-1K datasets. This highlights the superior ability of the Improved MobileNetV3 model to not only

classify instances accurately but also identify relevant instances effectively, making it a more reliable and efficient choice for image classification tasks compared to other methods. The significant improvements in performance across multiple datasets demonstrate the proposed model's robustness and its potential for real-world applications.

3.8 Evaluation performance of training and testing

In CBIR, evaluating training and testing performance entails a thorough procedure to determine how well retrieval algorithms retrieve relevant pictures according to their content properties. Algorithms learn feature representations from a labelled dataset to identify distinguishable features like color, texture, and shape. Metrics like precision and recall are then used during testing to assess how well these algorithms work. Recall evaluates the percentage of relevant pictures successfully retrieved, whereas precision measures the percentage of relevant images retrieved. These evaluations are essential for fine-tuning and optimizing algorithms to improve image retrieval systems' accuracy and effectiveness across various applications and datasets. Figures 14-16 demonstrate the evaluation performance of training and testing on each dataset.

3.9 Computational time complexity

The computational time complexity in CBIR is a crucial component that affects the scalability and efficiency of the system. The amount of processing time required by an algorithm to process data and produce results is referred to as time complexity. This usually entails procedures for feature extraction, similarity matching, and picture retrieval in the context of CBIR. In order to guarantee fast reaction times, efficient algorithms are necessary, mainly when working with big datasets and real-time applications. Reduced computing time complexity makes the system more useful for multimedia databases, security systems, and medical imaging applications by enabling faster retrieval results.

Table 7 shows how different methodologies' computational time complexity compares, highlighting how effective the proposed Improved MobileNetV3 model is. Using the Multi-class SVM approach, Rani et al. (2024) obtained a computational time complexity of 0.45 ms. Using their VCCINN approach, Mahalle et al. (2023) obtained a somewhat lower complexity of 0.37 ms. Using the GA + SVM technique, Khan et al. (2021) reported a time complexity of 0.41 ms. An even more effective ELNDP approach with a temporal complexity of 0.29 ms was reported by Kelishadrokhi et al. (2023). Salih et al. (2023) reported that their LBP + DWT approach had a calculation time of 0.35 ms. The proposed Improved MobileNetV3 model outperforms all previous techniques and shows the lowest computational time complexity of 0.21 ms. The proposed algorithm is particularly suitable for large-scale and real-time image retrieval applications due to its huge reduction in processing time, which shows its efficiency and superiority. Figure 17 illustrates the comparison of the proposed running time with the existing method.

