Fusion of CNN-QCSO for Content Based Image Retrieval

Sarva Naveen Kumar [1,]* and Ch. Sumanth Kumar [2]

[1] Department of Electronics and Communication Engineering, Vasavi College of Engineering, Hyderabad, India
[2] Department of Electronics and Communication Engineering, GITAM School of Technology, Gandhi Institute of Technology and Management (Deemed to Be University), Visakhapatnam, India
*Correspondence: snavin9ap@gmail.com (S.N.K.)

Abstract—**As the growth of digital images is been widely increased over the last few years on internet, the retrieval of required image is been a big problem. In this paper, a combinational approach is designed for retrieval of image form big data. The approach is CNN-QCSO, one is deep learning technique, i.e., Convolutional Neural Network (CNN) and another is optimization technique, i.e., Quantm Cuckoo Search Optimization (QCSO). CNN is used for extracting of features for the given query image and optimization techniques helps in achieving the global best features by changing the internal parameters of processing layers. The Content Based Image Retrieval (CBIR) is proposed in this study. In big data analysis, CNN is vastly used and have many applications like identifying objects, medical imaging fields, security analysis and so on. In this paper, the combination of two efficient techniques helps in identifying the image and achieves good results. The results shows that CNN alone achieves an accuracy of 94.8% and when combined with QCSO the rate of accuracy improved by 1.6%. The entire experimental values are evaluated using matlab tool.**

Keywords—**Content Based Image Retrieval (CBIR), Convolutional Neural Networks (CNN), cuckoo search optimization, Quantum Cuckoo Search Optimization (QCSO)**

I. Introduction

Multimedia material is essential in a variety of fields such as research, medical treatment, and social networking, among others. As a result, the retrieval of image system is a need for individual in the present scenario. The multimedia database comprises a massive quantity of information in many formats such as text, audio, pictures, and videos. Content-Based Image Recovery (CBIR) has been one of the most demanding study fields in the previous decade and in presents also due to increase in digital image in world wide web and the visual complexity of the images are high. Humans can comprehend and analyze picture content, but machines cannot. There is a huge distinction between human insight and machine portrayal, which is alluded to as semantic gap. Accessing multimedia databases is a difficult challenge for CBIR due to the semantic gap. There have been studies to limit the research gap. To close the semantic gap between human undeniable level insight and lower level of machine description, numerous methodologies have been created. Because digital pictures are rich in substance and lack linguistic constraints, the substance of a picture assumes a part in the quest for comparable pictures as suggested by Gosh, Agarwal, and Motwani [1].

One of the most efficient and effective way for retrieving an image's visual data is regarded as CBIR by Haji and Alkawaz *et al.* [2]. Military affairs, medical sciences, education, architecture, law, and agriculture are just a few of the fields where CBIR is used. CBIR is seeking for photos that are comparable to the query digital image in terms of content, without the usage of textual information, labels, or tags. Image retrieval may be divided into two approaches: based on text retrieval and based on content retrieval. The text-based technique has some obvious flaws since each person's interpretation of each written description might change. When working with really big databases, it is also time intensive. The retrieval of visual data based on content necessitates a paradigm that differs considerably from both standard databases and text-based image comprehension systems by Dubey and Choubey [3]. CBIR is a strategy for keeping away from these issues. "Content-based" demonstrates that the hunt will take a look at the picture's real substance as opposed to the metadata associated with it, like watchwords, labels, or potentially descriptions. The frequency of occurrence of each colour in a picture is shown by the Conventional Colour Histogram (CCH) by Suhasini and Krishna [4]. Another significant aspect of pictures is their texture. In pattern recognition and computer vision, many texture representations have been studied. Section II discusses many approaches that have been applied.

