An Extreme Learning Machine-Relevance Feedback Framework for Enhancing the Accuracy of a Hybrid Image Retrieval System

B. Shikha[1]*, P. Gitanjali[2], D. Pawan Kumar[2]

[1] Department of ECE, DCRUST, Murthal & UIET, Kurukshetra University, Haryana (India)
[2] Department of ECE, DCRUST, Murthal, Sonepat, Haryana (India)

unir

LA UNIVERSIDAD
EN INTERNET

Abstract

The process of searching, indexing and retrieving images from a massive database is a challenging task and the solution to these problems is an efficient image retrieval system. In this paper, a unique hybrid Content-based image retrieval system is proposed where different attributes of an image like texture, color and shape are extracted by using Gray level co-occurrence matrix (GLCM), color moment and various region props procedure respectively. A hybrid feature matrix or vector (HFV) is formed by an integration of feature vectors belonging to three individual visual attributes. This HFV is given as an input to an Extreme learning machine (ELM) classifier which is based on a solitary hidden layer of neurons and also is a type of feed-forward neural system. ELM performs efficient class prediction of the query image based on the pre-trained data. Lastly, to capture the high level human semantic information, Relevance feedback (RF) is utilized to retrain or reformulate the training of ELM. The advantage of the proposed system is that a combination of an ELM-RF framework leads to an evolution of a modified learning and intelligent classification system. To measure the efficiency of the proposed system, various parameters like Precision, Recall and Accuracy are evaluated. Average precision of 93.05%, 81.03%, 75.8% and 90.14% is obtained respectively on Corel-1K, Corel-5K, Corel-10K and GHIM-10 benchmark datasets. The experimental analysis portrays that the implemented technique outmatches many state-of-the-art related approaches depicting varied hybrid CBIR system.

Keywords

I. Introduction

WITH the tremendous advancements in the domain of smart-phones and various image capturing devices, there has been a great revolution in the field of image processing. Nowadays, various social media platforms are more-and-more utilized for sharing these captured images. Hence, it has led to the creation of many massive image repositories [1]. The credit for an enhancement in digital images also owes to different advanced satellite and industrial based cameras, usage of the web, divergent portable devices which are used for image capturing and storing. In these gigantic databases or repositories, the tasks of exploring, browsing, indexing and retrieving are too strenuous. Manual searching and retrieval of these images from massive databases is very time consuming and is also prone to human delusions [2]. Therefore, for the trouble-free management and retrieval of digital images from these huge databases, an efficient system is required and the key to this problem is Content-based Image retrieval (CBIR). This system performs an effective image retrieval based on various image attributes like; texture, shape, color, spatial information, etc.

Generally, an image retrieval process is divided into two forms:

Text-based and Content-based. Images are retrieved on the basis of text annotations in a text-based retrieval system. But, it suffers from many disadvantages like human annotation errors and moreover the usage of synonyms, homonyms, etc. lead to an inaccurate image retrieval [3].

In order to overcome these limitations, the best solution is CBIR systems. In these systems, feature vectors are created for each type of extracted feature which in turn represents the different attributes of an image. Similar feature vectors are created for the complete database also. Lastly, similarity matching is done based on the obtained feature matrix of the given input image and feature vectors obtained from the total images in the repository and afterward final results are achieved. Therefore, for the evolvement of a highly effective CBIR system, a low dimensional feature vector is one of the prime requirements.

A. Motivation

In literature, numerous retrieval systems, particularly for images, have been developed which are focused entirely on the single feature of an image like color, texture and shape. But, these systems are incapable of describing the images with complex appearance. Therefore, for the representation of complex images, an ideal combination of features is required which are finally transformed into a precise and single feature vector. This single feature vector contains additional comprehensive information related to an image and works magnificently as compared to the feature vector based on a solitary technique. Therefore, to

* Corresponding author.

E-mail address: shikpank@yahoo.com

represent the complicated images, a perfect amalgamation of basic image features (color, texture and shape) is required. Classification of massive datasets is also a desired and imperative process in order to obtain precise and accurate results. Different machine learning based classifiers have been used significantly for this purpose. But, they lack the features of a neural-network-based system. In the present era, the use of Deep-learning has a significant contribution in the image processing domain. The working of these deep structures depends primarily on the total number of hidden layers as well as the number of neurons in those layers. So, the drawbacks of these customary classifiers can be removed by using a classifier based on a neural network. This feed-forward single hidden layer classifier encapsulates the capabilities of both Deep learning by using hidden layer neurons and of machine learning by performing a more accurate classification task. Finally, to remove the gap between the basic image features and lofty human perception, particularly called as Semantic gap, Intelligent techniques are required. So, in this implementation, our main objective is the formation of a hybrid CBIR system by using best techniques for color, texture and shape retrieval. Then, the formation of a feature vector based on these combined features and by using a feed-forward single hidden layer neural network for the purpose of an efficient classification. Lastly, utilization of an intelligent technique which captures high-level semantics of an image.

B. Related State-of-the-Art Work

In the past decade, many techniques have been used in CBIR systems to extract the features of an image. CBIR is an efficient system which utilizes features of an image and thereby retrieves the required results by using these extracted features. Extraction of features is indeed the foremost phase in a majority of CBIR systems. General features like color, shape, texture, spatial information etc. can be extracted during this extraction process. However, the extraction of multiple or hybrid features is the demand of the current era. Among these hybrid features, color is the most distinctive cue that humans can visualize very promptly. Texture describes the discernible patterns of an image. Among the prominent features of an image, the shape is also considered as an important feature that describes the characteristics of an image. Fadaei et al. [2] propose a hybrid retrieval system which utilizes only color and texture attributes. For the extraction of color features, Dominant color descriptor (DCD) has been used. In order to extract texture features, wavelet and Curvelet features have been utilized. Finally, Particle swarm optimization (PSO) has been used which selects the optimum features of these techniques for an ideal combination.

Fusion of color and shape features have also been utilized in research. Color coherence vector (CCV) has been used as a color extraction technique. Various shape parameters have been used to define and extract the connected parts of the region or boundary of an image [3]. Numerous hybrid systems have been deployed in literature like that of Pavitra et al. [4] who develop a fusion based technique in which Color moment is utilized for the extraction of color features as well as a filter to select some specified images based on the value of a moment. Further, Local binary pattern (LBP) and Canny edge detector have been used to denote texture and edge information of an image.

Another hybrid system which utilizes the memetic algorithm has also been developed. In this system, color is extracted using a basic RGB color space model. For the shape analysis, the median filter has been used and finally, Gray level co-occurrence matrix (GLCM) has been employed to extract features related to texture [5]. For similarity calculation, memetic algorithm, which is the combination of Genetic and Great deluge algorithms, has been used. One more fusion system in which Statistical moments and 2D histograms have been used for color extraction and texture analysis has been carried out by using GLCM [6].

The basic necessity of any CBIR system is the brilliant selection of image parameters needed for extraction of features [7]. Color and edge directivity descriptor (CEDD) is utilized for the extraction of both color as well as texture features while a 2nd level of Discrete wavelet transform (DWT) is employed for the analysis of shape features. Lastly, in order to classify the images, Support vector machine (SVM) classifier has been used. A CBIR system developed in different levels of the hierarchy has been developed by Pradhan et al. [1]. Here, adaptive tetrolet transform is used to extract the textual information. Edge joint histogram and color channel correlation histogram have been used respectively to analyze shape and color features related to an image. This system is realized in the form of a three-level hierarchical system where the highest feature among the three is depicted at every level of the hierarchy.

There are primarily two main domains in which features of an image can be classified. One is the frequency domain and another is the spatial domain. A CBIR system associated with features pertaining to these two domains has been developed by Mistry et al. [8]. In this system, various spatial domain techniques like color auto-correlogram, HSV histogram, and color moments have been used. Under the frequency domain, methods like Gabor wavelet transform and SWT moments have been employed. Also, binarized statistical image features in combination with color and edge directivity descriptor have been used for further enhancing the performance of the system.

Indexing is considered as a prominent technique in order to lessen the memory requirements and to save the execution time. A hybrid CBIR system based on the technique of indexing is developed by Guo et al. [9]. This system utilizes Ordered-Dither Block Truncation Coding (ODBTC) in order to index color images. Color distribution and contrast of an image is represented by Color co-occurrence feature (CCF) and information regarding edges is given by Bit pattern feature (BPF).

Prashant et al. [10] develop a hybrid retrieval system which uses Local binary pattern (LBP) as a texture extraction technique. Here, the LBP calculation is based on the different scales which captures more prominent features of an image as compared to single-scale LBP. Lastly, Gray level co-occurrence matrix (GLCM) has been used efficiently for the computation of feature vectors. Chandan Singh et al. [11] describe a CBIR system where the Color histogram is utilized as a color descriptor. A Color histogram is a graphical characterization of pixels in an image. In addition to Color histogram, Block variation of local correlation coefficients (BVLC) and Block difference of Inverse Probabilities (BDIP) are adopted for texture extraction.

