Content-based Image Retrieval using Encoder based RGB and Texture Feature Fusion

Charulata Palai[1], Pradeep Kumar Jena[2*], Satya Ranjan Pattanaik[3], Trilochan Panigrahi[4], Tapas Kumar Mishra[5]
Department of Computer Science and Engineering, NIST Institute of Science and Technology, Berhampur-761008, India[1, 2]
School of Computing, Gandhi Institute for Technology, Bhubaneswar-752054, India[3]
Department of ECE, National Institute of Technology, Goa-403401, India[4]
Department of Computer Science and Engineering, SRM University-AP, Amaravati-522240, India[5]

Abstract—**Recent development of digital photography and the use of social media using smartphones has boosted the demand for image query by its visual semantics. Content-Based Image Retrieval (CBIR) is a well-identified research area in the domain of image and video data analysis. The major challenges of a CBIR system are (a) to derive the visual semantics of the query image and (b) to find all the similar images from the repository. The objective of this paper is to precisely define the visual semantics using hybrid feature vectors. In this paper, a CBIR system using encoded-based feature fusion is proposed. The CNN encoding features of the RGB channel are fused with the encoded texture features of LBP, CSLBP, and LDP separately. The retrieval performance of the different fused features is tested using three public datasets i.e. Corel-lK, Caltech, and 102flower. The result shows the class properties are better retained using the LDP with RGB encoded features, this helps to enhance the classification and retrieval performance for all three datasets. The average precision of Corel-lK is 94.5% and it is 89.7% for Caltech, and 88.7% for the 102flower. The average f1-score is 89.5% for Caltech, and 88.5% for the 102flower. The improvement in the f1-score value implies the proposed fused feature is more stable to deal the class imbalance problem.**

***Keywords*—CBIR; CNN Encoded Feature; LBP; CSLBP; LDP; feature fusion**

I. INTRODUCTION

Content-based image retrieval (CBIR) is the technique to retrieve similar images from a large image database using visual characteristics such as color, shape, structure, Zernike values, and histogram of the images[1, 2]. Nowadays it has an inevitable requirement in various application areas such as video surveillance, medical image retrieval, crime detection, military surveillance, remote sensing applications, the textile industry etc. [3-6]. The efficiency of the CBIR system greatly depends upon the visual feature selection. The high-level semantic features [3,7] of an image are its color, shape, structure, Zernike values, and histogram are used for manual image annotation and are less biased with noise [8,9]. Features represented using the spatial layout of the pixels within an image patch are referred as low-level features or local descriptors [10-12]. Some of the popular low-level image descriptors are Local Binary Patterns (LBP) [13-15], Orthogonal-Combination of Local Binary Patterns (OC-LBP), Center-Symmetric Local Binary Patterns (CS-LBP), Local Ternary Patterns (LTP), Local Directional Patterns (LDP)[15], Scale-Invariant Feature Transform (SIFT) [16] are used for image retrieval. The performance of a unique texture feature varies with different datasets. The major limitation of the texture feature is directly mapping the texture image to its histogram [1, 9, 17-19], which is represented on a scale of 0 to 255, so that all the information learned from the patches of images are not well preserved. With the implementation of Deep learning features a new breakthrough is achieved in the field of computer vision and its applications. It uses the Convolutional neural networks (CNNs) features [14, 20-23] as the image descriptor. The Deep learning technique requires adequate images for its training. The several layers of the CNN encoder represent the image features at different levels [11]. The lower layers contain the detailed image features, whereas the higher layers present the semantic information of the image [10, 11]. The fully connected layer extracts discriminative image features using an order-less quantization approach. Finally, these features are mapped to the class label using the dimension reduction technique and soft-max pooling [10, 24].

An effectual feature extraction technique precisely describes the image contents. It also helps to maintain a distinctive signature for the images of different classes. In recent years image retrieval using feature fusion has been emphasized by many researchers[3, 8] to build a more powerful image descriptor using the feature fusion technique [7,10,23,25-27]. These are more sensitive to noise and image resolution. Moreover mapping the low-level image features to the high-level visual semantics is challenging [7, 8, 28, 29]. Thus, there is a need to design an enhanced CBIR system.

In this work, a deep-learning feature fusion framework is proposed, where the auto-encoding features of the RGB channels are fused with the auto-encoding feature of the texture image. Here two different CNN models are trained independently. The first model usages the RGB channels data, which learns the spatial image information using automatic encoding. The second model usages the texture image data for the training to learn the auto-encoder-based texture features. The spatial and texture features extracted by CNN encoders are fused together to provide more precise feature descriptors for the image. The texture feature of an image i.e. the histogram of the texture image is biased by the background image textures, which impedes the learning ability of the classifier [9, 17]. Textures of similar images are expected to be alike. More effective learning can be possible from the texture image set, as the CNN uses the batch mode for the training. For the extensive analysis of the proposed fusion framework, a CBIR system is developed. Here the encoding features of three different textures such as LBP, CSLBP, and LDP are fused with the RGB channel encoding feature individually. The classification and retrieval performance of the different fused features are presented. These are also compared with encoding features of only RGB channels. The model is tested for three different datasets such as Corel-lK, Caltech, and 102flower. It is observed that the classification result of the LDP_RGB fusion outperforms the results of LBP_RGB, CSLBP_RGB, and RGB. Moreover, the proposed fusion features preserve more class-oriented properties, so that the retrieval rate is enhanced. The performance analysis for the top 80 images retrieval using the proposed auto-encoder-based feature fusion and the auto-encoder-based RGB channel feature are shown in the result section. The retrieval rate using LDP_RGB fusion also surpasses all the other methods discussed.

The major contributions of this work are mentioned below:

- A new enhanced feature fusion technique is used, where the texture image and RGB channel image features are fused.
- The auto-encoding features of the texture and RGB channels are extracted by two different CNN models to save the low variance pixel information of the texture image.
- The CBIR model is tested for three different textures i.e. LBP, CSLBP, and LDP textures with RGB channel encoding feature.
- The model is tested with three different datasets such as Corel-lK, Caltech, and 102flower.
- Improvement in the f1-score implies the proposed feature descriptor handles the class imbalance issue more precisely.
- The retrieval result is enhanced with the fusion of LDP and RGB encoded features.