Content-based image retrieval is a search strategy that makes use of the picture's structural and conceptual properties. Because of the growing demand for optimum picture retrieval in huge data bases, image retrieval research has received a lot of attention. The majority of research in this subject focuses on creating ways for

Manuscript received October 23, 2022; revised November 16, 2022; accepted February 6, 2023; published July 19, 2023.

doi: 10.12720/jait.14.4.668-673

quickly storing and retrieving pictures based on their content. In this article, a convolution neural network is utilized to extract features, which are then improved using the quantum cuckoo search. These enhanced characteristics contribute to a higher percentage of picture retrieval accuracy.

II. Related Work

Local characteristics comprised of spatial domains were assessed in [5] by Nazir and Ashraf *et al.* to show the consequences of comparing an image to an image collection. The segmentation method was required for overall feature extraction. There are two types of feature extraction: local features and global features. Color, shape, and texture are examples of local characteristics that may be used to detect things. For object categorization, global characteristics are used. To combine the colour and texture aspects, a novel CBIR technique was created. The colour information was mined using a Colour Histogram (CH). The Discrete Wavelet Transforms (DWT) and edge histogram descriptor was used to mine texture information. One significant drawback of this work was that it did not employ machine learning techniques such as Artificial Neural Network (ANN).

Devi and Hemachandran [6] described a novel approach for facial recognition that makes use of deep neural networks and classification by using support vector machine. From the pictures returned by the retrieval procedure, the recognizer of query image is done by using Support Vector Machine (SVM). Thus, in terms of recognition rate and retrieval time, the CBIR technique outperformed other traditional approaches. Lu and Yang *et al.* [7] developed a unique face recognition method that combines form and texture characteristics. The suggested method employs the representation of features i.e. shapes and texture, rapid retrieval of faced images, and finds other methods for retrieval of images. Face texture data was retrieved using a modified Google Net. Following that, these two characteristics were combined and balanced using Principal Component Analysis (PCA). A coarse-to-fine search technique was used to locate effectively comparable things in order to enhance efficiency. The scalability and speed of face characteristics have been enhanced. Wang and Liang *et al.* [8] discussed the difficulties that Content-Based Image Retrieval (CBIR) faces. A CBIR stores the image database in the index file so that it may be joined with the original picture. Vector format is used to express image descriptors. To enhance the picture retrieval rate, the form and texture characteristics were efficiently integrated. According on the experimental results, the suggested technique produced significant improvements on well-known datasets when compared to other standard methods.

Liu and Yang [9] supported a Colour Difference Histogram (CDH) image feature representation technique, which was used to characterize image characteristics for image retrieval. This technique showed strong discriminating power for colour, texture, form, and spatial arrangement characteristics. The performance of suggested method is evaluated by measuring the Precision and recall parameters. To get good outcomes, two features are combined. The authors of presented an effective approach for retrieving pictures of faces in [10] by Jian and Lam. Singular values and potential-field representation were used in this approach. The image is split for representation and employs rotation invariant, shift invariant, and scale-invariant properties. In content-based picture retrieval, Guo and Prasetyo *et al.* [11] employed error diffusion block truncation coding characteristics. To index video, an adaptation of the Error Diffusion Block Truncation Coding (EDBTC) image retrieval system was utilized, which saw the video as a sequence of pictures. Then, the Colour Histogram Feature (CHF), Bit pattern histogram detects a match between the query and the targeted picture. This approach yielded great accuracy, and the characteristics were incorporated into EDBT indexing. The work's tremendous intricacy was one of its drawbacks. The suggested approach efficiently created an image index for CIBR and image compression.

III. Proposed Work

This section describes how to use the image retrieval and classification approach to categories the available images in the database which is used to process the results. The process of proposed work is shown in Fig. 1. The pictures are initially obtained as input from a dataset for the purpose of pre-processing. Following that, the irrelevant data is eliminated using filter technique, i.e., Gaussian filter, and features are retrieved using Convolutional Neural Networks (CNN) to acquire the best traits of colour and texture. Quantum cuckoo search optimization is used to optimize the characteristics acquired. The obtained optimized features and database features are combined so that mini-Euclidian distance is calculated. Finally, the image is retrieved. It is possible to retrieve images depending on the query image and the text.