Dominant color descriptor (DCD) is considered as a vital color descriptor used in CBIR where course partitions are created by the division of a space utilized for color analysis. Each partition has a partition center and its percentage. Integration of DCD, Gray level co-occurrence matrix (GLCM) and Fourier descriptors is deployed for the constitution of a hybrid CBIR system [12]. Relevance feedback is also considered as an important intelligent technique which retrieves relevant images of interest based on the feedback obtained by the user. A hybrid system based on a single iteration of relevance feedback with the utilization of Non-dominated Sorting Genetic Algorithm with an exploitation algorithm has been designed by Miguel et al. [13].

Relevance feedback can also be used in combination with many meta-heuristics. Sequential forward selector (SFS) meta-heuristic is utilized in combination with relevance feedback using a single iteration. Many distance metrics have also been tested and analyzed here [14].

A color image retrieval system has been developed in two levels by using Color moment, Edge histogram descriptor (EHD) and Angular radial transform (ART) specifically for the extraction of color, texture and shape attributes of an image respectively. Color moment has been

used at the first level of the retrieval process and then followed by texture and shape descriptors in the second level [15]. An image retrieval system which utilizes only color and shape features by using RGB color model space and Edge directions has also been described [16].

Color co-occurrence matrix (CCM) computes the probability of occurrence of a pixel amongst a specific pixel and its neighbors which in turn is depicted as the extracted feature of an image. Difference between pixels of scan pattern (DBPSP) is based on the difference calculation between two pixels and its conversion into probability. These techniques, CCM and DBPSP, have been used for the extraction and analysis of color and texture attributes and finally, a hybrid system is formed [17]. Multiple features can be combined efficiently for the evolution of an effective image retrieval system. Pandey et al. [18] describe a retrieval system where Bi-cubic interpolation (BCI) is deployed for an initial pre-processing of an image. For the extraction of color features, color coding (CC) is used. Gray level difference method (GLDM) and Discrete wavelet transform (DWT) have been employed for texture feature extraction. Finally, Hu moments have been analyzed for shape feature extraction.

For the analysis and extraction of texture features Discrete cosine transform (DCT) [19] has also been used. Discrete wavelet transform has also been added to enhance the efficiency of the system. Then, the difference between these two techniques is computed and re-ranking of images is done. Lastly, for the extraction of color attributes from an image, Color coherence vector (CCV) has been employed. CBIR is also known by the name of Query by image content (QBIC) because the user's query is given in the form of an image. A different type of hybrid retrieval system is proposed in which a combination of shape and texture features is used. For shape extraction, Fourier descriptors have been used and Radial Chebyshev moments [20] have been used to analyze texture features. K-means Clustering has also been used to augment the classification accuracy of the system.

The extraction of features from an image can be done globally and locally. Global feature extraction is based on the whole image while local extraction is devised to be used for a specified region of an image. Based on this global and local feature extraction, a CBIR system is proposed specifically for shape feature extraction. In this system, Angular radial transform (ART) is used for global feature extraction while Histograms of spatially distributed points (HSDP) [21] is utilized for local content extraction.

Apart from these hybrid systems which are based on a certain feature extraction technique, Machine learning has also contributed fabulously in the domain of image retrieval. Many machine learning based classification algorithms have been successfully utilized in this area. Support vector machines (SVM), Naive Bayes, Random forest etc. are some of the common classifiers categorized under machine learning. A hybrid CBIR system described by Pradnya et al. [22] consists of color correlogram for color extraction and the combination of Gabor and Edge histogram descriptor (EHD) for texture feature extraction. SVM has also been used to obtain a precise classification accuracy. Segmentation based Fractal Texture Analysis (SFTA) can also be employed as a texture analysis technique [23]. Again, for classification purpose SVM classifier has been used.

In the current era, the focus of the researcher community has been shifted from machine learning to Deep learning. Many deep learning techniques have been used for the extraction of image features. Arun et al. [24] describes a Hybrid deep learning architecture (HDLA) which is capable of generating sparse representations used for reducing the semantic gap. This model uses Boltzmann machines in the upper layers and Softmax model in the lower levels. Another deep learning technique which is specified as deep belief network (DBN) [25] has also been utilized for feature extraction and classification in hybrid CBIR system.

Extreme learning machine (ELM) can be considered as a type of Deep learning network because it uses a neural network as its hidden layer. This network is a type of feed-forward neural system and has a solitary hidden layer and has excelled results as compared to other machine learning based classifiers [26]. Kaya et al. [27] describes a texture based hybrid retrieval system which utilizes two texture extraction techniques. For classification accuracy of butterfly images, ELM classifier has also been deployed. In the modern era, Convolutional neural networks (CNN) have also been utilized for the various functions related to image analysis [28].

In literature, the discussed hybrid systems lack in performance due to the utilization of either one or two visual attributes of an image. Due to this, the left-out and un-analyzed attribute cannot contribute to the formation of a final feature vector which is the desired condition for the retrieval of complex images. Moreover, the described systems in literature do not provide information concerning frequency relating to co-occurring of local patterns of an image. But, the proposed system is an intelligent fusion descriptor which contributes to producing a highly effective and accurate system by including spatial information and prominent interconnection among pixels of an image. The information concerning the shape attribute of an image is also included by analyzing some of its prominent parameters. The proposed system also includes human semantic information by using an intelligent technique of Relevance feedback.

C. Main Contributions

A predominant requirement of an effectual CBIR system is that the system should evaluate all the three visual media attributes i.e texture, shape and color in order to form an accurate retrieval system. CBIR systems based on a single feature extraction lack the description of complex images. Moreover, the usage of traditional machine learning based classifiers is deficient of desired retrieval accuracy. Lack of semantic information also reduces the performance of the system. To address these issues, an effective and novel CBIR is proposed where color is extracted using color moment, texture analysis is done using Gray level co-occurrence matrix (GLCM) and different region props are used for the extraction of shape attributes. To get an exact classification accuracy, an Extreme learning machine (ELM) classifier is used. Finally, to capture the high-level semantics of an image, Relevance feedback is used following various rounds to reformulate the training of an ELM classifier.

The remaining organization of this paper is given in the following way: Before the description of the implemented technique and various utilized techniques have been discussed in section II, designated as Preliminary section. Proposed work has been given in section III. Experimental simulation and analysis have been given in section IV and finally, in section V, the conclusion with future trends has been given.

II. Preliminaries

Various techniques used in the creation of the proposed system are described in this section.

A. Color Moment

In image retrieval systems, many techniques have been used for the extraction of color features. Among these, color histogram [29] is the conventional method of color feature extraction. Though it is very simple and is invariant to scale and angle rotation but it does not convey any spatial information regarding an image. Researchers have also used Color coherence vector (CCV) as a color feature descriptor but the feature vector produced by this method is of high dimensionality which is against the basic requirements of any CBIR system. Dominant color descriptor (DCD) also suffers from the lack of complete spatial

information and moreover, if the obtained feature vector is compact, then it is utilized feasibly, otherwise vague results can be produced.

Color auto-correlogram (CAC) has a high computation time and cost and it is very sensitive to noise also. Therefore, based on these facts and conclusions, the color moment has been chosen as an effective color feature extraction technique. It is robust, fast, scalable, consumes less time and space. Color moments are metrics which signify color distributions in an image. According to probability theory, its distribution can be effectively characterized by its moments. Color moment can be computed for any color space model. Different moments specify diverse analytical and statistical measures. This color descriptor is also scale and rotation invariant but it includes the spatial information from images [30].

If I_{ij} specifies the i_{th} color channel and j_{th} image pixel, the number of pixels in an image are N, then index entries associated with the particular color channel and region r is given by first color moment, which signifies average color in an image denoted as:

$$Mean(E_{r,i}) = \frac{1}{N} \sum_{j=1}^{N} I_{ij} \tag{1}$$

The next moment is given by Standard deviation which is obtained from a color distribution of a particular image and is defined as the square root of the variance. It is given as:

$$Standard\ deviation(\sigma_{r,i}) = \left(\frac{1}{N} \sum_{j=1}^{N} (I_{ij} - E_{r,i})^2 \right)^{\frac{1}{2}} \tag{2}$$

The third moment is called as skewness and signifies how asymmetric is the color distribution. It is given as:

$$Skewness\ (S_{r,i}) = \left(\frac{1}{N} \sum_{j=1}^{N} (I_{ij} - E_{r,i})^3 \right)^{\frac{1}{3}} \tag{3}$$

Fourth moment is denoted by Kurtosis and it signifies the color distribution shape, emphasizing particularly on tall or flat shape of the color distribution.

B. Gray Level Co-occurrence Matrix

In order to withdraw features related to textural patterns, various second-order statistical methods have been utilized. Among these, Gray level Difference matrix (GLDM) is error free in the calculation and is based on the difference between gray level pixel pairs. But, its main drawback is that with the change in gray level variance, it also becomes variant. On the other hand, Gray level run length matrix (GLRLM) is based on counting the number of runs which in turn depicts gray levels in an image. This technique suffers from meagerness to represent the pattern of an image and also its computational involvement is high. For analyzing a signal in time-frequency zone, initially Fourier based transforms like Discrete cosine transform (DCT) and Discrete fourier transform (DFT) were in trend. But, these traditional methods are deficient, as they do not convey the local information of an image. Also, the images regenerated by these techniques are of deprived standard, especially at the edges because of the presence of high-frequency bristly components.