The rest part of the paper is arranged in the following order: Section II presents a review of feature fusion and CBIR system. Section III shows the proposed feature fusion model, CNN encoding architecture, and performance evaluation metrics. The detailed results are shown in Section IV i.e. the results and discussions. The conclusion of the work is presented in Section V.

II. Related Work

Kayhan, N., et al. [1] build a weighted feature-based CBIR system using modified local binary patterns (MLBP), local neighbourhood differences patterns (LNDP), filtered gray level co-occurrence matrix (GLCM), and the quantization color histogram features. Khan, U. A., et al. [2] used hybrid classification model using three color moments, Haar Wavelet, Daubechies Wavelet and Bi-Orthogonal wavelets features. They have used genetic algorithm (GA) and SVM classification and L2 Norm is used for the similarity measure. Kashif, M., et al. [3] proposed a hybrid image descriptor using local ternary pattern, local phase quantization, and discrete wavelet transform. They used joint mutual information (JMI) based feature selection to derive the optimal feature for effective image retrieval. Carvalho, E. D., et al. [4] proposed a histopathological breast image classification model using phylogenetic diversity indexes. They have also used the phylogenetic diversity indexes to rank the images. Authors claim, it outperforms XGBoost, random forest, and support vector machine. Choe, J., et al. [5] proposed a medical image retrieval model for interstitial lung disease diagnosis using the deep learning features of CT images. Pradhan, J., et al. [7] proposed a regions-of-attention-based feature fusion technique for image retrieval, here authors used multi-directional texture features with spatial correlation-based color features to derive the image semantics. Pathak, D., et al. [9] proposed a retrieval system by concatenating the deep learning GoogleNet features with the hue, saturation, and intensity features of the HIS image, and Histogram of orientated gradient (HOG) feature of the RGB image. Here the authors claim this technique is used to reduce information loss due to image resizing. A. Latif, et al. [10] presented a comprehensive review of the recent development and the state-of-the-art CBIR systems. The study explored the major concepts of CBIR like image representation, image retrieval, low-level feature extraction, and recently used semantic deep-learning approaches, it also includes future research directions in CBIR. M. Sotoodeh, et al. [17] presented a local texture descriptor referred as Color Radial Mean Local Binary Pattern (CRMLBP). The CRMLBP is computed for the sign-difference, magnitude-difference, and central gray value patterns in the RGB color space and their histograms are concatenated. The feature weights are optimized using Particle Swam Optimization (PSO) technique. The performance of this feature vector is tested with various datasets such as Wang, Holidays, Corel data. Sampathila, N., et al. [18] presented an image retrieval method using Grey-level co-occurrence-based Haralik's features and histogram-based cumulative distribution function (CDF) for the brain MRI image retrieval. Here the KNN approach is used to find the distance between the query image and other images. Khan, M. A., et al. [19] proposed an intelligent human action recognition system using Hand-crafted and deep convolutional neural network features fusion. Here the histogram of oriented gradients (HoG) and deep features are fused. A multi-class support vector machine (M-SVM) is used for the classification. Ma, W., et al. [22] suggested a cloud-based privacy-preserving image retrieval service using deep convolutional features with from the encrypted image. For image encryption, a hybrid encryption method is adopted. Wang, S. H., et al. [23] suggested deep feature fusion technique using graph convolutional network and convolutional neural network features for Covid-19 classification. Here they used the CT images to test their model performance. L. T. Alemu, et al. [25] proposed a multi-feature fusion-based CBIR system, where various hand-crafted features with deep NN features and membership score is applied based on their probabilistic distribution. Then an incremental nearest neighbour (NN) selection is used to implement k-NN for dynamic query selection. Wang, W., et al. [26] presented a two-stage CBIR model using the fusion of global and local feature. Authors use a sparse coding for the sparse representation of the local features followed by feature pooling and the Euclidean distance measure is used to find the similarity between the sparse feature vectors. Bella, M. I. T. et al. [28] proposed the image retrieval system using information fusion technique, where the GLCM and HSV color moment features are fused the model is tested with Corel-1K, Corel-5K, and Corel-10K datasets.

Table I presents a survey on the different feature fusion techniques used for image classification and retrieval. However, there is a scope to define a better image descriptor using the strength of the texture feature with the deep CNN feature. In this work, intend to define a more precise feature vector by combining the CNN-encoded texture feature with the encoded RGB channel feature.

TABLE I. SURVEY TABLE FEATURE FUSION AND CBIR SYSTEM

Sl. No.	Research Study / Year	Feature Fusion Method	Classification / Retrieval	Database
1	H. Wang / 2020 [6]	Visual saliency based multi-feature fusion	Retrieval	Corel-1K
2	Pradhan, J. / 2021 [7]	Texture, and spatial correlation-based color features fusion	Retrieval	Corel, and GHIM
3	K. T. Ahmed /2021 [16]	Spatial color with shaped features fusion	Retrieval	Caltech, Corel, COIL, and ALOT
4	Khan, M. A. /2020 [19]	Hand-crafted, HoG, and deep features fusion	Classification	Weizmann, UCF11, IXMAS
5	Wang, S. H. /2021 [23]	Deep feature fusion	Classification	Chest CT images
6	Wang, W. /2022 [26]	Global and Local features fusion	Retrieval	Coil20, Caltech
7	Wang, Z. / 2019 [27]	Deep features, morphological features, texture, and density feature fusion	Classification	400 mammograms pathologica dataset
8	Bella, M. I. T./2019 [28]	Fused Feature of GLCM and HSV Color Moments features.	Retrieval	Corel-1K, Corel-5K, and Corel-10K

III. PROPOSED MODEL

In the case of Deep learning, the image features are fetched automatically using a CNN encoder. The features extracted from the RGB channel carry more information than the Gray-scale image, as it learns from three channels R, G, and B coherently. At the same time, computational complexity increases. Moreover, information stored in all three channels is highly correlated, which impedes the learning rate. In this work, a feature fusion technique is proposed where the CNN-encoded feature of the texture image is fused with the encoded feature of RGB channels. Two different CNN encoders are used to derive the texture and RGB features from an image. The motivation behind two different CNN encoders instead of adding the texture image in the 4th channel in addition to the R, G, and B is that the range of the pixel values of the texture image is comparatively smaller than the pixel values of the R, G, and B channels. So the texture information will not be suppressed during the recursive MAX pooling and ReLU operations.

