A Hybrid Deep Feature Extraction Framework with Quantum Grey Wolf Optimization for Enhanced Content-Based Image Retrieval

Sarva N. Kumar [1,*] and Ch S. Kumar [2]

[1] Department of Electronics and Communication Engineering, MLR Institute of Technology, Hyderabad, India
[2] Department of Electrical Electronics and Communication Engineering, School of Core Engineering, GITAM (Deemed to be University), Hyderabad, India
Email: navin9ap@gmail.com (S.N.K.); schennup@gitam.edu (C.S.K.)
*Corresponding author

***Abstract*—A key tool for organizing and extracting visual information from large picture databases is Content-Based Image Retrieval (CBIR). To make CBIR systems much more accurate and efficient, this research proposes a new combined approach that uses deep feature extraction along with Quantum Grey Wolf Optimization (QGWO). The suggested system captures complex visual patterns across a variety of images categories by utilizing pre-trained Convolutional Neural Networks (CNNs) for reliable and advanced feature extraction. This paper presents the retrieval of single objects and multi-objects. The deep learning techniques utilized in this work are Inception V1, Inception V2, and ResNet50, chosen as they represent progressively advanced CNN architectures with increasing depth and feature extraction capability. To make the process faster and more effective, these features are improved using QGWO, a method that combines ideas from quantum computing with the social behaviour and hunting strategies of grey wolves. Our combined method performs better than existing CBIR algorithms in precision, recall, accuracy, and F1-Scores, based on thorough testing with the Corel-1K, Corel-5K, Corel-10K image datasets. The fusion of deep learning with a quantum-inspired optimization approach resulted in a retrieval accuracy of 99.20%, and percentage of wrongly retrieved images obtained is 1.08% using Corel-1k. The accuracy achieved using Corel-5k is 99.12% and Corel-10K is 99.04%. The results prove the computational efficiency of contemporary CBIR system.**

***Keywords*—Content-Based Image Retrieval (CBIR), Convolutional Neural Networks (CNNs), Inception V1, Inception V2, ResNet50, Quantum Grey Wolf Optimization (QGWO)**

I. INTRODUCTION

The explosive proliferation of digital visual content across various domains—ranging from social media and surveillance to medical imaging and satellite data has necessitated the development of intelligent systems for fast and accurate image retrieval. Content-Based Image Retrieval (CBIR), which focuses on retrieving images based on their visual content rather than textual metadata, plays a central role in this context. Despite significant advancements, existing CBIR systems still face challenges related to high-dimensional feature spaces, semantic gaps between low-level features and high-level concepts, and scalability when applied to large-scale datasets.

Digital image search and retrieval have been revolutionised in several fields by Content-Based Image Retrieval (CBIR) [1]. These categories include medical imaging, satellite images, and multimedia applications, among others. Medical diagnostics, surveillance systems, and social media platforms are just a few examples of the niche industries where digital picture collections are expanding at an exponential rate, making effective retrieval methods for direct visual content analysis imperative [2]. For large-scale applications, traditional text-based picture retrieval systems are impracticable due to their heavy reliance on user annotations. To overcome this shortcoming, CBIR systems automate retrieval based on picture content instead of metadata by extracting relevant information using enhanced feature descriptors [3].

Earlier iterations of CBIR methods relied on more simplistic visual characteristics [4]. Shukran *et al*. [5] laid out a thorough framework, which examined form, texture, and colour aspects in picture retrieval. By systematically evaluating recall, response time, and accuracy measures across various feature combinations, this study established important standards. Although individual features might capture picture attributes, this ground-breaking study showed that they frequently failed to depict complicated semantic connections [6]. Combining several types of features might greatly enhance retrieval accuracy, according to an examination of feature extraction methods; nevertheless, computing economy was still an issue. Due to the shortcomings of more conventional methods, researchers began looking into deep learning solutions. However, the use of raw deep features often

Manuscript received June 12, 2025; revised July 24, 2025; accepted October 9, 2025; published 16, 2026.

leads to high-dimensional vectors, many of which are redundant or non-informative for the retrieval task. This results in unnecessary computational overhead and reduced retrieval precision.

To overcome this, feature selection and dimensionality reduction techniques are employed to refine the feature vectors. Among various optimization methods, metaheuristic algorithms have shown promise due to their flexibility and effectiveness in exploring high-dimensional search spaces. The Grey Wolf Optimizer (GWO), inspired by the leadership hierarchy and hunting strategy of grey wolves, has gained popularity for its simplicity and performance. However, like many conventional metaheuristics, GWO is susceptible to premature convergence and local optima, especially in complex and high-dimensional feature spaces.

In this paper, we propose a novel hybrid framework that combines deep feature extraction with a Quantum Grey Wolf Optimization (QGWO) algorithm to address these issues. The QGWO enhances the exploration and exploitation balance of the original GWO by incorporating principles from quantum mechanics, such as wave function-based probabilistic representation and position updates informed by quantum tunnelling behaviour. This integration enables more robust and diversified search dynamics, leading to better feature subset selection.

Our contributions can be summarized as follows:

1) Hybrid deep feature architecture: We utilize pre-trained Convolutional Neural Networks (CNNs) (Inception V1, Inception V2 and ResNet) for extracting deep semantic features, capturing complex spatial hierarchies from images.
2) Quantum-inspired optimization: We propose a quantum-enhanced GWO for selecting optimal feature subsets, reducing redundancy, and improving the retrieval relevance.
3) Comprehensive evaluation: We conduct extensive experiments on benchmark CBIR dataset Corel-1K, Corel-5K and Corel-10K using standard evaluation metrics to validate the performance gains achieved by our method and evaluated the computational efficiency.
4) Comparative analysis: The proposed method is benchmarked against conventional QCSA-AlexNet, QGWO-ResNet 101, QGWO-IRV2.

The remainder of this paper is structured as follows: Section II reviews related work in deep learning-based CBIR and metaheuristic optimization. Section III details the proposed hybrid framework, including the feature extraction and QGWO algorithm. Section IV describes the experimental setup, datasets, and evaluation metrics. Section V presents and discusses the results. Finally, Section VI concludes the paper with potential directions for future research.