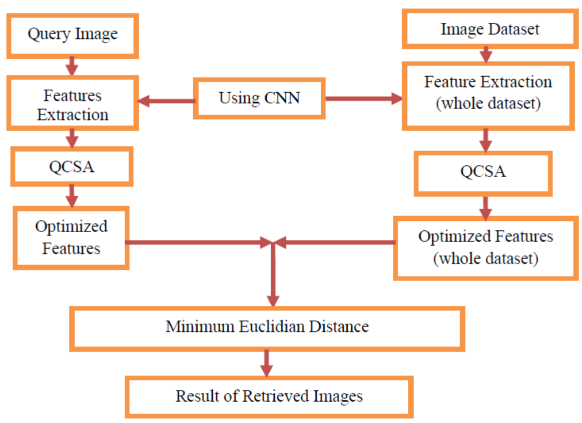

Figure 1. Proposed process flow of CBIR.

A. Database

The suggested work's performance is evaluated utilizing Corel [12] datasets. Corel image collection comprises 1000 pictures divided into twenty categories, including aero planes, bicycles, boats, birds, bottles, buses, cars, cats, chairs, cows, dining tables, dogs, horses, motorbikes, potted plants, sheep, sofas, trains, and television monitors. All picture dimensions are changed to 227×227 for feature extraction using CNN and feature optimization with Quantm Cuckoo Search Optimization (QCSO).

B. Extraction of Features

The suggested work employs CNN as a technique for extracting features rather than a traditional feature extractor. CNN has emerged as an important research subject in the fields of machine learning and system vision. So the primary goal of utilizing CNN as a feature extractor is to compare a CNN-based CBIR system to a traditional CBIR system to see whether it is superior in any manner. Aside from classification of images, for identification of objects CNN is effectively used. In this paper, we use CNN's Alex Net architecture to extract features suggest by Krizhevsky and Sutskever *et al.* [13]. The Alex Net is made up of eight layers which are trained, in which they are divided into two parts. First part comprises of five layers and is known as convolutional layers, whereas the second part comprises of three layers known as fully connected layers. With 4096 dimensions/features per picture, this study employs the architecture's 7th layer for feature extraction. The CNN's Alex Net method begins with the picture dataset for feature extraction. The features which are retrieved are saved in feature database. Later, the user sends a query image to evaluate the performance of proposed method based on the similarity of the characteristics that are extracted by employing a distance algorithm that ranks the results in descending order to obtain the top most relevant results. The layers available in CNN is shown in Fig. 2.

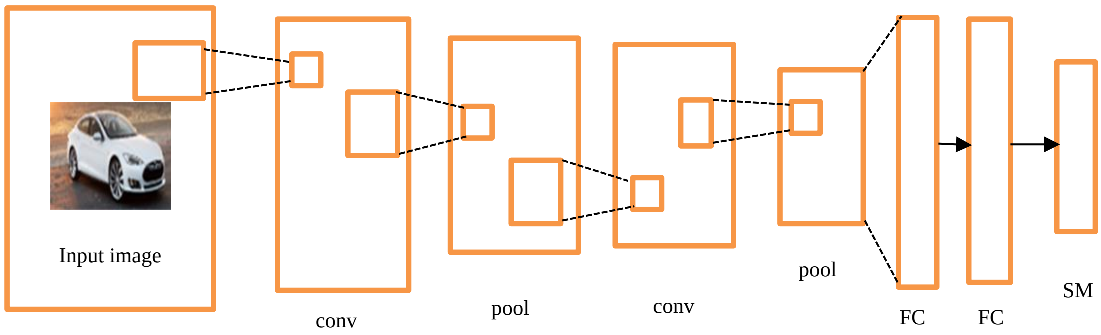

Figure 2. Architecture of CNN.