A. LBP Texture Image

The LBP texture of a 3 x 3 pixel block is achieved by thresholding the pixel values of the neighbours with its center pixel into binary values, where the value is 1 if the value of the neighbour pixels is greater or equal to the value of the center pixel, otherwise 0. The values of all the 8 neighbours are stored in an unsigned-byte form, here the range varies from 0 to 255. Eq. (1) shows the calculation of the LBP texture image, where R is the radius of the circle [14].

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(g_p - g_c)\, 2^P \qquad (1)$$

$$where\ \ S(z) = \begin{cases} 1, & if\ z \geq 0 \\ 0, & otherwise \end{cases}$$

B. CSLBP Texture Image

Center-Symmetric Local Binary Patterns are produced by computing the thresholding difference of pixel values with their symmetrically opposite pixels with respect to the canter of a pixel block. Here the thresholding difference is a smaller integer value T. The CSLBP labels generate shorter histograms, which is a more stable feature for the flat image regions. Eq. (2) represents the calculation of the CSLBP texture image [15].

$$CSLBP_{P,R,T(x,y)} = \sum_{p=0}^{\left(\frac{P}{2}\right)-1} S\left(g_p - g_{p+\left(\frac{P}{2}\right)} - T\right) 2^P \qquad (2)$$

$$where\ \ S(z) = \begin{cases} 1, & if\ z \geq 0 \\ 0, & otherwise \end{cases}$$

C. LDP Texture Image

The LDP pattern illustrates the response values of all eight directional edges of a center pixel. It is calculated using the Kirsch masks in the eight different orientations i.e. (M0 ~ M7) with respect to a 3x3 pixel block, and Eq. (3) and Eq. (4) show the LDP value calculation at a point (x, y).

$$m_i = \sum_{l=-1}^{1}\sum_{k=-1}^{1} I(x+l, y+k) \times M_i(l,k) \qquad (3)$$

$$LDP_{x,y}(m_{0,} \dots, m_7) = \sum_{i=0}^{7} S(m_i - m_k) \times 2^P \qquad (4)$$

Fig. 1 shows the Original images of Corel-lK, Caltech, and 102Flower datasets with their equivalent LBP, CSLBP, and LDP texture images. Here four images are shown from each dataset and the *Image Ref* is a combination of the folder name and the image name.

D. Auto-encoder-based CNN Feature

The deep CNN feature of an image is generated using an automatic encoding technique. The image feature is learned through batch mode training, hence it is expected that the feature preserves the class information. As the layers of CNN architecture are densely connected, the learning becomes faster with automatic weight adjustment for a particular class using supervised learning.

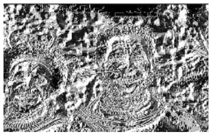
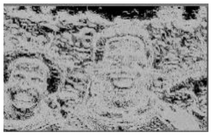
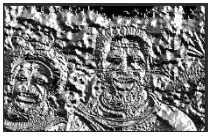
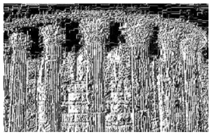
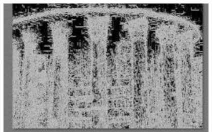
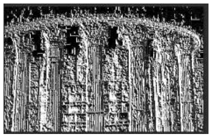
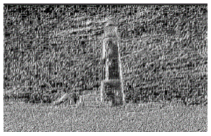
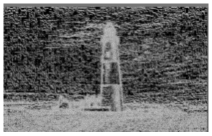
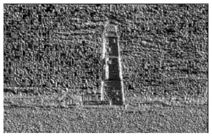
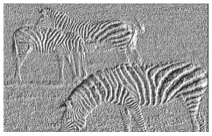
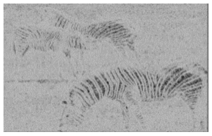
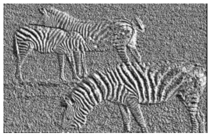
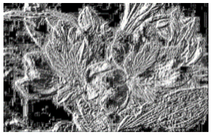
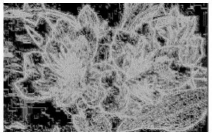
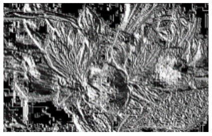
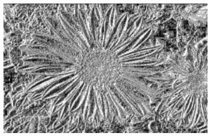
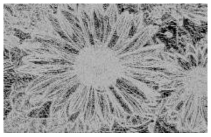
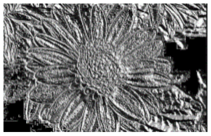
Database	Image Ref.	Original Image	LBP Texture	CSLBP Texture	LDP Texture
Corel Data	1_3				
	3_207				
Caltech Data	132_0024				
	250_0008				
102flower Data	18_image_04247				
	41_image_02204				

Fig. 1. The Original RGB image with their equivalent LBP, CSLBP and LDP texture image.

E. Feature Fusion Model

Fig. 2 shows the proposed model, here the size of the input image is 512 x 512 for both the CNNs i.e. the RBG and the texture input. The CNN architecture consists of seven layers and each layer contains a convolution operation followed by the ReLU and MAX pooling operations. Non-linearity property is introduced to the convolution output with the ReLU activation function. Whereas the image size reduction is done by the MAX pooling with each convolution operation. The flattened layer is used to reduce the image to a single-dimension feature vector of size 1 x 1024. Further dimension reduction is done with four fully connected layers. The soft-max operation is used to calculate the class label from the feature map using the energy function.

$$(I * f)_{x,y} = \sum_{s=1}^{H} \sum_{t=1}^{W} f_{s,t} . I_{x+s-1,y+t-1} + b \quad (5)$$

$$ReLU(x) = Max(0, x) \quad (6)$$

$$E(w, b) = \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(x_i)\big) + \alpha R(w) \quad (7)$$

Where:

L = model loss parameter.

R = regularization factor used to deal with the model complexity.

α = regularization strength control parameter.