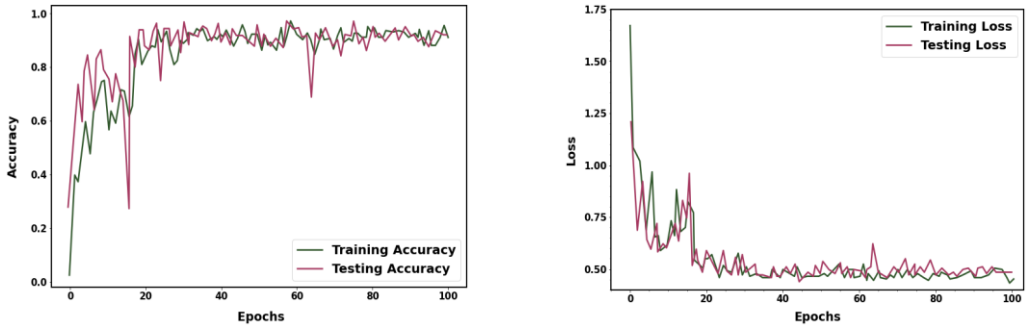

Figure 14. Performance of training and testing on the Kavsir dataset

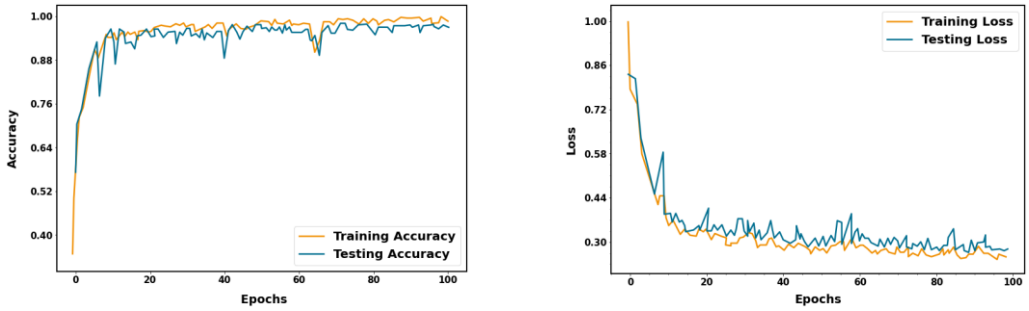

Figure 15. Performance of training and testing on the CIFAR-10 dataset

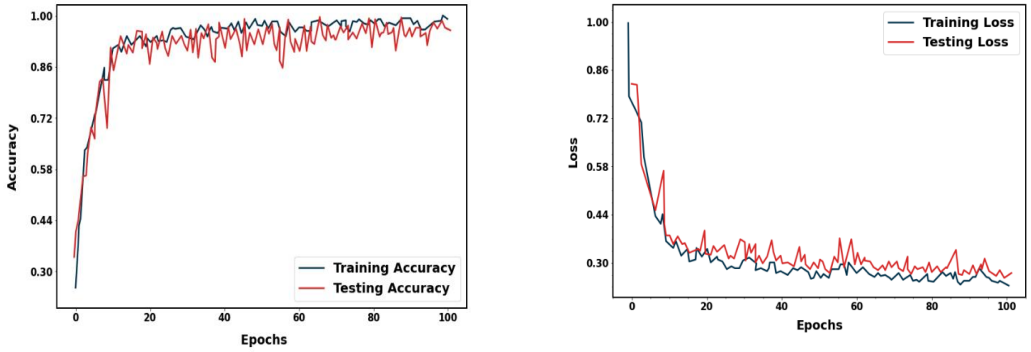

Figure 16. Performance of training and testing on Corel-1k dataset

Table 7. Comparison of computational time complexity of proposed and existing methods

Reference	Method	Computational Time Complexity (ms)
(Rani et al., 2024)	Multi-class SVM	0.45
(Mahalle et al., 2023)	VCCINN	0.37
(Khan et al., 2021)	GA + SVM	0.41
(Kelishadrokhi et al., 2023)	ELNDP	0.29
(Salih et al., 2023)	LBP + DWT	0.35
Proposed Model	Improved MobileNetV3	0.21

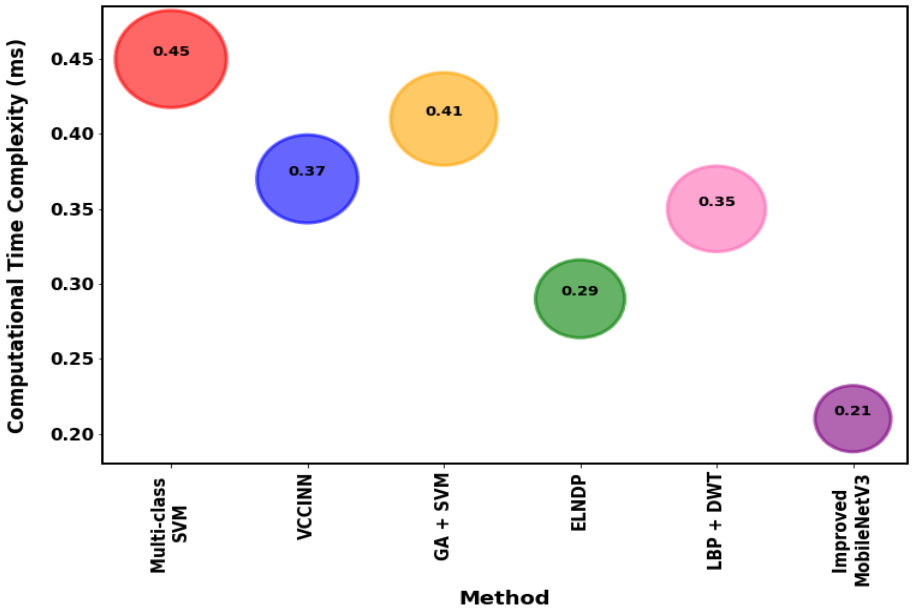

Figure 17. Comparison of computational time complexity of proposed and existing methods

3.10 Limitations of proposed model and future work

There are several limitations to this study. First, the average precision in the external verification dataset is lower, indicating that the system's generalization capacity still needs work. Second, the comparatively limited information in the picture set affects how well it can be retrieved and predicted. The aim of future research is to have data continuously added to the database, creating a situation that is similar to how radiologists learn new cases. Future improvements to this model are recommended, including extracting high-level features from images and reducing the semantic gap in image retrieval through saliency-based image techniques.

4. Conclusions

This work presented an efficient CBIR system that employs Improved MobileNetV3 to retrieve images from databases. The proposed CBIR system extracts feature from an image when the user submits a query image. The proposed research successfully implements a comprehensive pre-processing and classification pipeline for image retrieval. We achieved effective noise reduction using a median filter, normalization through the min-max method, and contrast enhancement using ACCLAHE. Feature extraction with a Modified ResNet152V2 model allowed us to capture detailed and discriminative features. Utilizing the Quantum Chaotic Honey Badger Algorithm efficiently selected relevant features and removed redundancies, enhancing the model's performance. Finally, the Improved MobileNetV3 technique demonstrated high accuracy and efficiency in classifying the images post similarity matching, validating the robustness of the proposed approach.

The proposed CBIR method outperformed the existing approaches and demonstrated excellent retrieval picture outcomes in terms of recall and precision rates. Additionally, the proposed approach had the highest overall performance outcomes regarding recall and precision rates compared to existing state-of-the-art techniques.