II. Related Work

The development of effective Content-Based Image Retrieval (CBIR) systems has evolved significantly over the past two decades. The key areas in CBIR are feature extraction, feature selection and optimization models for feature optimization.

An analysis of various feature extraction techniques demonstrated that the combination of multiple feature types could significantly improve retrieval accuracy, although computational efficiency remained a challenge. The limitations of traditional approaches led to the exploration of deep learning solutions. Significant progress was made by implementing pre-trained CNN models, specifically VGG16 and ResNet-50, through transfer learning [7]. A comparative analysis on the ImageNet dataset demonstrated how deep learning architectures could effectively capture hierarchical feature representations, substantially improving retrieval accuracy [8]. This work particularly highlighted the advantages of transfer learning in reducing training requirements while maintaining high performance across diverse image categories.

Hu and Bors [9] introduced an innovative co-attention mechanism to address the challenge of complex image scenarios by dynamically adapting to query content. This approach showed remarkable improvements in handling non-salient objects and complex backgrounds, particularly in scenarios where traditional methods failed.

Taheri *et al.* [10] developed a sophisticated semantic pyramid approach for handcraft feature fusion. This work introduced novel evaluation metrics through t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization and silhouette criterion analysis, providing deeper insight into feature behaviour and interpretability. The field further evolved with advanced feature representation techniques. Badiger *et al.* [11] proposed a comprehensive fuzzy graph model, incorporating deep representations and introducing adaptive learning mechanisms that improved retrieval accuracy., the persistent challenge of data imbalance was addressed through innovative fuzzy clustering-based feature normalization [12].

Deep learning, particularly CNNs, has revolutionized visual feature extraction by providing hierarchical, semantically rich representations that significantly outperform handcrafted features. Architectures such as have been widely used as feature extractors in CBIR systems, yielding promising results. With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), CBIR has seen a paradigm shift.

Rani *et al.* [13] proposed CBIR using separable CNN. Further deep learning models are CNNs like VGG [14], GoogLeNet [15], ResNet [16], and EfficientNet [17], ESA-ResNet34 [18], have demonstrated superior performance by automatically learning hierarchical features from raw image data. These models, typically pre-trained on large-scale datasets like ImageNet, are widely used as feature extractors by removing their classification heads and using the intermediate feature maps.

Zhu [19] proposed massive-scale image retrieval based on deep visual feature representation. The deep neural network played an important role in feature extraction process [20] when compared to exiting traditional model like proposed by Vu [21]. Nevertheless, one major drawback of deep features is their high dimensionality, which can cause storage overhead, slower similarity computation, and feature redundancy. This has led researchers to explore various strategies for feature selection and dimensionality reduction. The optimization of extracted features helps to improve the accuracy in process of retrieving the images. As an alternative, feature selection methods aim to identify and retain the most informative features directly from the original high-dimensional space. Metaheuristic optimization algorithms, such as Genetic Algorithms (GA) [22], Particle Swarm Optimization (PSO) [23], and Ant Colony Optimization (ACO) [24], have been extensively used for this purpose. These techniques enable adaptive and task-aware selection of features that contribute most significantly to retrieval performance. Among metaheuristic approaches, the Grey Wolf Optimizer (GWO) has gained attention due to its ability to balance exploration and exploitation based on the leadership hierarchy and hunting mechanism of grey wolves [25]. GWO has been applied in various domains including feature selection, scheduling, and medical image analysis. However, standard GWO may suffer from premature convergence and a limited ability to escape local optima in complex search spaces.

To address these limitations, quantum-inspired algorithms have emerged as promising alternatives. Quantum computing principles such as superposition, entanglement, and quantum tunnelling are leveraged to enhance diversity and search capabilities. The Quantum Grey Wolf Optimizer (QGWO) introduces quantum behaviour into the traditional GWO, where agents' positions are updated based on probabilistic wave functions rather than deterministic rules. This results in a more robust global search and better feature subset selection [26].

The motivation and novelty of proposed work despite the effectiveness of CNNs and metaheuristics individually, their integration—especially with quantum-inspired optimization—for CBIR-specific feature selection has not been sufficiently investigated. This paper fills that gap by introducing a hybrid deep feature-QGWO framework, leveraging the discriminative power of deep networks and the search efficiency of quantum-enhanced GWO. The proposed approach not only improves retrieval accuracy but also ensures computational tractability, making it suitable for large-scale image retrieval systems.

III. METHODOLOGY

CBIR is a technique that utilizes visual content or image features to search and retrieve relevant images from a database. CBIR operates by extracting low-level or high-level features (such as color, texture, or shape) from the images stored in the database. When a query image is provided, its features are extracted and compared against those in the database to identify the most similar images based on a defined distance metric, typically using the shortest feature distance.

The proposed framework consists of two main components: (1) deep feature extraction using pre-trained CNNs, and (2) feature selection using the proposed QGWO algorithm. Fig. 1 presents an overview of the entire architecture.

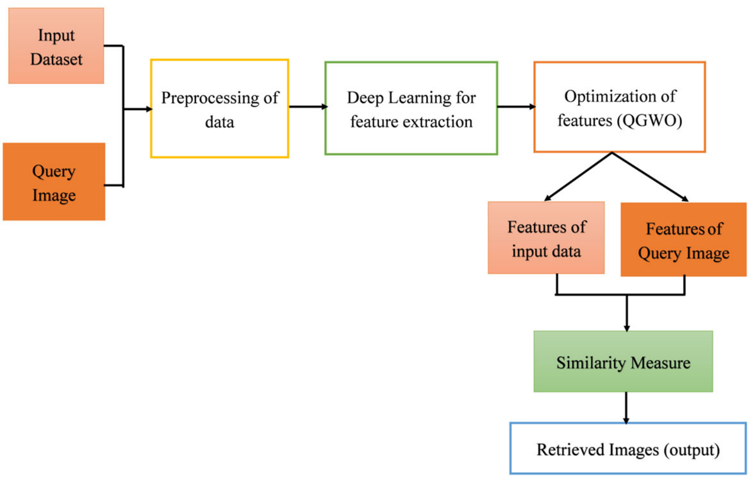

Fig. 1. Framework of model.