The cross-entropy loss is determined as the penalty value in each iteration using that energy function. Eq. (5) represents the convolution operation at point a (x, y) of an image I used the filter f, where the H and W represent the height and width of the image. The ReLU operation is defined using Eq. (6). Eq. (7) represents the regularized training error of an instance. Eq. (8) represents the sigmoid function Si used to map the output value within (0, 1). The cross-entropy loss for each iteration is defined by Eq. (9).

$$S_i(x) = \frac{1}{1+ e^{-x}} \quad (8)$$

$$LogLoss = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} logP_{ij} \quad (9)$$

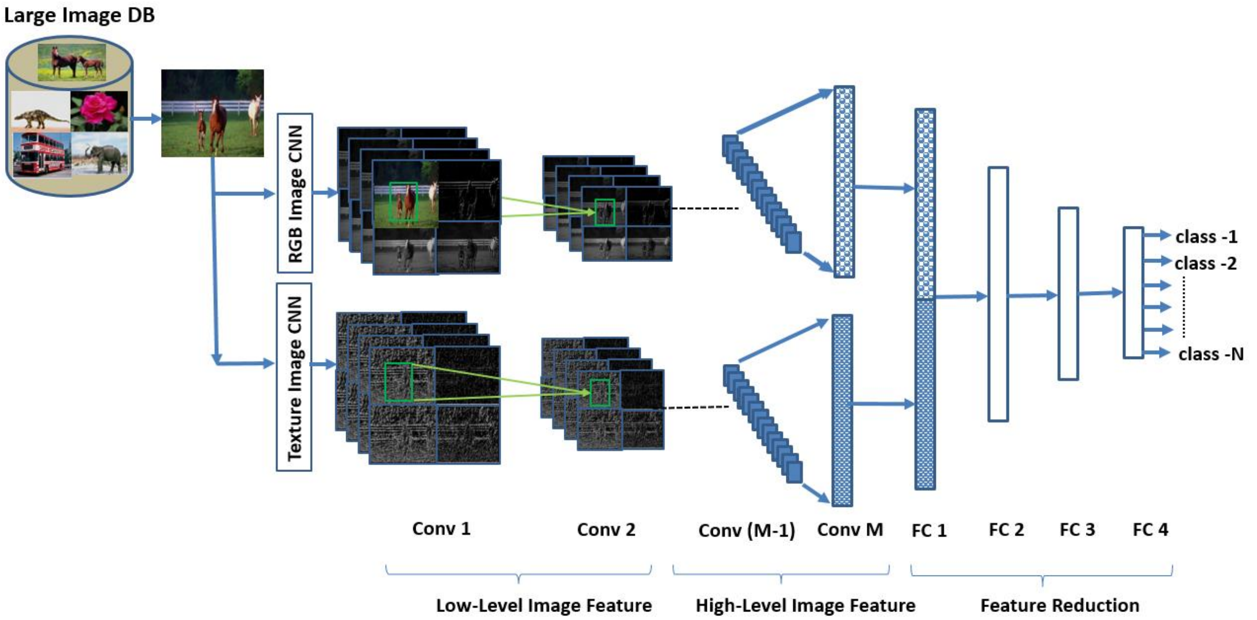


Fig. 2. Block diagram of the proposed CNN encoder-based RGB and texture features fusion.

Where N shows the number of samples and M is the number of labels, the y_{ij} represents if the label j is correctly classified as, for the instance, i. Here P_{ij} is the probability value of the model that assigns label j to the instance i.

The proposed feature fusion model using the CNN feature of the RGB image and the CNN feature of the texture image uses the standard learning rate with an early stop parameter value of 0.99. The model training is done using 80:20 holdout validation. Here a GTX 1650 graphics system with 16 GB RAM is used for the training and testing of the proposed model.

F. Performance Measures

The performance of the proposed feature fusion is evaluated using parametric quantifiers such as precision, recall, and f1-score [13], which are defined below using Eq. (10), Eq. (11), and Eq. (12) respectively.

$$Precision = \frac{True^{+}}{True^{+} + False^{+}} \qquad (10)$$

$$Recall = \frac{True^{+}}{True^{+} + False^{-}} \qquad (11)$$

$$f1 - Score = 2 \times \frac{Precision \times Recall}{Precision + \ Recall} \qquad (12)$$

Here the true positive (True+) value shows the number of images correctly identified into their belonging class by the system. The false positive (False+) shows the number of images falsely recognized by the system, and the false negative (False-) shows the number of images falsely rejected by the system. The precision shows the number of images correctly identified into their belonging class with respect to the total number of images identified by the system. Whereas recall represents the number of images correctly identified for a class with respect to all the images belonging to that class. Hence the average recall value is a significant performance measure for a retrieval system. The Caltech and 102flower datasets have a different number of total images in different classes. The harmonic mean of these classes i.e. f1-Score is also presented in addition to the precision [13, 29]. The receiver operating characteristics (ROC) curve, which is plotted using the true-positive rate vs. false-positive rate, illustrates graphically the classifier's performance. The City-block distance measure shown in the Equation (13) is used to measure the similarity between the images.

City-block distance measure:

$$D_{CT} = \sum_{i=0}^{L-1} \left|F_i^q - F_i^t\right| \qquad (13)$$

Where:

F_i^q = feature vector of the query image

F_i^t = feature vector of the database image

IV. Results and Discussion

The results of the proposed CBIR model using encoded texture feature fusion are discussed in this section. Here the CNN-based auto-encoding features of the RGB channels are fused with the auto-encoding features of three different texture features i.e. LBP, CSLBP, and LDP. The image retrieval model is tested with three different datasets such as Corel-lK, Caltech, and 102Flower. To avoid the extensive processing time, selective 15 classes of the 102Flower dataset have been considered. The precision, recall, and f1-score of each class are presented for all three datasets. The ROC curve shows the overall classification performance using the CNN encoding features of RGB, RGB_LBP, RGB_CSLBP, and RGB_LDP. The average retrieval performance is shown separately for all three datasets using all the above-discussed four encoding features for top 80 image retrieval. The class-wise retrieval performances are illustrated with the bar graph for all four encoding features. The detailed analysis results of individual classes are discussed in the sub-sections below.