5. Acknowledgements

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

6. Author's Contributions

Dr. G. Sai Chaitanya Kumar- Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Writing – Original Draf. Dr. V. Srilakshmi- Software, Formal Analysis, Investigation, Resources, Writing – Original Draft, Writing - Review & Editing, Visualization. Dr. G. N. Beena Bethel- Conceptualization, Writing - Review & Editing, Original Draft, Writing - Review & Editing, Visualization. Dr. Narendhar Mulugu- Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources. Dr. M V Kamal- Writing - Review & Editing, Original Draft, Writing - Review & Editing, Visualization.

ORCID

G. Sai Chaitanya Kumar https://orcid.org/0000-0002-2127-2086

7. 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.

References

Chen, Y., Ling, M., Liu, Y., Chen, X., Li, Y., & Tong, B. (2024). Enhancing MRI image retrieval using autoencoder-based deep learning: A solution for efficient clinical and teaching applications. *Journal of Radiation Research and Applied Sciences*, 17(3), Article 100932. https://doi.org/10.1016/j.jrras.2024.100932

Fadaei, S., Dehghani, A., & Ravaei, B. (2024). Content-based image retrieval using multi-scale averaging local binary patterns. *Digital Signal Processing*, 146, Article 104391. https://doi.org/10.1016/j.dsp.2024.104391

Hong, S. A., Huu, Q. N., Viet, D. C., Thuy, Q. D. T., & Quoc, T. N. (2023). Improving image retrieval effectiveness via sparse discriminant analysis. *Multimedia Tools and Applications*, 82(20), 30807-30830. https://doi.org/10.1007/s11042-023-14748-9

Kelishadrokhi, M. K., Ghattaei, M., & Fekri-Ershad, S. (2023). Innovative local texture descriptor in joint of human-based color features for content-based image retrieval. *Signal, Image and Video Processing*, 17(8), 4009-4017. https://doi.org/10.1007/s11760-023-02631-x

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

Khunsongkiet, P., Bootkrajang, J., & Techawut, C. (2024). Low-level feature image retrieval using representative images from minimum spanning tree clustering. *Multimedia Tools and Applications*, 83(2), 3335-3356. https://doi.org/10.1007/s11042-023-15605-5

Mahalle, V. S., Kandoi, N. M., & Patil, S. B. (2023). A powerful method for interactive content-based image retrieval by variable compressed convolutional info neural networks. *The Visual Computer*, 40(8), 5259-5285. https://doi.org/10.1007/s00371-023-03104-5

Rani, K. V., Prince, M. E., Therese, P. S., Shermila, P. J., & Devi, E. A. (2024). Content-based medical image retrieval using fractional Hartley transform with hybrid features. *Multimedia Tools and Applications*, 83(9), 27217-27242. https://doi.org/10.1007/s11042-023-16462-y

Ranjith, E., Parthiban, L., Latchoumi, T. P., Kumar, S. A., Perera, D. G., & Ramaswamy, S. (2024). An effective content-based image retrieval system using deep learning-based inception model. *Wireless Personal Communications*, 133, 811-829. https://doi.org/10.1007/s11277-023-10792-8

Rashad, M., Afifi, I., & Abdelfatah, M. (2023). RbQE: An efficient method for content-based medical image retrieval based on query expansion. *Journal of Digital Imaging*, 36(3), 1248-1261. https://doi.org/10.1007/s10278-022-00769-7

Salih, F. A. A., & Abdulla, A. A. (2023). Two-layer content-based image retrieval technique for improving effectiveness. *Multimedia Tools and Applications*, 82(20), 31423-31444. https://doi.org/10.1007/s11042-023-14678-6

Shetty, R., Bhat, V. S., & Pujari, J. (2024). Content-based medical image retrieval using deep learning-based features and hybrid meta-heuristic optimization. *Biomedical Signal Processing and Control*, 92, Article 106069. https://doi.org/10.1016/j.bspc.2024.106069

Vu, V. H. (2024). Content-based image retrieval with fuzzy clustering for feature vector normalization. *Multimedia Tools and Applications*, 83(2), 4309-4329. https://doi.org/10.1007/s11042-023-15215-1

Wang, S., Xia, Y., Xiang, N., Qian, K., Yang, X., You, L., & Zhang, J. (2024). Multi-colour sketch-based image retrieval with an explicable feature embedding. *Engineering Applications of Artificial Intelligence*, 135, Article 108757. https://doi.org/10.1016/j.engappai.2024.108757

Wang, Y., Chen, L., Wu, G., Yu, K., & Lu, T. (2023). Efficient and secure content-based image retrieval with deep neural networks in mobile cloud computing. *Computers & Security*, 128, Article 103163. https://doi.org/10.1016/j.cose.2023.103163

Zhang, H., Cheng, D., Kou, Q., Asad, M., & Jiang, H. (2024). Indicative Vision Transformer for end-to-end zero-shot sketch-based image retrieval. *Advanced Engineering Informatics,* 60, Article 102398. https://doi.org/10.1016/j.aei.2024.102398

Zhang, X., Bai, C., & Kpalma, K. (2023). OMCBIR: Offline mobile content-based image retrieval with lightweight CNN optimization. *Displays*, 76, Article 102355. https://doi.org/10.1016/j.displa.2022.102355