A. *Preprocessing*

The preprocessing stage is a critical initial step in any CBIR system, as it prepares the input images for effective feature extraction and retrieval. This stage ensures consistency, enhances image quality, and reduces variability caused by external factors. In this work, initially the images in the dataset are resized to a fixed resolution (e.g., 224×224) to ensure uniform input size for feature extraction or deep learning models. Secondly, noise in the image is filtered using median filter. This filter is applied to suppress unwanted noise and artifacts, which may otherwise degrade feature quality. Finally, images are normalised. The pixel intensity values are scaled to a common range (e.g., [0, 1] or [−1, 1]) to improve training stability and model generalization.

B. *Deep Feature Extraction*

In the first phase, input images are processed through pre-trained deep CNN models to extract high-level feature representations. Three well-established architectures, Inception V1, Inception V2 and ResNet-50 are employed due to their proven capability to encode rich semantic content across various domains. The approach of feature extraction is shown in Fig. 2.

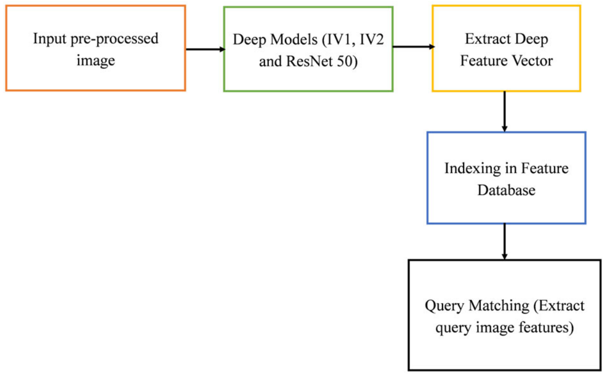

Fig. 2. Process of feature extraction.

1) Inception V1 (IV1)

InceptionV1 is a model which provides a powerful and efficient mechanism for extracting robust features. This method is suitable for CBIR due to its ability in extracting multiscale features in the inception module. The deeper network allows to extract high number of features. The process of proposed feature extraction using Inception V1 (IV1) is shown in Fig. 3.

The image is passed through convolutional and inception modules of the pre-trained network. Intermediate features capture spatial hierarchies, object textures, and contextual information. In this process the features are extracted from a fully connected layer. This layer provides a compact and discriminative feature vectors. All the extracted feature vectors for all database images are stored in a feature vector database. When a new image is provided as a query, it undergoes the same Inception V1-based feature extraction process. Each Inception module performs parallel convolutions with different filter sizes: 1×1 conv helps in reducing the dimensions. The 3×3 conv and 5×5 conv block capture medium and large spatial features. The 3×3 max pooling + 1×1 conv block preserves spatial info.

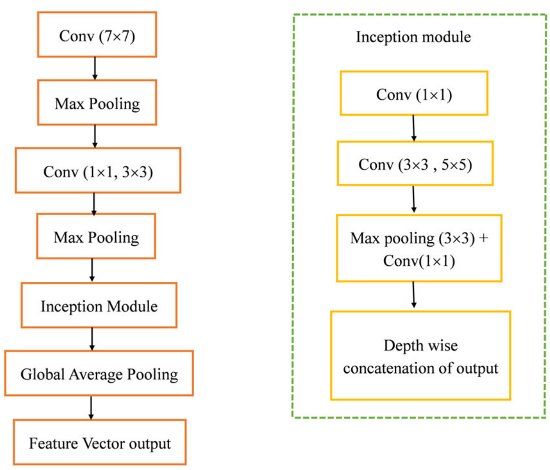

Fig. 3. Process of Inception V1.

2) Inception V2 (IV2)

It is an improved version of Inception V1 by performing factorization of convolution, batch normalization after every convolutional layer. The process of Inception V2 (IV2) is shown in Fig. 4.

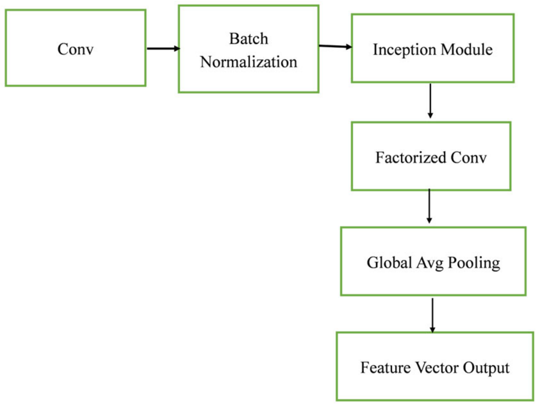

Fig. 4. Process of Inception V2.

In this process every convolution layer is followed by batch normalization by which the data training will be fast and provide better generalization. The inception module captures features at multiple scales simultaneously. The factorization layer involved in IV2 process helps to reduce the level of computation and maintain receptive field. Finally, the global average pooling produces compact feature vectors by averaging each feature map and obtain the feature vector output.

3) ResNet-50

A 50-layer residual network with identity shortcut connections that allow better gradient flow and deeper learning. Features are extracted from the penultimate average pooling layer. All input images are resized to a fixed resolution, normalized, and batch-processed to extract features. The resulting feature vectors are stored for subsequent optimization. The process of ResNet 50 in CBIR is shown in Fig. 5.

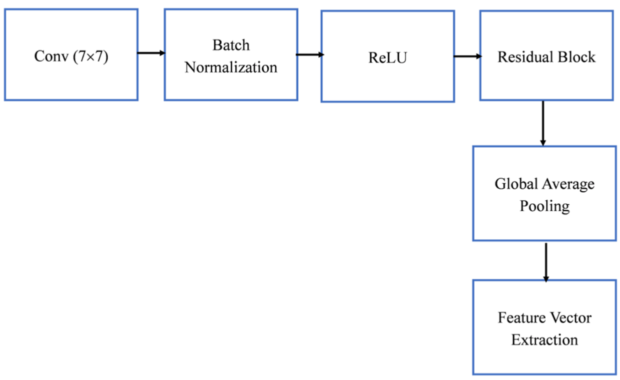

Fig. 5. Process of feature extraction using ResNet 50.