In the wavelet domain, both Gabor wavelet and Discrete Wavelet Transform are highly prominent. But, due to a large dimension of feature vectors, Gabor wavelet takes more time in image analysis. Discrete wavelet transform (DWT) also suffers from many disadvantages like ringing near discontinuities, variance with shift, the lack of directionality of decomposition functions, etc. Local binary pattern (LBP) is also considered as an important tool to extract texture information from an image. The main issues associated with this technique are the production of very long histograms which slows down

the process of image recognition, sensitivity to noise and the effect of the center pixel is sometimes not included. Based on these conclusions, the best second-order statistical texture feature extraction technique is Gray level co-occurrence matrix (GLCM) which has many dominant attributes like: (1) Rotation invariance (2) Diverse applications (3) Simple implementation and fast performance (4) Numerous resultant (Haralick) parameters (5) Precise inter-pixel relationship.

It is an analytical technique used to withdraw texture features for the retrieval and classification of different images. The spatial correlation enclosed by the pixel duplet in any given image is also computed by this geometrical method. A co-occurrence matrix denoted by P_{dis} *(m,n)* specifying grey levels, contains information about two pixels: Gray level content m is denoted by the first pixel and content n is denoted precisely by the second pixel which further is separated by a distance denoted by *dis*. These specifications are chosen according to a specific angle. Matrices produced by this technique give the gray level spatial frequencies which further gives association among pixels that are adjacent and have distinct distances amidst them [31].

Therefore, the description of GLCM is as follows:

$$P_{dis,\Theta}(m,n) = P_r(I(p_1) = m \wedge I(p_2) = n \wedge \|p_1 - p_2\| = d_{is} \tag{4}$$

In equation (4) I is the given image and the positional information in the same image I is given by p_1 and p_2, the probability is denoted by P and Θ denotes the range of different angle directions given by 0º, 45º, 90º and 135º. Therefore, a GLCM image is represented by d as its vector used for movement, δ as its radius and θ as orientation. A generalized GLCM matrix can be represented by Fig. 1.

Gray Tone	0	1	2	3
0	# (0,0)	# (0,1)	# (0,2)	# (0,3)
1	# (1,0)	# (1,1)	# (1,2)	# (1,3)
2	# (2,0)	# (2,1)	# (2,2)	# (2,3)
3	# (3,0)	# (3,1)	# (3,2)	# (3,3)

Fig. 1. A Generalized GLCM.

Therefore, for a given specific test image with gray tone values, a GLCM matrix is formed, which is the spatial co-occurrence dependence matrix.

Prominent four types of GLCM feature parameters which are subjected to be used for the extraction [32] of the textual content of an image and are denoted by:

$$Contrast = \sum_m \sum_n [P(m,n) * (m-n)]^2 \tag{5}$$

The disparity between the topmost and the bottom most conterminous pixel sets is given by Contrast intensity.

$$Correlation = \frac{\sum_m \sum_n (m-u_x) * (n-u_y) * P(m,n)}{\sigma_i \sigma_j} \tag{6}$$

The correspondence between a reference pixel and its adjoining pixels in an image is diagnosed by making use of Correlation. It considers the mean and standard deviation of a matrix by encapsulating both the row and column of that particular matrix.

$$Homogeneity = \sum_m \sum_n \frac{P(m,n)}{1+|m-n|} \tag{7}$$

In the spatial domain, the proximity among gray levels in an image is defined by the term homogeneity.

$$Energy = \sqrt{\sum_m \sum_n P(m,n)^2} \tag{8}$$

Energy of a texture denotes the cyclic consistency of gray level allocation in an image.

C. Region-props Process

Shape is also an important visual attribute which can be used to depict the information related to an image. Shape features of an object provide valuable information about the identity of the same object. These features can be categorized into two types: based on its region and based on its boundary [33]. Information about an object's internal regions is depicted by region-based descriptors while boundary based shape descriptors are based on the usage of boundary information of an object. Fourier descriptors are among the popular techniques of boundary-based shape descriptors. They are robust, contain perceptual characteristics but information about local features is not present because only magnitudes of the frequencies are present in Fourier transform and location information is missing.

Curvature scale space (CSS) is another shape descriptor which analyzes the boundary of an object as a 1D signal and finally represents the signal in scale space. The major issue with this technique is the superficial projections on the shape of an image. Angular radial transform (ART) produces a high dimensional feature vector which causes a hindrance in the performance of a CBIR system. Image moments, Zernike moments, Hu moments, Canny edge detector are also among the prominent shape and edge descriptors but suffer from some main functioning issues like prior image normalization to obtain scale invariance, more time consumption, computational complexity, etc. Properties of image regions are efficiently measured by using region props which measure the number of connected components in an image. Many types of region props such as Center of gravity (Centroid) [3], Mass, Dispersion, Eccentricity, Axis of least inertia, Hole area ratio, etc. can be used to find the shape related information in an image. In this paper, some of the region props are used to find the largest connected component. They are:

Mass: It is the total number of pixels present in one class. It is given as:

$$\sum_{mn} h(m,n) \tag{9}$$

Where $h = \begin{cases} 1 \ if \ s(m,n) \in C \\ 0 \ if \ s(m,n) \notin C \end{cases}$

Centroid: The center value of all the pixels is denoted by Centroid and is also known as the center of mass. C denotes the cluster, h specifies mask over the same cluster C over image $S(m,n)$. A Centroid is given by:

$$y_C = \frac{\sum_{mn} m * h(m,n)}{mass} \tag{10}$$

And $z_C = \frac{\sum_{mn} n * h(m,n)}{mass} \tag{11}$

In these equations (y_c, z_c) are the co-ordinates of the centroid.

Mean: It is defined as the average value of all pixels and is denoted by:

$$\mu_C = \frac{\sum_{mn} I_{mn} * h_C(m,n)}{mass} \tag{12}$$

Variance: It is an analysis which measures the distance of the spread from an average value of given random numbers.

$$\sigma_C^2 = \frac{\sum_{mn} (I_{mn} - \mu_C) * h_C(m,n)}{mass} \tag{13}$$

Dispersion: Dispersion is defined as the total distance from the centroid to every class present in an image. It is given as:

$$Dispersion = \sum_j dist(O_{D,} O_{j,D}) \tag{14}$$

Where $dist(O_D, O_{j,D})$ gives information related to the distance metric, centroid of class D is O_D, centroid of region j of class D is O_{jD}.

D. Extreme Learning Machine

To procure a precise accuracy during the classification of the image dataset, many types of classifiers have been used in the field of image retrieval. Support vector machines (SVMs) [34] are amongst the most widely used classifiers due to their ease of access. But, they are less accurate and take more training time. Moreover, no multi-class SVM is directly available. Naive's Bayes also suffers from various issues like: To deal with continuous features initially, a binning procedure is required to be adopted to translate those features into discrete features. Data scarcity, data assumptions are also some of the issues related to this classifier. Random forest, KNN are also used as classifiers but lack in accurate classification due to some issues. Therefore, an Extreme learning machine (ELM) classifier based on neural network is chosen which has many advantages as compared to these classifiers. The training speed of an ELM is very swift and also it has a rationalized performance based on its operation. ELM is well known as a single-layer feed-forward neural network (SLFN), designed to be used for classification and regression applications. It is a supervised learning model based on labeled data. It was initially introduced by Huang at el. [26]. ELM has been successfully utilized in many research problems like pattern recognition, classification, fault identification, over-fitting, etc. In this system, the input weights are selected without any conscious decision and by using a specified analytical technique, its output weights are decided. This algorithm indeed is based on a single hidden layer and the total number of nodes present in this layer is a principal parameter to be decided. ELM has many advantages in contrast to many related traditional techniques like: least human involvement, [35] swift learning speed, complimentary universal capability, convenience to use, varied kernel functions, etc. [36].

If there are M different samples denoted by (X_j, t_j), where $X_j = [x_1, x_2, x_3, \ldots x_n]^T \in R^n \times R^m$ (where j=1, 2...... M).

Then, it is a SLFN with $\widetilde{M}$ hidden nodes and f(x), where f(x) is an activation function, and can be represented as:

$$\sum_{j=1}^{\widetilde{M}} \beta_i f(u_j x_k + v_j) = o_k \tag{15}$$

In the above equation, the weight vectors, which connect the input nodes to the jth hidden nodes, are given by $u^j = [u^{j1}, u^{j2}, u^{j3}\ldots u^{jn}]_T$ and $\beta^j = [\beta^{j1}, \beta^{j2}, \beta^{j3}\ldots\beta^{jn}]_T$ represents the connecting vectors between the jth hidden nodes and the nodes depicting outputs. The value describing threshold for the various hidden nodes is given by v^j whereas u^j and x^k is the inner product. The basic diagram of ELM is shown in Fig. 2.
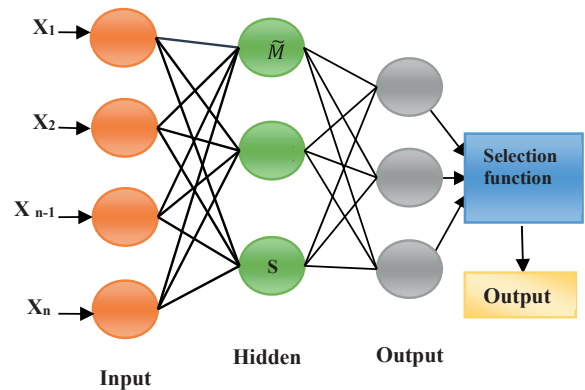

Fig. 2. A basic form of an Extreme learning machine.