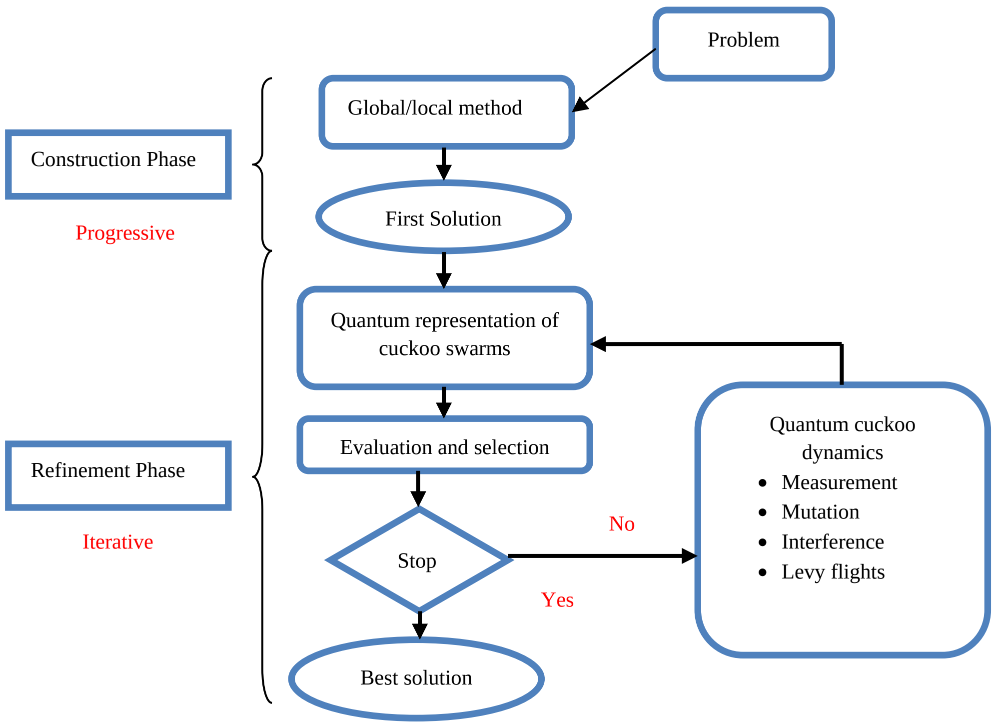

Figure 3. Process of Quantum Cuckoo Search (Kartous and Layeb *et al*. [15]).

C. Feature Optimization

For optimization, a Quantum Cuckoo Search (QCS) technique is used. A QCS method is utilised for getting optimal features from colour, shape, and texture characteristics that are extracted, which is the essential point in making those algorithms function. The goal function utilised in this case is evaluating the correlation for each extracted picture feature. The optimised features are separated into train and test features before being processed, and the outputs are classified as training and testing categories. These inputs are then analysed using a CNN classifier in order to get the relevant pictures. The quantum technique is based on cuckoo search, as demonstrated in the code below.

Algorithm 1 (Yang and Deb [14]):

Create a goal function $g(x), y = (y_1, \ldots\ldots, y_d)^T$
Generate an initial population of n number of host nests, $y_i (i = 1, 2, \ldots, n)$;
While (t <Max Generation) or (stop criterion);
– levy flights are used for getting a cuckoo (i) randomly;
– fitness function is evaluated Fit_i;
– nest is chosen and denoted as j randomly;
if (Fit_i*>Fit_j),
the new solution is replaced in place of j;
end
– Abandon a fraction (pa) of worse nests
–Lévy flights for building new nest at new locations;
– Best solution is saved (or nests with quality solutions);
– Assigning rank for the solutions and achieving the best;
end while

The proposed strategy is divided into two stages and is shown in Fig. 3. The phase one is a construction phase that is used to build initial solutions and is based on a stochastic progression of CBIR; the phase two is a refinement technique that is used to enhance the solutions from the phase one and is based on the dynamics of quantum cuckoo algorithm.