The output obtained after deep processing via global average pooling is used as a feature descriptor and is save in the databased. This vector which is stored is compared during retrieval using distance metrics. All the features achieved using the three deep learning models i.e., Multi scale features using IV1, optimized and normalized multi scale features using IV2 and deep residual feature with strong abstraction using ResNet-50 are combined and fetched to optimization model. The combination of all features improves the feature diversity and provide better semantic coverage which leads to improve the retrieval accuracy. The direct concatenation of features is given in Eq. (1).

$$F_{combined} = F_{IV1} + F_{IV2} + F_{ResNet50} \tag{1}$$

The features combined are fetched to optimization model for selection of best features. The best optimized features selection improves the accuracy of retrieving the images.

C. Problem Formulation for Feature Selection

Let $F = [f_1, f_2, ..., f_d]$ be the full deep feature vector of dimension d. The goal of feature selection is to find a subset $F^1 \subset F$, such that the selected features maximize the retrieval relevance while minimizing redundancy. This is framed as an optimization problem where each feature is either selected (1) or not selected (0). The objective function J is designed to:

$$J(z) = \lambda_1 \left(1 - mAP_{var}(z)\right) + \lambda_2 \frac{\|z\|_0}{d} \tag{2}$$

where $mAP_{var}(z)$ is the mean Average Precision computed using the selected subset z, $\|z\|_0$ counts the number of selected features, d is the total dimensionality. The weights are set such that $\lambda_1 \gg \lambda_2$. We minimize J.

D. Quantum Grey Wolf Optimization (QGWO)

The core of our approach lies in the novel QGWO algorithm, which enhances the original GWO by introducing quantum principles to diversify the search space and improve convergence. The behaviour of grey wolf needs to be addressed in which we considered candidate solutions (wolves) and update their positions based on the positions of three leading wolves alpha (α) beta (β) and delta (δ). The update rule is given in Eq. (3).

$$\vec{X}(t+1) = \frac{1}{3}\left(\vec{X}_\alpha + \vec{X}_\beta + \vec{X}_\delta\right) \tag{3}$$

where, $\vec{X}_\alpha = \alpha - A_1 \odot D_\alpha$, $\vec{X}_\beta = \beta - A_1 \odot D_\beta$, $\vec{X}_\delta = \delta - A_1 \odot D_\delta$. The term $D_\alpha = |C_1 \odot \alpha - x|$, $D_\beta = |C_2 \odot \beta - x|$, $D_\delta = |C_3 \odot \delta - x|$. The term A and C is given as, $A = 2ar_1 - a$; $C = 2r_2$. r_1, r_2 are random vectors.

In quantum enhancement the positions are encoded as probabilistic wave functions.

$$\psi_i = e^{-\frac{(z-\mu_i)^2}{2\sigma_i^2}} \tag{4}$$

where, μ_i is the current position of the i^{th} wolf, σ_i controls the spread (exploration), the new position is sampled from ψ_i using quantum tunnelling concepts. In additional the quantum mutation is applied to escape the local minima.

$$x_i^{new} = x_i + \eta \cdot randn() \cdot e^{-\gamma t} \tag{5}$$

where the term η is the mutation scale and γ is the decay rate over iterations.

Using QGWO, the Quantum behavior enables global search and avoids local minima. Removes uninformative dimensions from deep features, and reduce vector size which helps in speed up similarity computations.

Since deep feature vectors are continuous, we encode feature selection using a binary representation: each bit of z corresponds to a feature dimension, where 1 = selected and 0 = not selected. The continuous wolf positions in QGWO are mapped to binary using a sigmoid transfer function followed by thresholding ($\geq 0.5 \rightarrow 1$, $< 0.5 \rightarrow 0$). This ensures that QGWO optimizes over binary subsets while exploring the continuous search space.

QGWO Pseudo code:

Algorithm: QGWO Pseudo code

Input: d = feature dimension; N = number of wolves (population size); T = maximum iterations
Output: z_star = best feature subset (binary vector)
Step 1: Initialize population
Initialize population $X(N, d)$ with random values in [0,1]
for $i = 1 : N$
Step 2: Binary conversion using sigmoid transfer function
$Z(i,:) = \left(1./(1+\exp(-X(i,:))) >= 0.5\right)$;
Step 3: Compute fitness using objective function J (Eq. 2)
$Fitness(i) = J\left(Z(i,:)\right)$;
end
Step 4: Identify alpha, beta, delta wolves (best 3 solutions)
$[alpha, beta, delta] = select\_top(Fitness, Z)$
Step 5: Main loop
for t = 1: T
for i =1: N
Position update based on alpha, beta, delta wolves
$X(i,:) = update\_position(X(i,:), alpha, beta, delta, t, T)$;
Apply quantum mutation for diversification
$X(i,:) = quantum\_mutation(X(i,:), t, T)$
Binary conversion
$Z(i,:) = (1./(1+\exp(-X(i,:))) > 0.5)$;
Evaluate fitness
$Fitness(i) = J\left(Z(i,:)\right)$
end
Update alpha, beta, delta wolves
[alpha, beta, delta] = select_top3(Fitness, Z);
End
Step 6: Return best feature subset
$z\_star = alpha$;

E. Similarity Computation

Once the optimal feature subset F' is selected using QGWO, image similarity is computed using distance metrics. In this work a Euclidean Distance (ED) metric is utilized to calculate the distance measure. For each query image, the similarity scores are computed between its selected features and all database images, then ranked to produce retrieval results. The feature vectors represent the content of the image. Euclidean distance computes how far apart two feature vectors are in a multi-dimensional space. A shorter distance implies that both vectors (images) are very close in this feature space meaning the images are visually or semantically similar. The ED is evaluated using Eq. (6).

$$ED = \|F_1 - F_2\|_2 \tag{6}$$

where F_1 and $F_2 \in \mathbb{R}^n$ are feature vectors of length n.

Depending upon the measure the images are retrieved. The experimental results obtained using the proposed methodology is discussed in Section IV.

IV. Experimental Setup and Evaluation

This section outlines the experimental settings used to validate the proposed framework, including dataset descriptions, evaluation metrics, and implementation details.

A. Datasets

We evaluate our framework on three widely-used benchmark datasets for CBIR: Corel-1K: Consists of 1,000 images equally divided into 10 categories (e.g., dogs, persons, buses). Corel-5k consists of 5000 images from 50 categories and Corel-10K consists of 10,000 images form 50 categories. Each category contains images of 256×256 resolution. For the dataset, a subset of images is used as queries, while the remaining serve as the retrieval database.