The Equation (15) can be briefly re-written as:

$$H = \beta O \tag{16}$$

Where

$$H = \begin{bmatrix} f(u_{1,}x_{1,}v_1) & \cdots & f(u_{\tilde{M},}x_{1,}v_{\tilde{M}}) \\ \vdots & \vdots & \vdots \\ f(u_{1,}x_{M,}v_1) & & f(u_{\tilde{M},}x_{M,}v_{\tilde{M}}) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{M}}^T \end{bmatrix} \text{ and } O = \begin{bmatrix} O_1^T \\ \vdots \\ O_1^T \end{bmatrix}$$

Here, H represents the output matrix of the hidden layer. Also, the H matrix becomes invertible, if the total number of given samples $\tilde{M}$ becomes equal to hidden node parameters H. However, the learning attributes of ELM, u_j, v_j and the hidden nodes can be allocated randomly, in the absence of any input data, therefore the output weights β of a linear system can be calculated by implementing least square technique as follows:

$$\hat{\beta} = H^\dagger T \qquad (17)$$

Where Moore-Penrose conception in one of its versions is given by $H^\dagger$. So, from equation (17), we can see that the calculation of the output weights is done by using a straightforward mathematical equation, thereby avoiding any lengthy procedures. Thus, the algorithm of ELM can be summarized in three steps, which are as follows:

(1) Assigning of hidden node parameters u_j and v_j, where i=1,2……. $\tilde{M}$.

(2) Calculation of H, i.e matrix of a hidden layer by using an activation function.

(3) Output weights calculation by using $\beta = H^\dagger T$

Thus, ELM transforms a complex problem into a simpler and linear function. Fast speed, more accuracy and many more advantages contribute to making this technique more sophisticated and precise as compared to many other customized methods.

E. Relevance Feedback

It is a strategy which helps to refine a particular image based on the feedback obtained by the user. To search the system with massive database images, text or a combination of images annotated with text can be used as a query. Then, a set of relevant images are obtained based on a specific query image. These retrieved images are analyzed by the user and finally, the query image is refined by using Relevance feedback which selects the best-matched images, based on some common features. This process works iteratively until the desired results are obtained or the user gets satisfied. This intelligent technique can also be used in combination with many other concepts like Support vector machines (SVM), Neural networks with different training algorithms [37], Deep learning, machine learning, etc. [38].

The query input given by the user can be broadly classified into three types: The first category consists of a system in which a query image is composed of only keyboard text letters. This technique has some limitations like polysemy, synonymy, homonymy, etc. So, finding the desired images based on the user's intention is a major issue of concern. A query can also be given in the form of an image, which is the second medium of inputting a query image. This technique has removed many ambiguities, which were present in the traditional method of the query by text. Also, this method has gained vast popularity in recent times due to its numerous applications in image processing. Relevance feedback can be considered as the third category of providing a query image, indeed through the iterative refinement of a user's query image. The three basic ways of a query refinement are as follows:

(1) Extension of Query: In this technique, the neighboring images of an actual query image are also included in it, based on the feedback obtained by the user. Thus, in a way, an expansion of an original query image is done.

(2) Query Re-Weighting: This method enhances the weights of some prominent attributes of an image and simultaneously reduces the weights of some un-important attributes. In this way, a query becomes more refined.

(3) Movement of Query: A query is moved close to the required images by the adjustment in the attributes of a utilized distance function.

In this paper, to make the hybrid textural system more effective, Relevance Feedback is utilized. It works on the relevant images obtained after classification by ELM. It works on the refinement process until the satisfactory results are obtained.

III. Proposed Methodology

In the proposed work, a unique image retrieval system has been described which is an amalgamation of texture, color and shape features depicting an image. Color moment has been utilized for color feature extraction while Gray level co-occurrence matrix (GLCM) and varied region props have been adopted respectively for texture and shape feature extraction. These three techniques together present a great combination. Color moment effectively captures spatial information of a particular image and is also invariant to rotation, scale and angle. Similarly, GLCM has highly factual results and takes very little computation time in its execution. For the shape feature extraction, Mass, Centroid, Mean, Variance and Dispersion parameters have been calculated. These parameters effectively describe the shape of an object, specified in the given image. Hence, our proposed framework consists of an amalgamation of these three techniques. For classification accuracy, the incorporation of an Extreme Learning Machine (ELM) has also been done whose training images have been reformulated or updated based on the condition of relevance feedback. For the detailed explanation of the proposed work, this approach has been divided into two major subsections: The first section explains the working of both the main phases of the system, i.e., Training stage and the Testing stage. The other stages related to the proposed model have been given in another subsection.

A. Training and Testing Stage

The stage related to training of the system is basically concerned with the training of an ELM model. Every neural network is required to be trained through the total images in the database. In this phase, the three types of features are withdrawn from the whole dataset and three independent feature vectors are formed. These three feature vectors are normalized to form a hybrid feature vector (HFV). To train the ELM model, this HFV is given as an input to it and different categories of a dataset are formed and gets trained as a classifier. Normalization brings the feature dimensions into a common range. Here, minimum-maximum normalization is used and is denoted by:

$$Obtained\ normalized\ features = \frac{y - maximum}{minimum - maximum} \qquad (18)$$

Where y is the value of a particular feature, *minimum* is the bottom value of every single feature vector and *maximum* is the highest value of every single feature.

Decimal scaling and Z-score are also among the prominent Normalization techniques. In Decimal scaling, the normalization is achieved by moving the decimal point of input values. But, it is generally based on calculating the least and the largest value of given data, which is difficult in some cases. Moreover, if the assumption about these values is done, then also the results could be impermissible. The main disadvantage of Z-score Normalization is that it always assumes a normal distribution. But, if this condition is not met, then vague
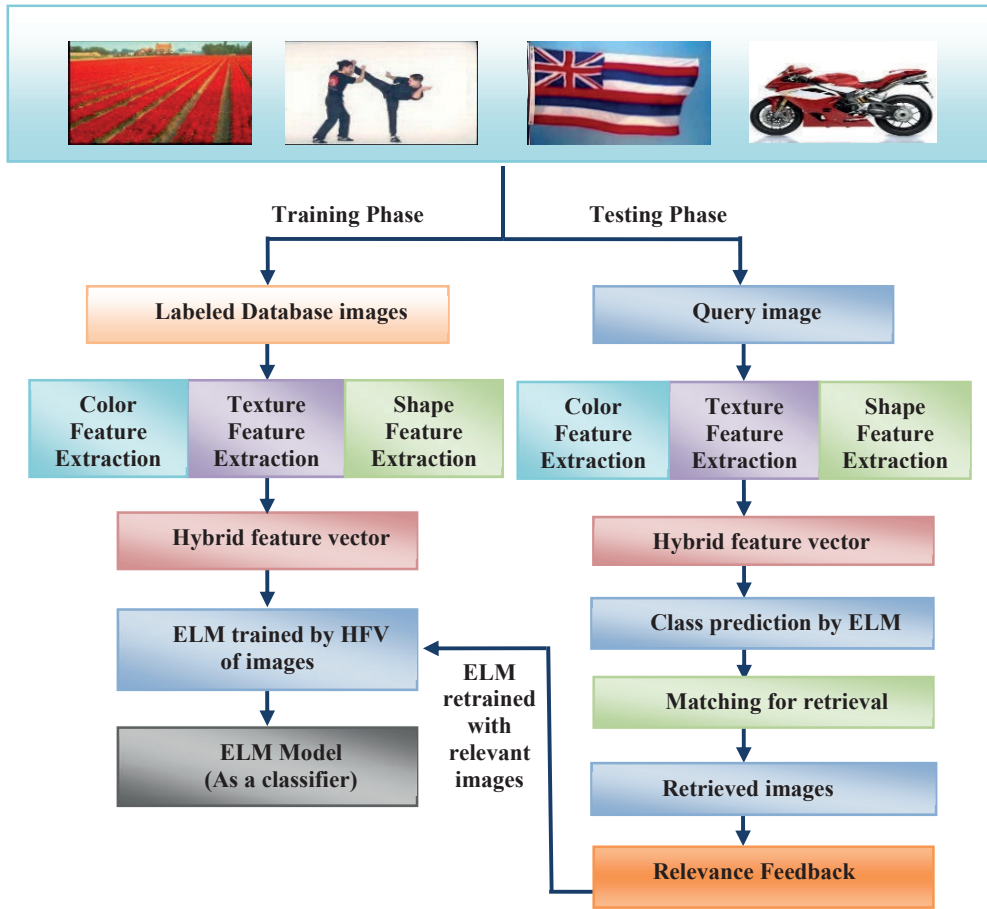
Fig. 3. Architectural framework of the proposed system.

results are produced. Therefore, on the basis of these conclusions, Min-Max Normalization is preferred because the correspondence among all the data values is conserved, without any bias introduction. The basic architectural implementation is given in Fig. 3.