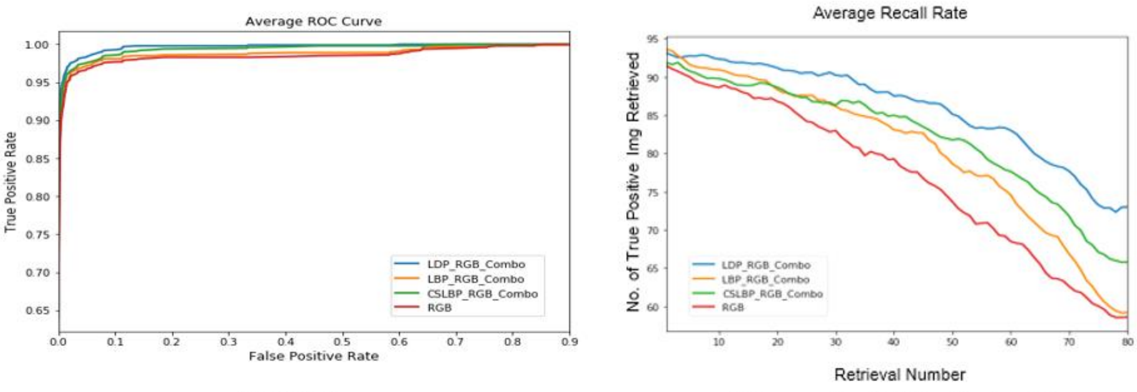
A. Results Analysis of Corel-lK

Table II shows the performance analysis of the Corel-1K dataset maximum average precision value is 94.5% using the LDP_RGB encoder feature. It is 94.2% using LBP with RGB, 94.3% for CSLBP with RGB, and 94.2% using the RGB encoder feature. The classification performance is presented using the ROC curve in Fig. 3(a).

The Recall and f1-score are 94.4%, and 94.5% respectively using the LDP_RGB feature, which is maximum in comparison to the other features. Though there is a small difference in the classification rate, the average retrieval rate is significantly enhanced using the LDP_RGB feature in comparison to the other features for retrieving the top 80 images shown in Fig. 3(b), and the class-wise retrieval analysis is shown in Fig. 4 for the top 10 images. In this dataset, each class consists of 100 images, so there is no major difference in the precision and f1-score values.

TABLE II. CLASS-WISE PERFORMANCE OF THE COREL-1K DATASET

Class Name	**LDP_RGB_Combo**			**LBP_RGB_Combo**			**CSLBP_RGB_Combo**			**RGB**			**No. of Images**
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	
African Tribes	0.94	0.9	0.92	0.92	0.91	0.91	0.88	0.96	0.92	0.87	0.95	0.91	100
Beaches	0.93	0.91	0.92	0.94	0.89	0.91	0.92	0.89	0.9	0.84	0.95	0.89	100
Buildings	0.88	0.87	0.87	0.97	0.87	0.92	0.93	0.89	0.91	0.96	0.85	0.9	100
Buses	0.99	0.97	0.98	0.97	0.96	0.96	0.99	0.97	0.98	0.97	0.97	0.97	100
Dinosaurs	0.99	1	1	1	0.97	0.98	1	1	1	1	1	1	100
Elephants	0.98	0.97	0.97	0.95	0.95	0.95	0.94	0.9	0.92	0.93	0.92	0.92	100
Flowers	0.99	1	1	0.96	1	0.98	1	1	1	1	1	1	100
Horses	0.99	0.94	0.96	0.95	0.99	0.97	0.97	1	0.99	0.98	1	0.99	100
Mountains	0.85	0.94	0.9	0.86	0.95	0.9	0.83	0.91	0.87	0.92	0.87	0.89	100
Foods	0.91	0.94	0.93	0.9	0.92	0.91	0.97	0.89	0.93	0.95	0.88	0.91	100
Avg. Accuracy %	**94.5**	**94.4**	**94.5**	94.2	94.1	93.9	94.3	94.1	94.2	94.2	93.9	93.8	



(a) ROC Curve

(b) Average Recall Rate of the top 80 Image

Fig. 3. Analysis of the Corel 1K images using RGB and Texture with RGB CNN feature.

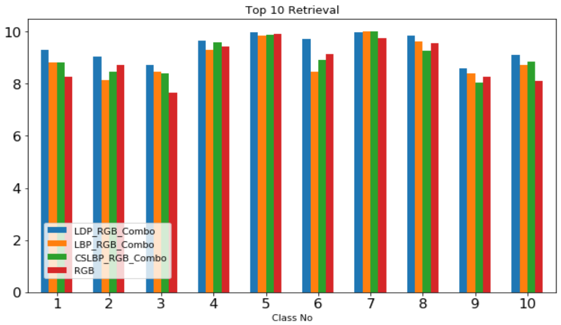
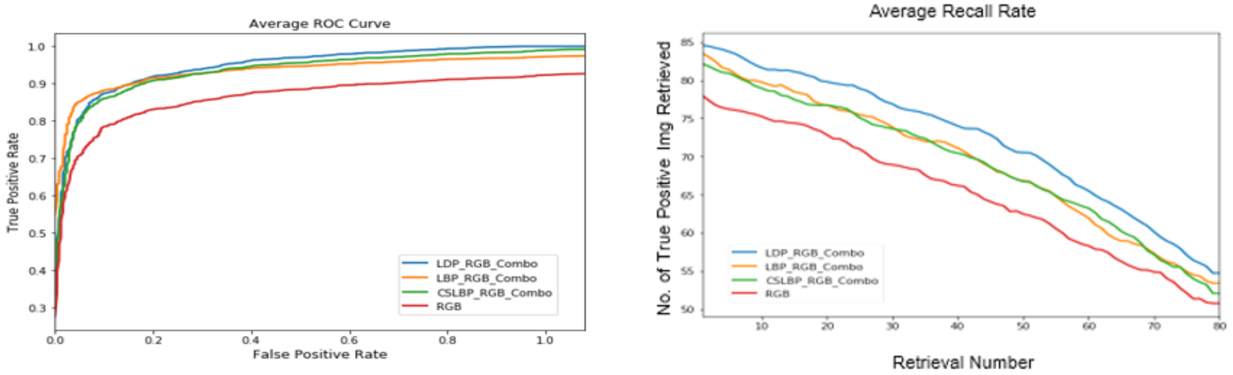


Fig. 4. Class-wise retrieval of the top 10 images of the Corel 1K dataset.