D. Measure of Distance

The measure of distance between the optimized features and the feature database available in dataset is calculated using the distance of Euclidean. It is used to compare query picture features to dataset features as shown in Eq. (1). A set of relevant pictures is collected and organized in decreasing order of their Euclidean distance score. In Eq. (1), "x" and "y" termed to be the two dimensions of a picture and dimension of image available in dataset is termed as "a" and "b".

$$dis\big((x, y), (a, b)\big) = \sqrt{(x-a)^2 + (y-b)^2} \tag{1}$$

IV. Results and Discussion

A fully unsupervised retraining approach is provided that takes use of the data's geometrical structure to improve the deep CNN descriptor efficiency in the CBIR assignment. Following that, the use of max pooling method to retrieve the feature representations from the activations of the final convolutional layer is proposed. In this work, it is suggested to retrain the weights of the CNN model which is pre-trained on the supplied information, with the goal of reducing the distance, i.e., Euclidean Distance (ED) between each picture description and its closest description. This procedure is carried out in the fully linked layer.

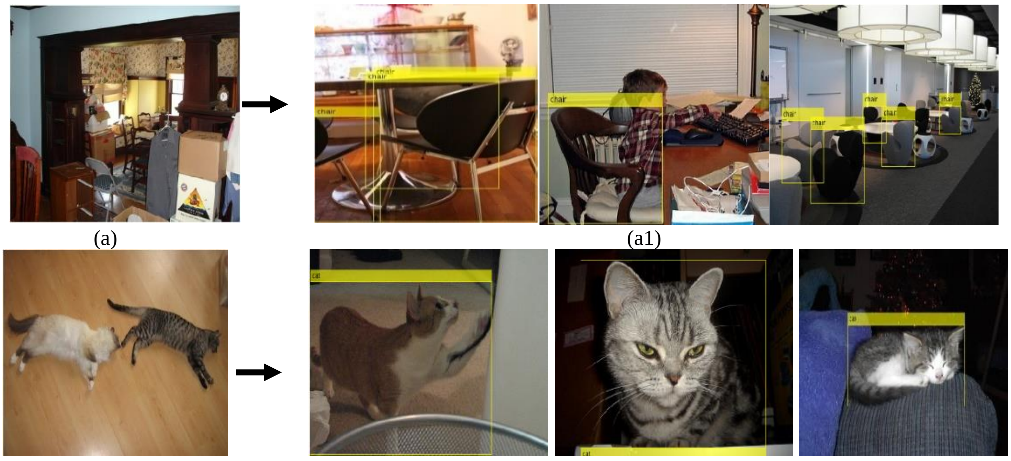

Figure 4. (a) Input chair image; (a1) retrieved images; (b) Input cat image; (b1) retrieved cat images.

From Corel database, all images have been used as query images (the tested 20 semantic class includes aero plane, bicycle, boat, bird, bottle, bus, car, cat, chair, cow, dining table, dog horse, motorbike, potted plant, sheep, sofa, train, television monitor) and then the first 20 most similar images are retrieved. The images retrieved using proposed technique is shown in Fig. 4.

The results using Euclidean distance calculation in Fully Connected (FC) layer has obtained, cuckoo search based, quantum cuckoo search based and other optimization techniques will be compared. The parameters are evaluated and the results will be compared. The performance of proposed methodology is evaluated using parameters like Sensitivity (Se), Specificity (Sp) and Accuracy (A).

$$Se = \frac{Tuely\ Positive}{Truely\ Positive + Falsely\ Negative} \tag{2}$$

$$Sp = \frac{Truely\ Negative}{Truely\ Negative + Falsely\ Positive} \tag{3}$$

$$A = TP + TN / TP + TN + FP + FN \tag{4}$$

$$Precision = \frac{Number\ of\ retrived\ relevant\ images}{No\ of\ relevent\ images} \tag{5}$$