The main aim of the testing stage is concerned with the testing of the proposed system by using a pre-trained ELM model with a specific query image and obtaining the desired top N images.

B. Prominent Stages During the Working of the Proposed System

The various prominent stages during the working of the proposed system are as follows:

1. Feature Extraction Stage

In this stage, texture, shape and color features are withdrawn from a query image using GLCM, region props process and color moment respectively and three respective feature vectors are obtained. Again, a hybrid feature vector (HFV) is finally obtained by combining the three independent feature vectors by using the process of normalization.

2. Classification Stage

The second important stage of working is based on classification. Here, the obtained HFV is applied in the form of an input to a pre-trained ELM model. This ELM model has been initially trained in the training phase with the complete specific database.

To train the ELM model, different types of activation functions can be used and here, Radial basis function (RBF) is utilized. RBF is a type of multi-layer perceptron (MLP) which can use one or more hidden layer besides input and output layer. RBF has only a single hidden layer of neurons where each neuron of the hidden layer calculates the RBF function. These hidden nodes project the lower dimension

of feature vector to a higher dimension. The output nodes contain the classifying neurons. The number of neurons in the output layer are equal to the number of classes of the dataset. The output layer nodes decide to which class, the input feature vector should lie. If the output of first node = 1 and all other nodes = 0, then it means that the query image belongs to first category.

The output of this classification step is the images with meticulous categorization and class prediction by an ELM classifier. Various working parameters of the proposed system are given in Table I. Here, 9 features of color moment are used as input, 44 of GLCM and Region props contain 5 features. To form a hybrid system, these features are added and 58 input features are obtained. For, Corel-1K dataset, the output neurons are 10, 50 for Corel-5K and so on. Single layer feed forward neural network is used here as it contains a single hidden layer and it is a feed forward neural network. Radial basis function (RBF) is utilized as an activation function where 100 denotes the number of neurons in the hidden layer.

TABLE I. Various Working Parameters of the Proposed System

Number of input features	9 (Color), 44 (Texture), 5 (Shape)
Fused features	9+44+5=58
Number of Output (Classes)	10 (Corel-1K), 50 (Corel-5K), 100 (Corel-10K), 20 (GHIM-10)
ELM Kernel Type	Radial basis function (RBF)
Kernel Parameters	100
Algorithm	Feed-forward single hidden layer neural network

3. Similarity Matching Stage

After an accurate class prediction by an ELM network, the whole dataset images are successfully classified into varied categories. Now, based on a query image, a similarity matching is done between a given query image and the respective category to which it belongs. These categories are formed as a result of an ELM classification. After the results of similarity calculation are obtained, the resultant images are arranged in increasing order based on the utilized distance metric. A result of zero with regard to distance metric exhibits accurate resemblance between two images. Many types of distance metrics are utilized for the purpose of calculating similarity. Some of the prominent distance metrics which are used in similarity calculation are given under:

$$Distance_{Euclidean} = \sqrt{\sum_{j=1}^{n}(|I_j - D_j)^2|} \quad (19)$$

$$Distance_{Manhattan} = \sum_{j=1}^{n}|I_j - D_j| \quad (20)$$

$$Distance_{Minkowski} = \left[\sum_{j=1}^{n}(|I_j - D_j)^{1/P}|\right] \quad (21)$$

Here, I_j denotes the input query image and D_j depicts all database images.

4. Relevance Feedback Stage

The main aim of this step is to encompass the user feedback to check the relevancy of the retrieved images after classification by ELM. The images retrieved after classification are divided into two groups: Relevant and Non-Relevant based on the feedback obtained by the user. These set of Relevant images are again considered to improve and reformulate the classification procedure of ELM. The novelty of the proposed approach is that the obtained set of predicted semantics by ELM is improved to a great extent by using two iterations of relevance feedback. This process rejects the non-relevant images based on user's feedback and finally, based on these refined images, final top 10 images are retrieved. This process enhances the accuracy of the proposed system to a significant value.

IV. Experimental Methods and Results

To analyze the retrieval efficiency of the implemented technique, various benchmark datasets for CBIR system has been utilized. There are a diversity of images present in these databases which have been used successfully in many retrieval techniques. All the experiments have been performed on Windows 10 operating system (OS) by utilizing version R2017a of MATLAB with core i3 processor, 4 GB RAM and 64-bit windows.

1st Dataset: Corel-1K: The initial database analyzed for experimentation is Corel-1K which is composed of 1000 different images. It has 10 categories which are Africa, Buildings, Bus, Beach, Dinosaur, Elephant, Flowers, Horse, Food and Mountains. 100 images are present in each category and each image bears the size of either 256 × 384 or 384 × 256. (http://wang.ist.psu.edu/docs/related/)

2nd Dataset: Corel-5K: The next dataset is Corel-5K and consists of 5000 images. There are 50 categories in this dataset and again every group has 100 images. Some of the group types of this dataset are butterfly, cards, vegetables, dogs, judo, etc. The size of each image is either 256 × 384 or 384 × 256. (http://www.ci.gxnu.edu.cn/cbir/)

3rd Dataset: Corel-10K: The third dataset is Corel-10K and there are 10,000 images in this dataset. This database has different groups of 100 images and every set has further 100 images in it. Every image bears the size of either 256 × 384 or 384 × 256. Some of the images of this dataset belong to cars, dolls, lamp-posts, flags, etc. category. (http://www.ci.gxnu.edu.cn/cbir/)

4th Dataset: GHIM-10: The last database for the experimental analysis is GHIM-10. It also consists of 10,000 images with a total of 20 categories. Each category has 500 images in it consisting of bikes, car, aeroplane, grasshopper, etc. Every image size is of either 300 × 400 or 400 × 300. (http://www.ci.gxnu.edu.cn/cbir/)

For the formation of a query or input image, each and every single image of all the databases is utilized. If the obtained resultant images correspond to the concordant native category of the input image then, the proposed system has effective results and the retrieval is considered as a successful retrieval. The role of Extreme learning based-Relevance feedback framework contributes to enhancing and upgrading the retrieval efficiency of the implemented technique. Few of the sample images from all the datasets is shown in Fig 4.
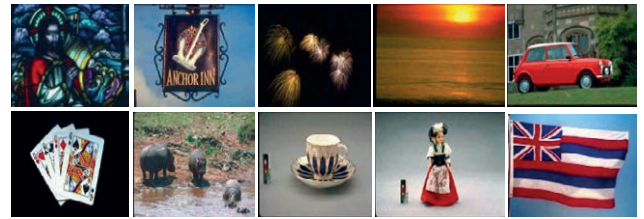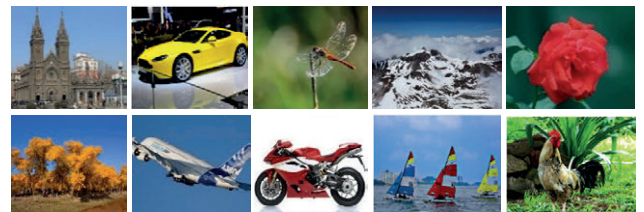
Corel-1K

Corel-5K
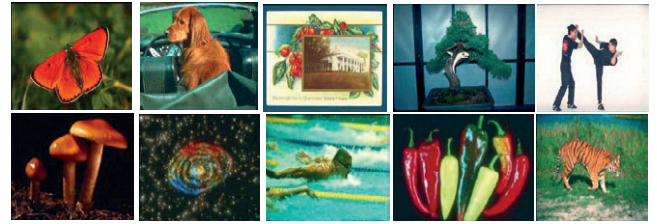
Corel-10K

GHIM-10

Fig. 4. Sample images from all datasets.

A. Methods

In CBIR systems, the capability of a particular system can be concluded with respect to many evaluation parameters [39]-[40]. Precision and Recall are the most well-known evaluation metrics. These are defined using the given equations:

$$Precision(P_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \quad (22)$$

$$Recall(R_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ database} \quad (23)$$

Here, the total recovered images are 10 while the number of relevant images of a particular dataset depends on the number of images present in each category of that dataset. Corel-1K, Corel-5K and Corel-10K, has 100 images as relevant while in the case of GHIM-10, it is 500.

B. Results

Results based on the retrieval of desired images are obtained by taking each and every image from all the datasets i.e Corel-1K, Corel-5K, Corel-10K and GHIM-10 as a query or an input image. Then, Color moment, GLCM and various region props are used for the withdrawal of color, texture and shape attributes and the hybrid feature vector is formed by combining independent feature vectors of the three techniques. The same procedure applies to all database images also. After feature extraction, the pre-trained ELM model performs the accurate class prediction and finally, a similarity is calculated between the entered input image and the classified images of the complete dataset. Similarity calculation is done by utilizing three distance metric techniques. The results in terms of Average precision on all the employed datasets, obtained after classification by ELM using the three distance metrics are shown in Table II.