TABLE III. CLASS-WISE PERFORMANCE OF THE CALTECH DATASET

Class Name	LDP_RGB_Combo			LBP_RGB_Combo			CSLBP_RGB_Combo			RGB			No. of Images
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	
Backpack	0.88	0.93	0.9	0.88	0.91	0.9	0.84	0.94	0.88	0.98	0.86	0.92	151
Billiards	0.91	0.88	0.9	0.92	0.93	0.92	0.84	0.91	0.88	0.91	0.81	0.86	278
Bonsai-101	0.9	0.87	0.88	0.9	0.89	0.89	0.96	0.85	0.9	0.98	0.82	0.89	122
Boxing-glove	0.86	0.97	0.91	0.94	0.89	0.91	0.81	0.94	0.87	0.87	0.9	0.89	124
Eiffel-tower	0.88	0.9	0.89	0.91	0.89	0.9	0.84	0.92	0.87	0.9	0.8	0.85	83
Fern	0.87	0.94	0.9	0.9	0.89	0.89	0.91	0.95	0.93	0.89	0.85	0.87	110
Fighter-jet	0.83	0.84	0.83	0.85	0.88	0.87	0.89	0.81	0.85	0.79	0.78	0.79	99
Fire-truck	0.94	0.87	0.91	0.87	0.85	0.86	0.97	0.6	0.74	0.98	0.89	0.93	118
Gorilla	0.93	0.89	0.91	0.85	0.9	0.87	0.89	0.9	0.89	0.89	0.9	0.9	212
Iris	0.81	0.85	0.83	0.79	0.85	0.82	0.91	0.86	0.89	0.56	0.91	0.69	108
Light-house	0.86	0.9	0.88	0.89	0.88	0.88	0.93	0.81	0.87	0.83	0.92	0.87	190
Sunflower-101	0.99	0.95	0.97	0.97	0.95	0.96	0.97	0.95	0.96	0.99	0.93	0.95	80
Watch-101	0.95	0.93	0.94	0.95	0.93	0.94	0.91	0.9	0.9	0.81	0.95	0.87	201
Waterfall	0.91	0.87	0.89	0.81	0.83	0.82	0.87	0.81	0.84	0.95	0.79	0.86	95
Zebra	0.93	0.85	0.89	0.91	0.81	0.86	0.71	0.94	0.81	1	0.84	0.92	96
Avg. Accuracy%	**89.7**	**89.6**	**89.5**	88.9	88.5	88.6	88.3	87.3	87.2	88.9	86.3	87.1	



(a) ROC Curve

(b) Average Recall Rate of the top 80 Image

Fig. 5. Analysis of the Caltech images using RGB and texture with RGB CNN feature.

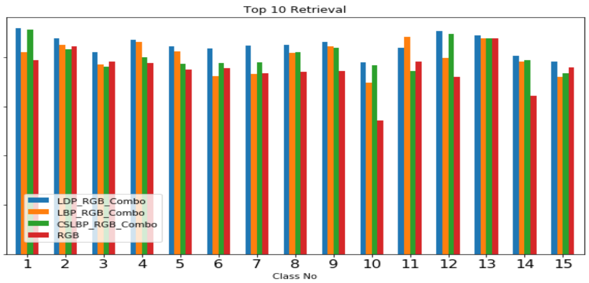


Fig. 6. Class-wise retrieval of the top 10 images Caltech dataset.

B. *Results Analysis of Caltech*

The result analysis of the Caltech dataset is shown in Table III, in this case, value of maximum average precision is 89.7% using the LDP_RGB encoder feature. Whereas it is 88.9% using LBP with RGB, 88.3% for CSLBP with RGB, and 88.9% using RGB encoder feature. The classification performance is presented using the ROC curve in Fig. 5(a). The Recall and f1-score are 89.6%, and 89.5% respectively with the LDP_RGB feature, which is more in comparison to the other encoding features.

Though there is a visible difference in the classification rate, the average retrieval rate is also enhanced using the LDP_RGB feature in comparison to the other features for retrieving the top 80 images shown in Fig. 5(b), and Fig. 6 show the class-wise retrieval analysis for the top 10 images. As in this dataset, the number of images in the different classes varies the precision, and f1-score values are also different, moreover these values are more stable using the LDP_RGB encoder feature.

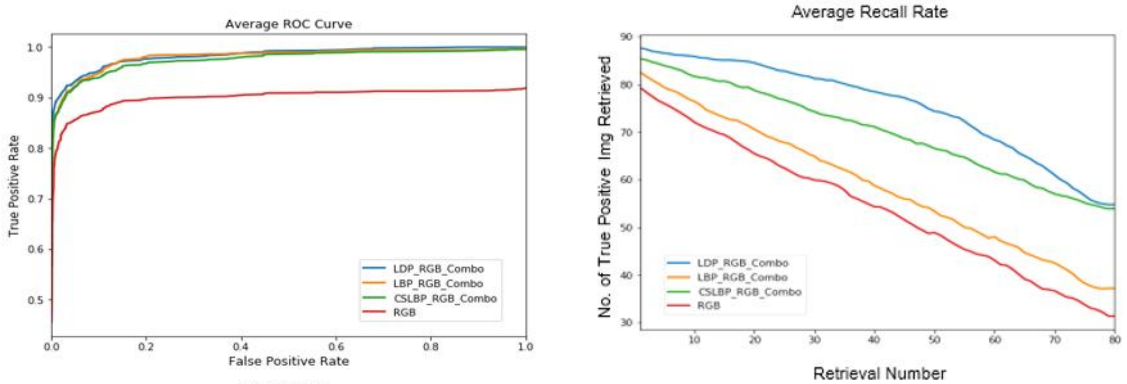
C. Results Analysis of 102Flower

The classification performance of the 102Flower dataset is shown in Table IV. Here 15 classes are selected, and the name of the flower and the number of images available for each class is mentioned in the table in the first and last columns respectively. Here the value of maximum average precision is 88.7% using the LDP_RGB encoder feature. Whereas it is 87.9% using the LBP with RGB, 85.6% for CSLBP with RGB, and 84.5% using the RGB encoder feature. The classification performance of the 102Flower dataset is presented using the ROC curve in Fig. 7(a).