B. Evaluation Metrics

To quantify the effectiveness of the retrieval process, standard evaluation metrics such as precision, recall, and F1-Score are used to evaluate the performance of the proposed model.

Precision measures the proportion of correctly retrieved relevant images among all images retrieved. The precision is evaluated using Eq. (7).

$$Precision = \frac{True\ positives}{True\ positives + False\ positives} \tag{7}$$

Recall measures the proportion of correctly retrieved relevant images out of all relevant images in the dataset.

$$Recall = \frac{True\ positives}{True\ positives + False\ positives} \tag{8}$$

Specificity indicates the model's ability to correctly identify non-relevant images.

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \tag{9}$$

Accuracy represents the overall correctness of the retrieval model by measuring the ratio of correctly retrieved images (both relevant and non-relevant).

$$Accuracy = \frac{True\ positives + True\ negatives}{Total\ number\ of\ cases} \tag{10}$$

Error percentage reflects the proportion of incorrect retrievals made by the model.

$$Error\ \% = \left(\frac{False\ positives + False\ negatives}{Total\ number\ of\ cases}\right) \tag{11}$$

Total processing time: The processing time of proposed model is evaluated using Eq. (12).

$$T_{total} = T_{extract} + T_{GWO} + T_{similarity} + T_{retrival} \tag{12}$$

C. Implementation Details

The proposed model is implemented in MATLAB using MATLAB toolbox for feature extraction and for vector operations. Optimization modules are custom-coded for both GWO and QGWO. The hardware parameters utilized for conducting our experiments are run on a workstation with Intel Core i9 CPU, 64GB RAM, and NVIDIA RTX 3080 GPU. The optimization parameters involved are shown in Table I.

TABLE I. Optimization Parameters

Number of wolves	**30**
Max iterations	50
Quantum mutation rate (η)	0.05
Exploration decay (γ)	0.01

All parameters are fine-tuned via grid search for each dataset. This section presents the empirical results of the proposed hybrid framework, comparing its retrieval performance with baseline methods. We also analyze the impact of Quantum Grey Wolf Optimization (QGWO) on feature selection, retrieval efficiency, and scalability. The evaluated results for the given input query are shown in Fig. 6.

(a)

(b)

Fig. 6. Input query image. (a) Sunrise, (b) Jet.

Table II presents the evaluation metrics obtained using the proposed model. The dataset encompasses multiple image categories, and the retrieval accuracy has been analyzed on a per-class basis.

The results in the Table III demonstrate that the proposed QGWO-IV1IV2-ResNet method achieves superior performance in terms of accuracy, precision, recall, and other evaluated metrics for CBIR. The Parameter Counts (PC) and Floating-point Operations Per Second (FLOPS) are shown in Table III.

TABLE II. OPTIMIZED QWO- IV1IV2RESNET

Class Name	Se/Recall (%)	Precision (%)	Specificity (%)	Error Percent (%)	Accuracy (%)
Africa	99.56	88.8	99.50	0.87	99.13
Beach	98.35	90.8	98.30	0.75	99.24
Buildings	99.35	93.85	99.30	0.70	99.29
Buses	99.58	86.4	99.50	0.81	99.18
Bear	99.84	90.74	98.80	0.89	99.11
Dinosaurs	99.49	86.4	99.40	0.97	99.03
Elephant	99.45	96.6	99.40	0.75	99.25
Flower	99.61	92.8	99.60	0.73	99.27
Horses	99.14	94.37	99.10	0.82	99.18
Mountain	98.91	86.15	98.90	0.74	99.26
Food	99.52	86.74	99.50	0.94	99.06
Tiger	99.36	94.09	99.30	0.76	99.24
Lion	99.17	90.33	90.10	0.93	99.07
Sports	99.62	90.81	99.60	0.74	99.26
Sunset	98.89	98.7	98.8	0.82	99.18
Snow	99.25	90.97	99.20	0.81	99.19
River	98.58	81.74	98.50	0.75	99.25
Planes	90.9	99.60	90.9	0.76	99.24
Bikes	98.15	92.40	98.10	0.70	99.30
Fruits	99.56	90.25	99.50	0.73	99.27
Overall Percentage	98.81	91.13	98.27	0.798	99.20

TABLE III. COMPARISON OF PROPOSED METHOD WITH EXISTING METHODS

Index	CART-Decision Tree [27]	CNN-ED [28]	QCSA-Alexanet [29]	QWO-ResNet 101 [30]	QWO-IRV2 [30]	ViT-GA [31]	Proposed QGWO-IV1IV2ResNet
Recall (%)	91.3	94.99	97.13	97.10	98.47	-	98.81
Precision (%)	75.0	71.46	77.37	77.60	88.48	-	91.13
Specificity (%)	95.6	94.85	97.13	97.28	98.17	-	98.27
Accuracy (%)	82.3	94.98	96.91	96.98	98.33	99.3	99.20
PC and FLOPS	-	-	-	-	-	-	25.6M Params and 4.1GFLOPs