TABLE II. Average precision using three distance metrics

Dataset	Euclidean	Manhattan	Minkowski
Corel-1K	92.1	88.45	85.3
Corel-5K	79.6	73.6	69.9
Corel-10K	72.4	68.7	65.3
GHIM-10	89.95	83.2	78.62

As evident from Table II, the average precision obtained by using Euclidean distance metric outperforms the other distance metrics. Since Manhattan distance is a distinctive case of Minkowski distance, it produces innumerable false negatives and does not yield accurate results. Euclidean distance metric is based on weighted and normalized attributes and has speedy computational performance. Therefore, the Euclidean distance metric gives precise results and is being used here.

The average precision of the proposed system is obtained in the form of three phases. In the first phase, precision is calculated only after the fusion of texture, color and shape parameters. In the second phase, the results are obtained from the combination of hybrid features and ELM classifier and, in the last phase, the results of the total proposed system are obtained which is the combination of hybrid attributes, ELM and Relevance feedback. Fig. 5 plot shows the Average precision vs Datasets plot for all the four databases in the stepping of three levels.
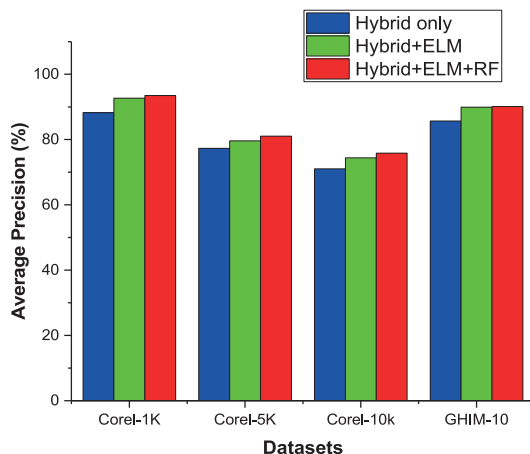

Fig. 5. Average Precision of the proposed system depicting three levels of functioning.

From Fig. 5, it is clear that the average precision of the system increases as more and more intelligence is being added to the proposed system. With the increase in intelligence of the system, more high-level semantics are captured and the system becomes more efficient.

Precision and Recall are the prominent measures to check the effectiveness of any particular CBIR system. Therefore, Precision vs Recall curves of the implementation by varying the number of retrieved images from 10 to 50 on all the four datasets is shown in Fig. 6 and 7.
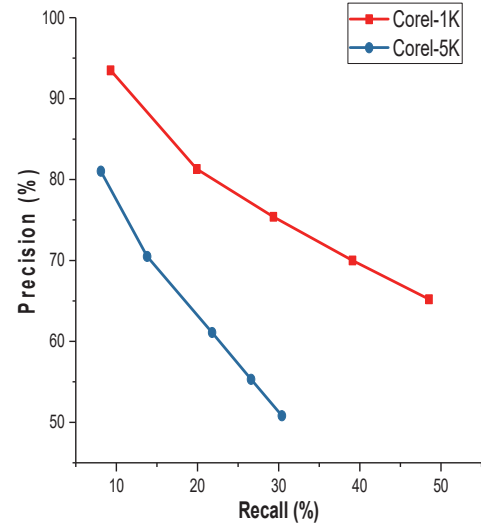

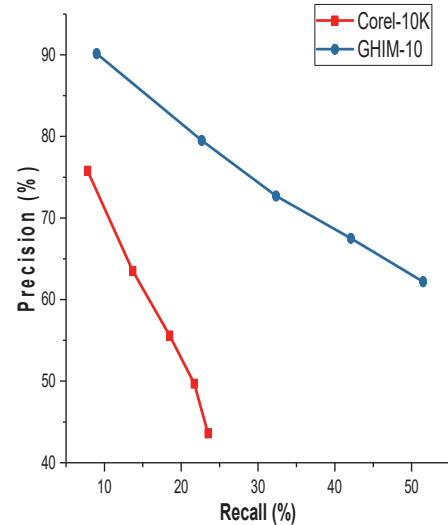Fig. 6. Precision Vs Recall plot on Corel-1K and Corel-5K datasets.

Fig. 7. Precision Vs Recall plot on Corel-10K and GHIM-10 datasets.

From Fig. 6 and 7, it can be seen that with the rise in the number of images retrieved, the value of precision decreases while recall increases. Separate Graphical user interface (GUI's) have been designed for each of the four datasets. Top ten images are retrieved through each utilized dataset. The GUI's depicting the retrieval results for each of the four datasets based on a specific query image are shown in Fig. 8 (a-d).

From Fig 8, it can be concluded that the top ten images are retrieved from the desired category of the dataset which in-turn belongs to the native category of the input image. Thus, the gross accuracy of the implemented technique is undoubtedly outstanding as compared to the other state-of-the-art techniques based on intelligent and hybrid feature retrieval. The accuracy of the presented system is also given in Table III.
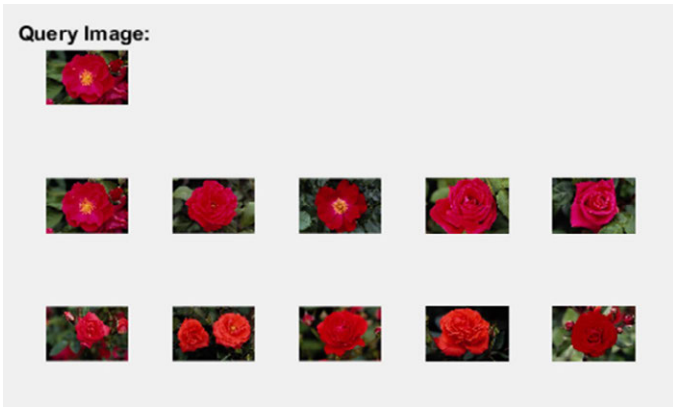
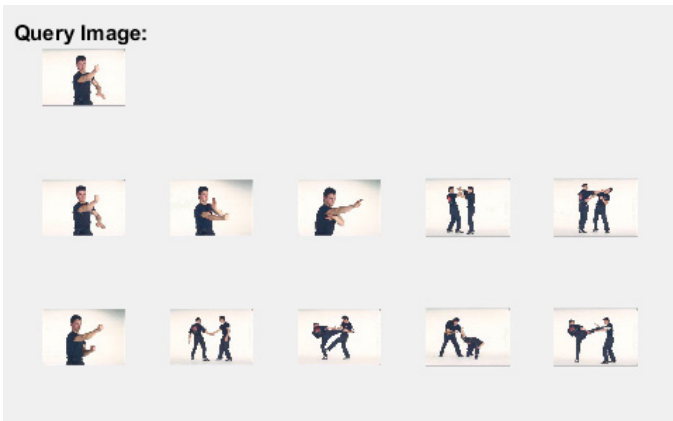Fig. 8(a). Retrieval results from Corel-1K dataset.
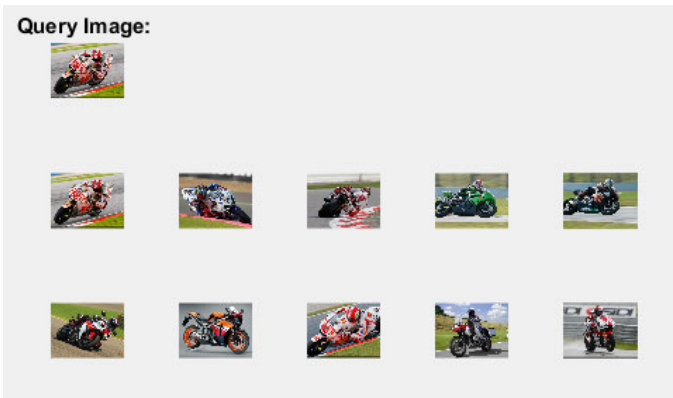

Fig. 8(b). Retrieval results from Corel-5K dataset.

Fig. 8(c). Retrieval results from Corel-10K dataset.

Fig. 8(d). Retrieval results from GHIM-10 dataset.

TABLE III. Accuracy of the Proposed System

Dataset	Accuracy (%)
Corel 1K	96.5
Corel 5K	94.67
Corel 10K	94.85
GHIM-10	99.02

The obtained accuracy can also be verified by checking the diagonal elements of the generated confusion matrix during classification by ELM classifier. Confusion matrix of one of the four datasets (GHIM-10) is also given in Table IV.

TABLE IV. Confusion Matrix of Ghim-10 Dataset

496	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
2	496	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	495	0	1	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0
0	0	0	498	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	498	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	495	0	0	1	1	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	496	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	0	0	0	0	0	497	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	1	0	497	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	1	0	1	0	496	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	1	1	0	0	0	496	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	2	2	0	0	495	0	0	0	0	0	0	0
3	1	0	0	0	0	1	0	2	0	0	0	0	493	0	0	0	0	0	0
1	1	1	0	2	0	1	0	0	0	0	0	0	0	494	0	0	0	0	0
1	0	0	0	0	2	0	1	1	0	0	0	0	0	0	495	0	0	0	0
0	0	0	0	1	0	0	0	2	2	0	0	0	0	0	0	495	0	0	0
0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	494	0	0
1	0	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	494	0
1	0	0	0	0	2	0	0	2	2	0	0	0	0	0	0	0	0	0	493

From the diagonal elements of the confusion matrix, it can be seen that out of 500 images present in 20 categories of the dataset, a handful of images are present in every single group, which corresponds to an accuracy of 99.02%.