There is a significant difference in the classification rate. The result shown in Fig. 7(b) claims that average retrieval rate using LDP_RGB feature is better than other fusion techniques for retrieving the top 80 images. Fig. 8 shows the class-wise retrieval of the top 10 images. The value of the f1-score is also enhanced using the LDP_RGB feature. Table V presents a state-of-art, where the performance of five other works available in the literature, using the same dataset are compared with the proposed feature fusion model. It shows the accuracy of Corel-1k is 94.5% and Caltech256 is 89.7%. A significant improvement is achieved for both datasets using the proposed feature fusion model.

TABLE IV. CLASS-WISE PERFORMANCE OF THE CALTECH DATASET

Class Name	**LDP_RGB_Combo**			**LBP_RGB_Combo**			**CSLBP_RGB_Combo**			**RGB**			**No. of Images**
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	
Purple Coneflower	0.99	0.88	0.93	0.95	0.9	0.93	0.9	0.83	0.86	0.92	0.8	0.85	84
Peruvian Lily	0.94	0.83	0.88	0.76	0.81	0.79	0.86	0.83	0.84	0.9	0.7	0.79	81
Cape Flower	0.85	0.88	0.86	0.86	0.85	0.85	0.87	0.82	0.84	0.9	0.83	0.86	106
Barbeton Daisy	0.88	0.84	0.86	0.86	0.85	0.85	0.78	0.86	0.82	0.94	0.85	0.89	126
Sword Lily	0.79	0.84	0.82	0.84	0.84	0.84	0.92	0.84	0.87	0.8	0.8	0.8	129
Pink-Yellow Dahlia	0.92	0.9	0.91	0.85	0.98	0.91	0.8	0.83	0.81	0.74	0.81	0.77	108
Californian Poppy	0.94	1	0.97	0.91	0.99	0.95	0.87	0.87	0.87	0.76	0.96	0.85	101
Azalea	0.85	0.89	0.87	0.84	0.85	0.84	0.96	0.85	0.91	0.79	0.85	0.82	95
Rose	0.93	0.87	0.9	0.99	0.84	0.9	0.75	0.89	0.81	0.75	0.84	0.79	170
Lotus	0.87	0.93	0.9	0.95	0.9	0.92	0.88	0.88	0.88	0.79	0.9	0.84	136
Anthurium	0.8	0.83	0.82	0.77	0.84	0.8	0.71	0.84	0.77	0.84	0.86	0.85	104
Frangipani	0.96	0.95	0.95	0.94	0.97	0.95	0.88	0.84	0.86	0.82	0.88	0.85	165
Hibiscus	0.8	0.87	0.83	0.86	0.85	0.85	0.9	0.82	0.85	0.85	0.8	0.83	130
Cyclamen	0.9	0.86	0.88	0.92	0.88	0.9	0.93	0.84	0.88	0.95	0.82	0.88	152
Foxglove	0.89	0.89	0.89	0.88	0.88	0.88	0.83	0.84	0.84	0.93	0.82	0.87	161
Avg. Accuracy%	**88.7**	**88.4**	**88.5**	87.9	88.2	87.7	85.6	84.5	84.7	84.5	83.5	83.6	



(a) ROC Curve

(b) Average Recall Rate of the top 80 Image

Fig. 7. Analysis of the 102Flowerset images using RGB and Texture with RGB CNN feature.

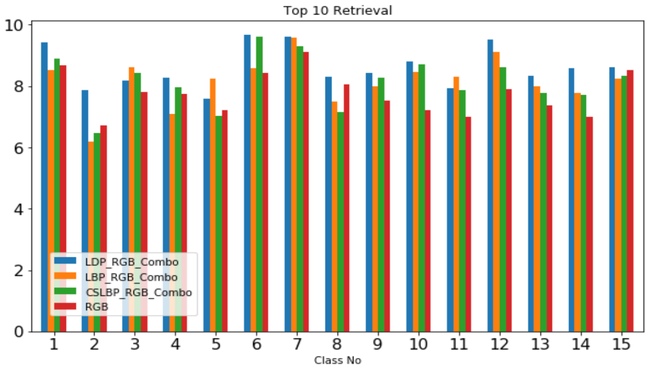


Fig. 8. Class-wise retrieval of the top 10 images of the 102 flowerset dataset.

TABLE V. COMPARISON OF CLASSIFICATION RESULTS OF EXISTING METHODS AND THE PROPOSED METHOD

References	Dataset	Methods Used	Accuracy%
Kayhan, N.[1]	Corel-lK	Weighted Color & Texture feature fusion	82.52%
Khan, U. A.[2]	Corel-lK	Wavelet features with GA and SVM	90.5%
K. T. Ahmed [16]	Corel-lK	Color and Object feature fusion	92.3%
K. T. Ahmed [16]	Caltech	Color and Object feature fusion	71.3.%
M. I. Thusnavis Bella [28]	Corel-lK	HSV Color and GLCM feature	83.3%
ElAlami, M. E. [30]	Wang-lK	GLCM feature using ANN	76.1%
Proposed Model	Corel-lK	LDP and RGB encoded feature	94.5%
Proposed Model	Caltech	LDP and RGB encoded feature	89.7%

V. CONCLUSION

This paper proposed a CBIR model using the feature fusion technique. Here the CNN-encoded features of the image are fused with the encoded features of the RGB image. As the range of pixel values in the texture image is comparatively smaller than that of the RGB image, two different encoders are employed to extract the CNN features separately. These two features are fused to define a more significant image descriptor.

The proposed model is tested with three public datasets i.e. Corel-lK, Caltech, and 102flower. It is observed that the classification performance is improved by the proposed feature fusion model as compared to the RGB channel encoding feature. The result shows the performance of the LDP with RGB feature fusion is better with respect to the LBP with RGB and CSLBP with RGB features. There is a significant improvement in the retrieval system for the top 10 as well as top 80 image retrieval. Moreover, the enhancement of the f1-score using the proposed feature fusion technique illustrates the class property is better retained using the fused features. The f1-score value improved significantly using the encoder-based LDP with RGB feature fusion for the dataset having class imbalance issues such as Caltech, and 102flower. In future, the model can be tested using the fusion of other textures like LTP, GLCM.