TABLE I. RESULTS OBTAINED USING CNN-EUCLIDIAN DISTANCE

Class Name	Se/Recall	Precision	Specificity	Error Percent	Accuracy
"Aeroplane"	96.13	68.74	96.16	4.98	95.01
"bicycle"	94.36	68.00	94.34	5.39	94.60
"bird"	96.73	76.28	96.64	4.39	95.60
"boat"	93.49	65.44	93.44	4.63	95.36
"bottle"	93.35	71.34	93.41	4.88	95.11
"bus"	94.64	67.82	94.66	4.58	95.41
"car"	95.31	83.18	95.37	5.26	94.74
"cat"	96.30	72.17	96.31	5.50	94.49
"chair"	95.82	77.78	95.84	4.85	95.14
"cow"	93.60	61.77	93.59	5.72	94.27
"dining table"	94.26	65.76	94.19	5.47	94.52
"dog"	96.33	76.75	96.34	4.77	95.22
"horse"	94.10	71.64	94.06	4.79	95.20
"motor bike"	94.24	69.25	94.28	5.01	94.98
"person"	94.36	92.02	94.35	5.72	94.27
"pottedplant"	95.94	72.14	95.95	4.47	95.52
"sheep"	95.68	60.71	95.68	4.60	95.39
"sofa"	95.40	69.19	95.43	4.71	95.28
"train"	95.09	69.49	95.09	5.26	94.74
"tv monitor"	94.69	69.60	94.65	5.23	94.76
"Overall Percentange"	94.99	71.46	94.99	5.01	94.98

TABLE II. RESULTS OBTAINED USING QCSA-CNN

Class Name	Se/Recall	Precision	Specificity	Error Percent	Accuracy
"Aeroplane"	96.85	72.47	96.82	3.68	96.31
"bicycle"	97.22	72.59	97.26	3.79	96.20
"bird"	96.40	81.73	96.38	2.87	97.12
"boat"	97.32	71.32	97.39	2.92	97.07
"bottle"	97.47	78.17	97.53	2.92	97.07
"bus"	95.88	72.36	95.85	3.19	96.8
"car"	95.60	86.66	95.64	3.93	96.06
"cat"	97.06	78.13	97.08	3.44	96.55
"chair"	96.68	81.52	96.70	3.60	96.39
"cow"	97.83	69.19	97.77	2.98	97.01
"dining table"	97.23	75.78	97.23	2.46	97.53
"dog"	97.29	82.17	97.25	3.09	96.90
"horse"	97.13	77.83	97.09	3.00	96.99
"motor bike"	98.22	77.15	98.25	2.73	97.26
"person"	96.92	96.16	96.90	2.82	97.18
"pottedplant"	97	76.81	96.96	3.11	96.88
"sheep"	96.7	66.76	96.70	2.52	97.47
"sofa"	97.86	73.29	97.84	3.44	96.55
"train"	97.54	78.66	97.56	2.57	97.42
"tv monitor"	98.36	78.58	98.27	2.603	97.42
"Overall Percentange"	97.13	77.37	97.13	3.08	96.91

In Table I, the values shown are without using optimization technique. In this the parameters evaluated are with the help of CNN and Euclidian distance. The overall retrieval accuracy is around 94.98%.

TABLE III. COMPARISON OF PROPOSED METHOD WITH EXISTING METHODS

	CNN-ED ([16]	CART- Decision Tree [17]	**Proposed QCSA-CNN**
Recall	94.99	91.3	**97.13**
Precision	71.46	75.0	**77.37**
Specificity	94.99	95.6	**97.13**
Accuracy	94.98	82.3	**96.91**

In Table II, the values shown are with using optimization technique. In this the parameters evaluated are with the help of QCSA- CNN and Euclidian distance. The overall retrieval accuracy is around 96.91%. The overall accuracy and other parameter are compared with the existing techniques and the values are shown in Table III.

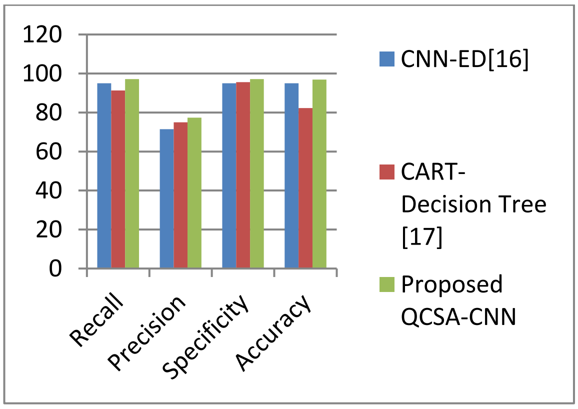

Figure 5. Comparison of different techniques.