C. Comparison of the presented system with the related techniques

The implemented system has been initially compared to many state-of-the-art related techniques regarding average precision obtained on Corel-1K, Corel-5K and Corel-10K datasets. In comparison, the main considerations are:

The majority of the hybrid systems are based only on the extraction of either one or two attributes of an image in spite of all the three basic visual attributes. Due to this, those systems lack in the recognition of complex and large dataset images.

The methods which are used for feature extraction by the related hybrid systems are deficient in one or the other ways as compared to the proposed system. For eg., the color histogram is used for color extraction but it lacks the spatial information and, moreover, two different images can produce the same histograms.

In order to classify the images, generally Support vector machine (SVM) has been utilized but it has less classification accuracy as

TABLE V. Comparison of the Proposed System with the State-of-the-Art Techniques

Database	Semantic Name	Average Precision (%)							
		Ref. [2]	Ref. [5]	Ref. [7]	Ref. [8]	Ref. [1]	Ref. [9]	Ref. [10]	Proposed
Corel-1K	Africa	72.4	81	68	74.5	95	81	84.7	84.5
	Beach	51.5	66	65	69.3	60	92	45.4	98
	Building	59.55	78.75	80	85.1	55	79	67.8	79
	Bus	92.35	96.25	90	95.4	100	93	85.3	86
	Dinosaur	99	100	100	100	100	99	99.3	100
	Elephant	72.7	70.75	80	83.3	90	79	71.1	100
	Flower	92.25	95.75	85	98	100	99	93.3	99
	Horse	96.6	98.75	95	94.2	100	80	95.8	100
	Mountain	55.75	67.75	75	75.9	75	85	49.8	84
	Food	72.35	77.25	85	92.6	100	88	80.8	100
	Average	76.5	83.225	82.3	77.8	87.5	87.5	77.3	93.05
Corel-5K	Average	55.25	68.6	64.5	59.5	72.56	75.4	58	81.03
Corel-10K	Average	49.58	59.98	57.54	52.45	65.45	68.9	72	75.8

compared to Extreme learning machine (ELM) employed in the presented system.

To capture the high-level semantic features of an image, Relevance feedback has been used in the proposed system but the related systems lack this concept of human intelligence.

These comparisons have been given in Table V where the precision of the implemented system outshines the other state-of-the-art compared techniques and performance plot has been given in Fig. 9.
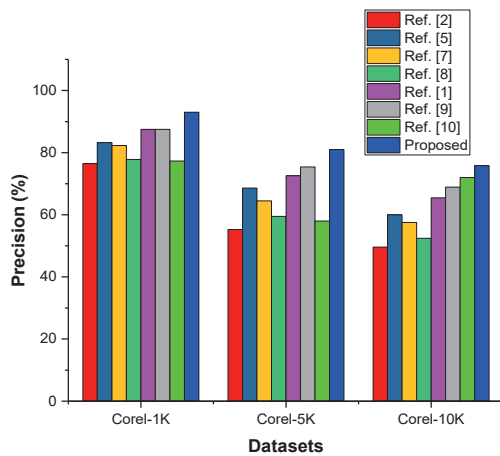

Fig. 9. Comparison plot of the proposed technique with the related techniques on Corel-1K, Corel-5K and Corel-10K dataset.

Again from the comparison of GHIM-10 dataset, it can be concluded that the proposed system has enhanced and accurate results as compared to many related techniques based on this database. The average precision of the proposed system obtained on GHIM-10 dataset is shown in Table VI and its comparative plot in Fig. 10.

TABLE VI. Comparison of the Proposed System Based on GHIM-10

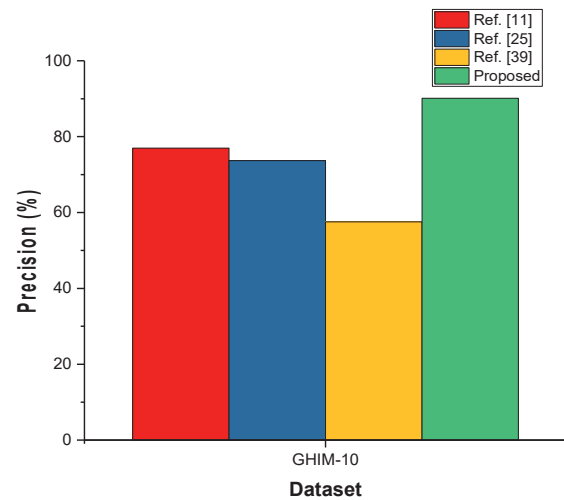| Database	Average Precision (%)			
GHIM-10	Ref. [11]	Ref.[25]	Ref. [39]	Proposed
	76.99	73.7	57.51	**90.14**

Fig. 10. Comparison plot of the proposed technique with the related techniques on GHIM-10 dataset.

The comparison of the proposed system in terms of Recall is also presented in Table VII for Corel-1K, Corel-5K and Corel-10K datasets. Table VIII gives the comparative analysis for GHIM-10 dataset for the Recall parameter. These recall results are based on the retrieval of 20 images.

TABLE VII. Comparison of the Proposed System Based on Corel-1K, Corel- 5K and Corel-10K

Database	Average Recall (%)				
	Ref. [1]	Ref. [5]	Ref. [7]	Ref. [8]	Proposed
Corel-1K	17.50	16.64	16.60	7.58	**19.92**
Corel-5K	14.67	13.72	14.45	5.33	**16.54**
Corel-10K	12.34	11.99	13.65	4.65	**13.75**

TABLE VIII. Comparison of the Proposed System Based on GHIM-10

Database	Average Recall (%)			
GHIM-10	Ref. [11]	Ref.[39]	Ref. [25]	Proposed
	22.33	1.38	25.65	**28.98**

Thus, the proposed method has superior results both in terms of Precision and Recall, in contrast to many state-of-the-art techniques and can work accurately and precisely on both small and large datasets.

D. Time Performance Analysis

In order to increase the accuracy of the proposed system, time performance analysis is an important parameter to be considered. Here, analysis of time is done during training and testing the model. This time analysis during training phase is divided into Feature extraction time and ELM training time while during testing phase it is based on the testing time of the complete proposed model. The time analysis for all the four utilized datasets is given in Table IX.

TABLE IX. Time performance Analysis of the Proposed System

Dataset	Time (in seconds)
Corel 1K	Feature extraction time = 348.64 ELM training time = 0.1781 Total training time = 348.81 Testing time = 0.389
Corel 5K	Feature extraction time = 595.733 ELM training time = 1.3029 Total training time = 597.03 Testing time = 0.955
Corel 10K	Feature extraction time = 2845.00 ELM training time = 245.67 Total training time = 3090.67 Testing time = 4.342
GHIM-10	Feature extraction time = 2543.00 ELM training time = 205.78 Total training time = 2748.78 Testing time = 3.044

From Table IX, it can be concluded that as the used dataset becomes more and more complicated i.e. number of images increases, some more time is utilized for the total training of the model but the testing time is much less. Thus, the proposed system is very effective in testing both smaller and larger image datasets.

V. Conclusion and Future Work

This paper describes a novel and an efficient technique for Content-based image retrieval (CBIR) system which is focused on the formation of a hybrid feature vector (HFV). This HFV is formed utilizing the independent feature vectors of three visual attributes of an image, namely texture, shape and color which are extracted by using Gray level co-occurrence matrix (GLCM), region props procedure employing varied parameters and color moment respectively. The proposed system is the combination of these three techniques which has many advantages as GLCM has precise inter-pixel and inter-pattern relationship, as compared to many basic texture extraction methods. Color moment captures spatial information of an image and is also invariant to scale, angle and rotation. Shape parameters can be used to detect the connected components in an image. These hybrid features are applied to an Extreme learning machine (ELM) deputing as a classifier which is a feed-forward neural network having one hidden layer. After that, to retrieve the higher level semantic attributes of an image, Relevance feedback is used in the form of some iterations based on the user's feedback. This extreme learning based-Relevance feedback framework helps in the evolution of an intelligent and modified system for learning and classification. Four benchmark datasets have been tested on the proposed system with respect to Precision, Recall and Accuracy. The average precision for the presented implementation is 93.05%, 81.03%, 75.8% and 90.14%,

respectively, on Corel-1K, Corel-5K, Corel-10K and GHIM-10 datasets, which is significantly larger than that of many state-of-the-art related methods of hybrid CBIR system. The proposed work does not consider the information of the desired region of an image but is based on an entire image. Therefore, our future work will concentrate on the Region of interest (ROI) of an image by using local as well as global information of an image by using Deep learning techniques for feature extraction. Internet of Things (IoT) will be used for the online creation and transfer of database images.

References

[1] J. Pradhan, S. Kumar, A. K. Pal, and H. Banka, "A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features," *Digit. Signal Process. A Rev. J.*, vol. 82, pp. 258–281, 2018.

[2] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "New content-based image retrieval system based on optimised integration of DCD, wavelet and curvelet features," *IET Image Process.*, vol. 11, no. 2, pp. 89–98, 2017.

[3] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 20-54, 2015.

[4] R. Chaudhari and A. M. Patil, "Content Based Image Retrieval Using Color and Shape Features," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 1, no. 5, pp. 386-392, 2012.