REFERENCES

[1] Kayhan, N., & Fekri-Ershad, S. (2021). Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns. *Multimedia Tools and Applications*, *80*(21), 32763-32790.

[2] 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.

[3] Kashif, M., Raja, G., & Shaukat, F. (2020). An efficient content-based image retrieval system for the diagnosis of lung diseases. *Journal of digital imaging,* 33(4), 971-987.

[4] Carvalho, E. D., Antonio Filho, O. C., Silva, R. R., Araujo, F. H., Diniz, J. O., Silva, A. C., ... & Gattass, M. (2020). Breast cancer diagnosis from histopathological images using textural features and CBIR. *Artificial intelligence in medicine,* 105, 101845.

[5] Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M. J., ... & Kim, B. (2022). Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology,* 302(1), 187-197.

[6] H. Wang, Z. Li, Y. Li, B. B. Gupta, and C. Choi, "Visual saliency guided complex image retrieval," *Pattern Recognition Lett.*, vol. 130, pp. 64–72, 2020, doi: 10.1016/j.patrec.2018.08.010.

[7] Pradhan, J., Pal, A. K., Banka, H., & Dansena, P. (2021). Fusion of region based extracted features for instance-and class-based CBIR applications. *Applied Soft Computing,* 102, 107063.

[8] Salih, S. F., & Abdulla, A. A. (2022). An effective bi-layer content-based image retrieval technique. *The Journal of Supercomputing,* 1-24.

[9] Pathak, D., & Raju, U. S. N. (2022). Content-based image retrieval for super resolutioned images using feature fusion: Deep learning and hand crafted. *Concurrency and Computation: Practice and Experience,* e6851.

[10] A. Latif et al., "Content-based image retrieval and feature extraction: A comprehensive review," *Math. Probl. Eng.,* vol. 2019, 2019, doi: 10.1155/2019/9658350.

[11] W. Yu, K. Yang, H. Yao, X. Sun, and P. Xu, "Exploiting the complementary strengths of multi-layer CNN features for image," *Neurocomputing,* vol. 237, pp. 235–241, 2017, doi: 10.1016/j.neucom.2016.12.002.

[12] G. M. Galshetwar, L. M. Waghmare, A. B. Gonde, and S. Murala, "Local energy oriented pattern for image indexing and retrieval," *J. Vis.*

Commun. Image Represent., vol. 64, p. 102615, 2019, doi: 10.1016/j.jvcir.2019.102615.

[13] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, “Medical image retrieval using deep convolutional neural network,” *Neurocomputing*, vol. 266, pp. 8–20, 2017, doi: 10.1016/j.neucom.2017.05.025.

[14] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, “A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval,” *Expert Syst. Appl.*, vol. 100, pp. 224–233, 2018, doi: 10.1016/j.eswa.2018.01.056.

[15] Chakraborty, S., Singh, S. K., & Chakraborty, P. (2019). R-theta local neighborhood pattern for unconstrained facial image recognition and retrieval. *Multimedia Tools and Applications,* 78(11), 14799-14822.

[16] K. T. Ahmed, S. Ummesafi, and A. Iqbal, “Content based image retrieval using image features information fusion,” *Inf. Fusion*, vol. 51, pp. 76–99, 2019, doi: 10.1016/j.inffus.2018.11.004.

[17] M. Sotoodeh, M. R. Moosavi, and R. Boostani, “A novel adaptive LBP-based descriptor for color image retrieval,” *Expert Syst. Appl.*, vol. 127, pp. 342–352, 2019, doi: 10.1016/j.eswa.2019.03.020.

[18] Sampathila, N., & Martis, R. J. (2022). Computational approach for content-based image retrieval of K-similar images from brain MR image database. *Expert Systems*, 39(7), e12652.

[19] Khan, M. A., Sharif, M., Akram, T., Raza, M., Saba, T., & Rehman, A. (2020). Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Applied Soft Computing,* 87, 105986.

[20] A. Alzu’bi, A. Amira, and N. Ramzan, “Content-based image retrieval with compact deep convolutional features,” *Neurocomputing*, vol. 249, pp. 95–105, 2017, doi: 10.1016/j.neucom.2017.03.072.

[21] S. Pang, M. A. Orgun, and Z. Yu, “A novel biomedical image indexing and retrieval system via deep preference learning,” *Comput. Methods Programs Biomed.*, vol. 158, pp. 53–69, 2018, doi: 10.1016/j.cmpb.2018.02.003.

[22] Ma, W., Zhou, T., Qin, J., Xiang, X., Tan, Y., & Cai, Z. (2022). A privacy-preserving content-based image retrieval method based on deep learning in cloud computing. *Expert Systems with Applications,* 117508

[23] Wang, S. H., Govindaraj, V. V., Górriz, J. M., Zhang, X., & Zhang, Y. D. (2021). Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion*, 67, 208-229.

[24] K. R. Kruthika, Rajeswari, and H. D. Maheshappa, “CBIR system using Capsule Networks and 3D CNN for Alzheimer’s disease diagnosis,” *Informatics Med. Unlocked*, vol. 14, pp. 59–68, 2019, doi: 10.1016/j.imu.2018.12.001.

[25] L. T. Alemu and M. Pelillo, “Multi-feature fusion for image retrieval using constrained dominant sets,” *Image Vis. Comput.*, vol. 94, p. 103862, 2020, doi: 10.1016/j.imavis.2019.103862.

[26] Wang, W., Jiao, P., Liu, H., Ma, X., & Shang, Z. (2022). Two-stage content based image retrieval using sparse representation and feature fusion. *Multimedia Tools and Applications*, 81(12), 16621-16644.

[27] Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*, 7, 105146-105158

[28] Bella, M. I. T., & Vasuki, A. (2019). An efficient image retrieval framework using fused information feature. *Computers & Electrical Engineering*, *75*, 46-60.

[29] Jena, P. K., Khuntia, B., Palai, C., Nayak, M., Mishra, T. K., & Mohanty, S. N. (2023). A Novel Approach for Diabetic Retinopathy Screening Using Asymmetric Deep Learning Features. *Big Data and Cognitive Computing*, *7*(1), 25.

[30] ElAlami, M. E. (2014). A new matching strategy for content based image retrieval system. *Applied Soft Computing*, *14*, 407-418.