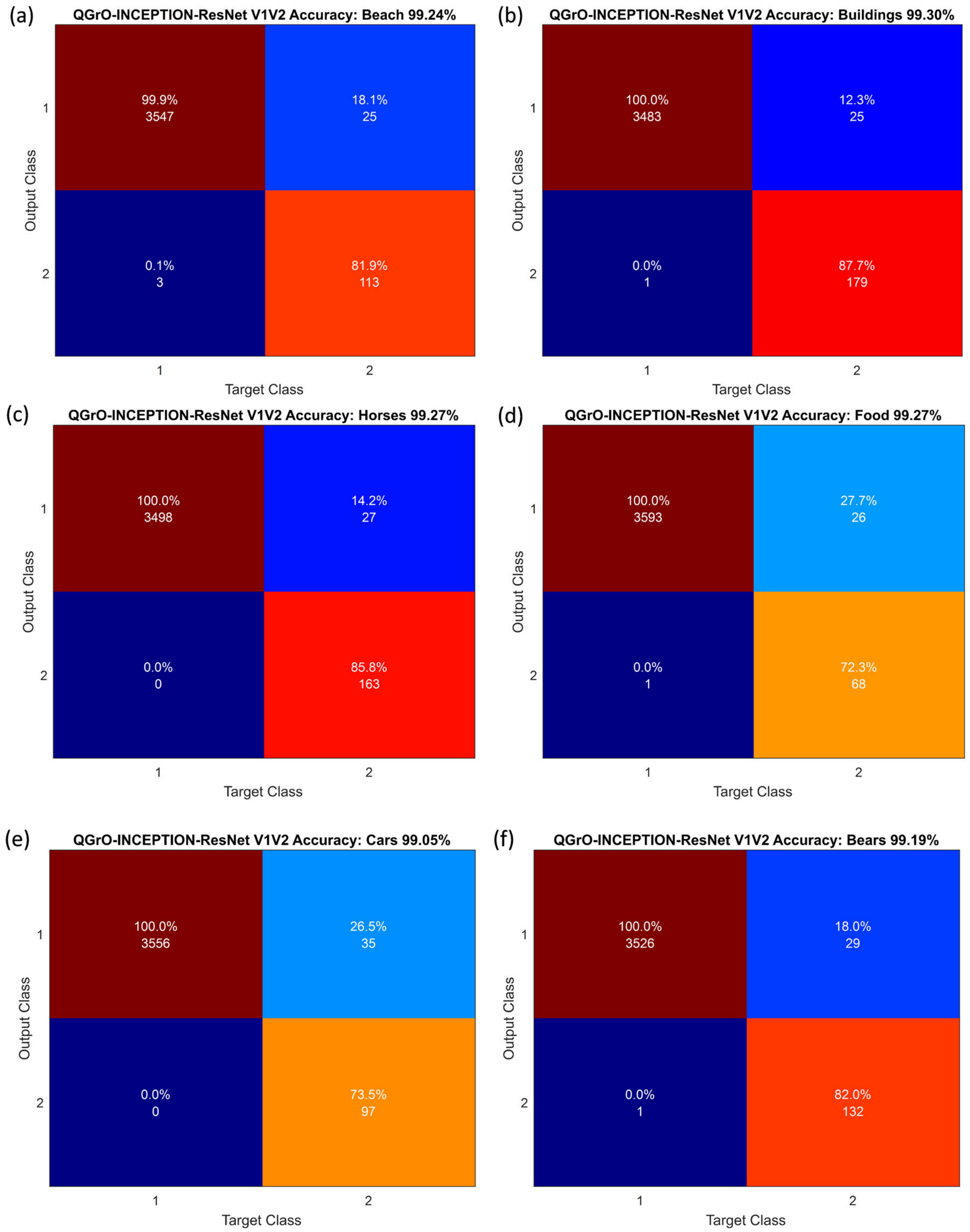

Fig. 7. Confusion matrices for different image classes. (a) Beach, (b) Buildings, (c) Horses, (d) Food, (e) cars, (f) Bear.

TABLE IV. COMPARISON OF PROPOSED METHOD WITH DIFFERENT DATASETS

Metrics	COREL-1K	COREL-5K	COREL-10K
Recall (%)	98.81	99.1	99.02
Precision (%)	91.13	90.21	90.18
Specificity (%)	98.27	99.10	99.02
Accuracy (%)	99.20	99.12	99.04
Error Percentage	1.08%	1.1%	1.16%
Retrieval Time	0.8 s	1.02 s	1.06 s
mAP	0.98	0.93	0.90
App Precison@K (K = 1)	0.99	0.99	0.99
App Recall@K (K = 1)	1.0	0.99	0.99