The graphical comparison is shown in Fig. 5. The QCSA-CNN model obtains better results in terms of rate of accuracy, precision rate and recall rate than the CNN-ED and CART-Decision tree owing to the accurate clustering model, which is shown in Fig. 5.

The accuracy of proposed method is better when compared to existing techniques CNN-ED and CART-Decision tree.

V. Conclusion

A novel content-based picture retrieval strategy with suitable shape and colour characteristics is provided in this study. The feature extraction phase is designed with a convolutional neural network in such a manner that the distance between features of the same image class is minimised while it is maximised for pictures of different classes. This approach enhanced performance. To acquire the convolutional representations and generate the target representations, a CNN model is used. For the retrieval task, the CNN model generates more efficient and concise picture descriptors. Furthermore, Quantum Cuckoo search optimization removes all unnecessary and redundant characteristics before selecting the most relevant features from the whole feature set. The results demonstrate that our proposed QCSA-CNN CBIR schema outperforms CNN-based CBIR systems in terms of precision, recall, and accuracy for several picture categories.

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 and all authors had approved the final version.

References

[1] N. Ghosh, S. Agrawal, and M. Motwani, "A survey of feature extraction for content-based image retrieval system," in *Proc. International Conference on Recent Advancement on Computer and Communication*, 2018, pp. 305–313.

[2] M. S. Haji, M. H. Alkawaz, A. Rehman, and T. Saba, "Content-based image retrieval: A deep look at features prospectus," *International Journal of Computational Vision and Robotics*, vol. 9, no. 1, pp. 14–38, 2019.

[3] R. Dubey, R. Choubey, and S. Dubey, "Efficient image mining using multi feature content based image retrieval system," *Int. Jr. of Advanced Compute Engineering and Architecture*, vol. 1, 2011.

[4] P. S. Suhasini, K. S. R. Krishna, and V. M. Krishna, "CBIR using color histogram processing," *Journal Theoretical and Applied Information Technology (NWTIT)*, pp. 110–115, 2009.

[5] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, "Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor," in *Proc. International Conference on Computing, Mathematics and Engineering Technologies, IEEE*, 2018, pp. 1–6.

[6] N. S. Devi and K. Hemachandran, "Content based feature combination method for face image retrieval using neural network and SVM classifier for face recognition," *Indian Journal of Science and Technology*, vol. 10, no. 24, pp. 1–11, 2017

[7] Z. Lu, J. Yang, and Q. Liu, "Face image retrieval based on shape and texture feature fusion," *Comput. Vis. Media*, vol. 3, no. 4, pp. 359–368, 2017.

[8] X. Y. Wang, L. L. Liang, Y. W. Li, H. Y. Yang, "Image retrieval based on exponent moments descriptor and localized angular phase histogram," *Multimedia Tools Appl.*, vol. 76, no. 6, pp. 7633–7659, 2017.

[9] G. H. Liu and J. Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[10] M. Jian and K. M. Lam, "Face-image retrieval based on singular values and potential-field representation," *Signal Process*, pp. 9–15, 2014.

[11] J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Trans Circuits System Video Technology*, vol. 25, no. 3, pp. 466–481, 2016.

[12] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1–9, 2012.

[14] X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congress on Nature and Biologically Inspired Computing*, India, 2009, pp. 210–214.

[15] W. Kartous, A. Layeb, and S. Chikhi, "A new quantum cuckoo search algorithm for multiple sequence alignment," *Journal of Intelligent Systems*, vol. 23, no. 3, pp. 261–275, 2014.

[16] 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.

[17] G. W. Jiji and P. S. J. D. Raj, "Content-based image retrieval in dermatology using intelligent technique," *IET Image Process*, pp. 1–12, 2014.