[5] L. K. Pavithra and T. S. Sharmila, "An efficient framework for image retrieval using color, texture and edge features," *Comput. Electr. Eng.*, vol. 0, pp. 1–14, 2017.

[6] M. K. Alsmadi, "An efficient similarity measure for content based image retrieval using memetic algorithm," *Egyptian J. of Basic and Appl. Sci.*, vol. 4, no. 2, pp. 1-11, 2017.

[7] E. Mehdi, E. Aroussi, and N. El Houssif, "Content-Based Image Retrieval Approach Using Color and Texture Applied to Two Databases (Coil-100 and Wang)". *Springer International Publishing, https://doi.org/10.1007/978-3-319-76357-6_5*, pp. 49-59, 2018.

[8] M. A. Ansari, M. Dixit, D. Kurchaniya, and P. K. Johari, "An Effective Approach to an Image Retrieval using SVM Classifier" *Int. J. of Computer Sciences and Engineering*, vol. 5, no. 6, pp. 64-72, March 2018.

[9] Y. Mistry, D. T. Ingole, and M. D. Ingole, "Content based image retrieval using hybrid features and various distance metric," *J. Electr. Syst. Inf. Technol.*, no. 2016, pp. 1–15, 2017.

[10] J. M. Guo, H. Prasetyo, and H. Su, "Image indexing using the color and bit pattern feature fusion," *J. Vis. Commun. Image Represent.*, vol. 24, no. 8, pp. 1360–1379, 2013.

[11] P. Srivastava and A. Khare, "Utilizing multiscale local binary pattern for content-based image retrieval," *Mutimedia Tools and App.,* DOI 10.1007/s11042-017-4894-4, pp. 1-27, 2017.

[12] C. Singh and K. Preet Kaur, "A fast and efficient image retrieval system based on color and texture features," *J. Vis. Commun. Image Represent.*, vol. 41, no. October, pp. 225–238, 2016.

[13] M. V. Lande, P. Bhanodiya, and P. Jain, "An Effective Content-Based Image Retrieval Using Color, Texture and Shape Feature," *Intelligent Computing, Networking, and Informatics, Advances in Intelligent Systems and Computing 243, DOI: 10.1007/978-81-322-1665-0_119,* pp. 1163–1170, 2014.

[14] M. Arevalillo-Herráez, F. J. Ferri, and S. Moreno-Picot, "A hybrid multi-objective optimization algorithm for content based image retrieval," *Appl. Soft Comput.*, vol. 13, no. 11, pp. 4358–4369, 2013.

[15] M. Mosbah and B. Boucheham, "Distance selection based on relevance feedback in the context of CBIR using the SFS meta-heuristic with one round," *Egypt. Informatics J.*, vol. 18, no. 1, pp. 1–9, 2017.

[16] E. Walia, S. Vesal, and A. Pal, "An Effective and Fast Hybrid Framework for Color Image Retrieval," *Sens Imaging*, vol. 15, no. 1, pp. 1-23, 2014.

[17] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233-1244, 1996.

[18] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image Vis. Comput.*, vol. 27, no. 6, pp. 658–665, 2009.

[19] D. Pandey, "An Efficient Low Level Features of CBIR using Wavelet,

GLDM and SMOSVM Method," *Int. J. of Signal Processing, Image Processing and Pattern Recog.,* vol. 10, no. 3, pp. 13–22, 2017.

[20] C. S. Parkhi and A. Gupta, "A comparative implementation of Content Based Image Retrieval techniques in Wavelet and Cosine domains," *IOSR J. Electron. Commun. Eng.*, vol. 9, no. 2, pp. 25–30, 2014.

[21] N. Neelima, E. Sreenivasa Reddy, and N. Kalpitha, "An Efficient QBIR System Using Adaptive Segmentation and Multiple Features," *Procedia Comput. Sci.*, vol. 87, pp. 134–139, 2016.

[22] P. Sharma, "Improved shape matching and retrieval using robust histograms of spatially distributed points and angular radial transform," *Opt. - Int. J. Light Electron Opt.*, vol. 145, pp. 346–364, 2017.

[23] P. Vikhar, R. Scholar, and P. Karde, "Improved CBIR System using Edge Histogram Descriptor (EHD) and Support Vector Machine (SVM)." in *Proc. of IEEE Int. Conf. on ICT in Business, Industry and Government (ICTBIG),* Indore, India, pp. 1-5, 2016.

[24] R. Usha and K. Perumal, "Content Based Image Retrieval using Combined Features of Color and Texture Features with SVM Classification," *Int. J. of Comp. Science & Comm. Networks* vol. 4, no. 5, pp. 169–174.

[25] K. S. Arun, V. K. Govindan, "A Hybrid Deep Learning Architecture for Latent Topic-based Image Retrieval," *Data Sci. Eng.*, vol. 3, no. 2, pp. 166–195, 2018.

[26] R. R. Saritha, V. Paul, and P. G. Kumar, "Content based image retrieval using deep learning process," *Cluster Comput.*, no. February, pp. 1–14, 2018.

[27] M. Mansourvar, S. Shamshirband, R. G. Raj, and R. Gunalan, "An Automated System for Skeletal Maturity Assessment by Extreme Learning Machines," *PLOS One,* doi:10.1371/journal.pone.0138493, pp. 1–14, 2015.

[28] Y. Kaya, L. Kayci, R. Tekin, and Ö. Faruk, "Evaluation of texture features for automatic detecting butterfly species using extreme learning machine," *J. of Experimental & Theoretical Artificial Intelligence,* vol. 26, no. 2, pp. 37–41, 2015.

[29] R. Fu, B. Li, Y. Gao, and P. Wang, "Content-Based Image Retrieval Based on CNN and SVM," In *Proc. of IEEE Int. Conf. on Computer and Comm., Chengdu, 14-17*, pp. 1-5, 2016.

[30] K. Hemachandran, A. Paul, and M. Singha, "Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram," *IET Image Process.*, vol. 6, no. 9, pp. 1221–1226, 2012.

[31] S. M. Singh and K. Hemachandran, "Content- Content - Based Image Retrieval using Color Moment and Gabor Texture Feature," *Int. J. of Comp. Science Issues*, vol. 9, no. 5, pp. 299–309, 2012.

[32] W.D. Carlos, M.C.R. Renata and A. L. B. Candeias, "Texture classification based on co-occurance matrix and self-organizing map," In *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, Istanbul, Turkey, 10-13, October, pp. 2487–2491, 2010.

[33] M. P. Kushwaha and R. R. Welekar, "Feature Selection for Image Retrieval based on Genetic Algorithm," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 4, pp. 16–21, 2016.

[34] J. Annrose and C. S. Christopher, "An efficient image retrieval system with structured query based feature selection and filtering initial level relevant images using range query," *Optik*, vol. 157, pp. 1053–1064, 2018.

[35] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 3, pp. 6–14, 2016.

[36] S. Ding and X. Xu, "Extreme learning machine: algorithm, theory and applications," *Neural comuting and Applications,* vol. 25, no. (3-4), pp. 1-11, June, 2013.

[37] A. Suliman and B. S. Omarov, "Applying Bayesian Regularization for Acceleration of Levenberg-Marquardt based Neural Network Training," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 1, pp. 68-72, 2018.

[38] S. Liu, H. Wang, J. Wu, and L. Feng, "Incorporate Extreme Learning Machine to content-based image retrieval with relevance feedback," In *Proc. of IEEE 11th World Congress on Intelligent Control and Automation,* Shenyang, China, 29 june-4 july, pp. 1010-1013, March 2015.

[39] G. Liu, "Content-Based Image Retrieval Based On Visual Attention And The Conditional Probability," In *proc. of Int. Conf. on Chemical, Material and Food Engg.*, pp. 838–842, 2015.

[40] S. Bhardwaj, G. Pandove, and P. K. Dahiya, "An Intelligent Multi-

resolution and Co-occuring local pattern generator for Image Retrieval," *EAI Endorsed Transactions on Scalable Information Systems,* vol. 4, no. 1, e1, pp. 1–12, 2019.

Shikha Bhardwa

She received her B.Tech degree in ECE from Kurukshetra University, Haryana, India in 2006, Master's degree in ECE from M.M University, Ambala, Haryana, India in 2009. Now, pursuing Ph.D in Electronics and Communication Engineering Department from DCRUST University, Murthal, Haryana, India. Since 2011, she is currently working as an Assistant Professor in University Institute of Engineering & Technology, Kurukshetra University, Haryana, India. Her major research areas are Image processing, Artificial intelligence, Machine learning etc.

Gitanjali Pandove

She received her Ph.D degree in ECE. She is working as Professor in ECE Deptt., in DCRUST, Murthal, Haryana, India. Her research areas are optical communication, digital image processing, power systems, etc. She has more than 70 publications in various international reputed journals.

Pawan Kumar Dahiya

He received his Ph.D degree in ECE. He is working as a Professor in ECE Deptt., in DCRUST, Murthal, Haryana, India. His research areas are Digital and Embedded System Design, ANPR, VANETs, Soft Computing Techniques, Image Processing, Internet of Things (IoT), etc. He has more than 75 publications in various national and international reputed journals.