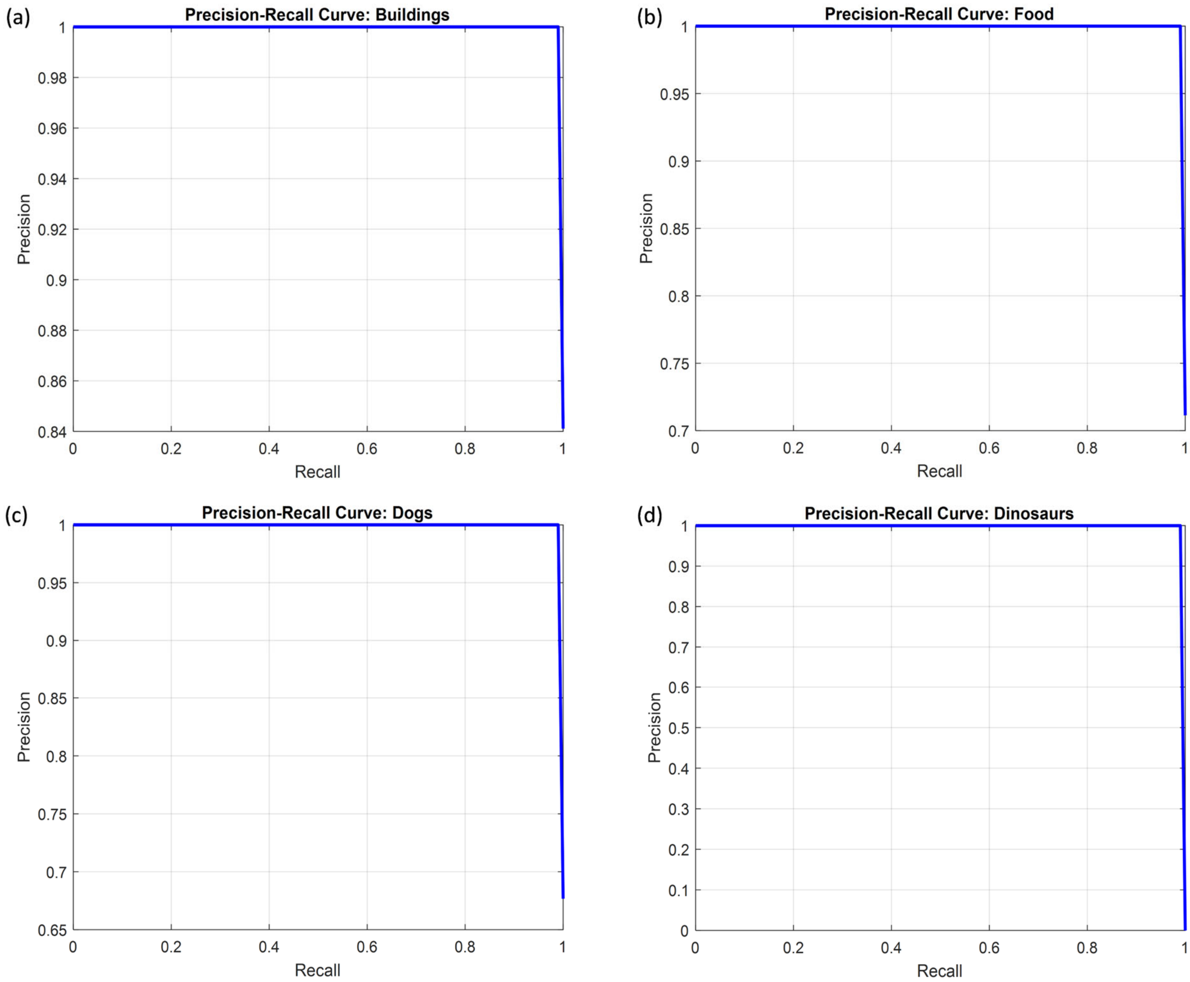

Fig. 8. PR curves. (a) Buildings, (b) Food, (c) Dogs, (d) Dinosaurs.

Individual confusion matrices are shown in Fig. 7 for the image classes: Bea`ch, Buildings, Horses, Food, cars, bear. Overall, the proposed model demonstrates high accuracy across most categories, with strong diagonal dominance indicating correct predictions. However, minor misclassifications are observed in visually similar classes for instance, some cars are occasionally misclassified as buses due to structural similarities. Similarly, a few cow images are confused with dogs, likely due to overlapping textures or backgrounds. Despite these cases, the model maintains consistent class-wise accuracy, confirming its robustness in handling diverse image categories.

The metrics evaluated for another dataset utilized are compared and shown in Table IV.

The precision recall curves are been evaluated and are shown in Fig. 8.

The retrieval time of image is evaluated. A typical CBIR system using deep CNN + Grey Wolf Optimization has a query processing time which is very less for all the three datasets utilized and is shown in Table V. The processing time will depend on system complexity, number of features, and optimization iterations. The comparison is of time is shown using optimization (QGWO selected subset) and without optimization (uses fully concatenated vectors).

TABLE V. PROCESSING TIME FOR RETRIEVAL OF IMAGE

Dataset Name	Time without optimization	Time with optimization
Corel 1K	3.22 s	0.8 s
Corel 5K	4.01 s	1.02 s
Corel 10K	5.04 s	1.06 s

D. Limitation of Work

While the proposed QGWO framework excels in feature optimization, its performance is still bounded by:

- The representational power of the base CNN (suggesting future use of Vision Transformers or CLIP).
- Manual parameter tuning in QGWO, which could be automated using reinforcement learning or meta-optimization

V. Ablation Study

We isolate the contribution of each design choice in our CBIR pipeline CNN backbone, feature post-processing, optimizer (QGWO vs. baselines), similarity metric, and handling on retrieval quality and efficiency. The ablation study conducted on Corel 1k dataset. The results achieved shown in Table VI.

TABLE VI. Results of Ablation Study

M/M	Re (%)	Pr (%)	Sp (%)	Acc (%)	Error (%)
ResNet	88.4	79.8	87.6	88.8	8.02
IV1	90.8	80.4	90.2	90.6	6.2
IV2	92.2	81.5	91.8	92.4	4.45
IV1+IV2+ResNet	94.5	84.6	93.8	94.2	2.96
IV1+IV2+ResNet+ QGWO	99.15	90.37	99.2	99.25	1.08

Inception V2 produces stronger deep features than Inception V1 and ResNet. Incorporating optimization significantly improves retrieval performance. Proposed Quantum Grey Wolf Optimization (QGWO) provides the best results, achieving the highest accuracy of 99.25% and lowest error rate of 1.08%.

As the optimization technique is important factor, the ablation study conducted by modifying search agents and iterations. The accuracy achieved using 10 search agents is 98.6% and 20 search agents is 99% and achieved 99.25% accuracy with 30 search agents. Increasing the number of agents improves performance up to a point (optimal at 30 agents), after which gains plateau. Increasing number of iterations to 100 and the accuracy achieved is 99.3% but the processing time increased to 1.4 s. Higher iteration counts slightly improve retrieval accuracy but at the cost of increased runtime.

VI. Conclusion

In this paper, we proposed a novel hybrid deep feature extraction framework integrated with QGWO for enhanced CBIR. By leveraging the powerful representational capacity of deep CNNs Inception V1, Inception V2, ResNet-50 and combining it with an advanced quantum-inspired optimization algorithm, our approach effectively addressed two critical challenges in CBIR: high-dimensional feature redundancy and suboptimal retrieval relevance. Extensive experiments across the benchmark dataset—Corel-1K, Corel-5K and Corel-10K demonstrated that the proposed method significantly outperforms baseline models in terms of retrieval precision, recall and retrieval accuracy, all while reducing the feature dimensionality. Comparative studies further showed that QGWO surpasses existing models in terms of accuracy, recall and precision. The percent of wrongly retrieved images obtained using the proposed model is 1.08%. The integration of quantum behaviour—through probabilistic sampling and tunnelling-inspired mutation—into the optimization process proved highly effective for deep feature selection, offering both robust convergence and adaptive search dynamics.

Despite its strong performance, the current framework opens several directions for further scope of work. Incorporating textual and audio cues using multimodal embeddings (e.g., CLIP or BLIP) could enhance semantic retrieval across domains. Leveraging Neural Architecture Search (NAS) or reinforcement learning could automate the tuning of QGWO parameters and improve generalizability across datasets. Ultimately, the proposed QGWO-based framework provides a scalable, accurate, and adaptive solution for modern CBIR systems, with strong potential for real-world deployment in fields such as digital asset management, medical imaging, and surveillance.

Conflict Of Interest

The authors declare no conflict of interest.

Author Contributions

Sarva Naveen Kumar conducted the research work, collected the data, and wrote the paper. Ch Sumanth Kumar supervised the work. All the authors had approved the final version.

References

[1] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *Journal of Digital Imaging*, vol. 24, no. 2, pp. 208–222, 2011. https://doi.org/10.1007/s10278-010-9290-9

[2] H. Müller, W. Müller, D. M. Squire *et al.*, "Performance evaluation in content-based image retrieval: Overview and proposals," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, 2001. https://doi.org/10.1016/S0167-8655(00)00118-5

[3] A. Shakarami and H. Tarrah, "An efficient image descriptor for image classification and CBIR," *Optik*, vol. 214, 164833, 2020. https://doi.org/10.1016/j.ijleo.2020.164833

[4] M. A. Hanif, H. Kaur, M. Rakhra, and A. Singh, "Role of CBIR in a different fields: An empirical review," in *Proc. 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST)*, 2022, pp. 1–7. https://doi.org/10.1109/AIST55798. 2022.10064825

[5] M. A. M. Shukran, M. N. Abdullah, and M. S. F. M. Yumus, "New approach on the techniques of Content-Based Image Retrieval (CBIR) using color, texture and shape features," *Journal of Materials Science and Chemical Engineering*, vol. 9, no. 1, p. 51, 2021. https://doi.org/10.4236/msce.2021.91005

[6] R. Bibi, Z. Mehmood, R. M. Yousaf, T. Saba, M. Sardaraz, and A. Rehman, "Query-by-visual-search: Multimodal framework for content-based image retrieval," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 5629–5648, 2020. https://doi.org/10.1007/s12652-020-01923-1

[7] G. Gautam and A. Khanna, "Content based image retrieval system using CNN based deep learning models," *Procedia Computer Science*, vol. 235, pp. 3131–3141, 2024. https://doi.org/10.1016/ j.procs.2024. 04.296

[8] N. Keisham and A. Neelima, "Efficient content-based image retrieval using deep search and rescue algorithm," *Soft Computing*, vol. 26, no. 4, pp. 1597–1616, 2022. https://doi.org/10.1007/s00500-021-06660-x

[9] Z. Hu and A. G. Bors, "Co-attention enabled content-based image retrieval," *Neural Networks*, vol. 164, pp. 245–263, 2023. https://doi.org/10.1016/j.neunet.2023.04.009

[10] F. Taheri, K. Rahbar, and Z. Beheshtifard, "Content-based image retrieval using handcraft feature fusion in semantic pyramid," *International Journal of Multimedia Information Retrieval*, vol. 12, no. 2, p. 21, 2023. doi.org/10.1007/s13735-023-00292-7

[11] R. M. Badiger, R. Yakkundimath, G. Konnurmath, and P. M. Dhulavvagol, "Deep learning approaches for age-based gesture classification in south Indian sign language," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13255–13260, 2024. https://doi.org/10.48084/etasr.6864

[12] V. H. Vu, "Content-based image retrieval with fuzzy clustering for feature vector normalization," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 4309–4329, 2024.

[13] S. Rani, G. Kasana, and S. Batra, "An efficient content-based image retrieval framework using separable CNNs," *Cluster Computing*, vol. 28, no. 1, p. 56, 2024. https://doi.org/10.1007/s10586-024-04731-w

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv Print, arXiv:1409.1556, 2014. doi.org/10.48550/arXiv.1409.1556

[15] C. Szegedy, W. Liu, Y. Jia *et al*., "Going deeper with convolutions," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. International Conference on Machine Learning*, *PMLR*, 2019, pp. 6105–6114.

[18] Y. Yuan, J. Sun, and Q. Zhang, "An enhanced deep learning model for effective crop pest and disease detection," *Journal of Imaging*, vol. 10, no. 11, p. 279, 2024. doi: 10.3390/jimaging10110279

[19] H. Zhu, "Massive-scale image retrieval based on deep visual feature representation," *Journal of Visual Communication and Image Representation*, vol. 70, 102738, 2020. https://doi.org/10.1016/j.jvcir.2019.102738

[20] G. Gautam and A. Khanna, "Content based image retrieval system using CNN based deep learning models," *Procedia Computer Science*, vol. 235, pp. 3131–3141, 2024. https://doi.org/10.1016/j.procs. 2024. 04.296

[21] V. H. Vu, "Content-based image retrieval with fuzzy clustering for feature vector normalization," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 4309–4329, 2024.

[22] U. A. Khan and A. Javed, "A hybrid CBIR system using novel local tetra angle patterns and color moment features," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 7856–7873, 2022. https://doi.org/10.1016/j.jksuci.2022.07.005

[23] M. Subramanian, V. Lingamuthu, V. Venkatesan *et al*., "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*, vol. 2022, no. 1, 3211793, 2022.

[24] A. Gautam and R. Bhatia, "A novel method for CBIR using ACO-SVM with DTCWT and color features," *Materials Today: Proceedings*, vol. 5, no. 1, pp. 1439–1446, 2018. https://doi.org/10.1016/j.matpr.2017.11.231

[25] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw*. vol. 69, pp. 46–61, 2014.

[26] S. K. Ali and D. Boughaci, "Quantum inspired grey wolf optimizer for convolutional neural network hyperparameter optimization," in *Proc. International Symposium on Quantum Sciences: Applications and Challenges*, 2023, pp. 50–64. https://doi.org/10.1007/978-3-031-59318-5_5

[27] G. W. Jiji and P. S. J. D. Raj, "Content-based image retrieval in dermatology using intelligent technique," *IET Image Process*, vol. 9, no. 4, pp. 1–12, 2014.

[28] M. Tzelepi and A. Tefas, "Fully unsupervised optimization of CNN features towards content based image retrieval," in *Proc. IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2018, pp. 1–5.

[29] S. N. Kumar and C. S. Kumar, "Fusion of CNN-QCSO for content-based image retrieval," *Journal of Advances in Information Technology*, vol. 14, no. 4, pp. 668–673, 2023.

[30] S. N. Kumar and C. S. Kumar, "QWO-IRV2 deep learning model for performing retrieval of multi object images," *International Journal of Engineering Trends and Technology,* vol. 72, no. 10, pp. 149–158, 2024. https://doi.org/10.14445/22315381/IJETT-V72I10P115

[31] P. Deekshita, V. Bonu, A. Ramyasri, V. Rani, K. Bodduru, and N. Karyemsetty, "A hybrid CBIR framework using vision transformers and genetic algorithm for enhanced image retrieval," *JASTT*, vol. 6, no. 2, pp. 277–289, 2025.