Content Based Image Retrieval Using Pigeon Inspired Based Optimization and Artificial Neural Network

*N.Parvin
Adhiyaman Arts and Science College for Women, Uthangarai, Tamil Nadu, India.
*Corresponding author Email:srsurajvarsi@gmail.com

Abstract. Content-Based Image Retrieval (CBIR) is also known as Query by Image Content (QBIC) that presents the technologies allowing to organize digital pictures by their visual features. They are based on the application of computer vision techniques to the image retrieval problem in large databases. CBIR consists of retrieving the most visually similar images to a given query image from a database of images. Retrieval of images is based not on keywords or annotations but on features extracted directly from the image data. Therefore, in this research study, the final output is retrieved from the database using feature extraction from the input query images. In this research study, Haralick Features are extracted, which is also known as Gray-Level Co-Occurrence Matrix (GLCM) along with Local Binary Pattern (LBP), and histogram of oriented gradients (HOG) features. When more number of features are used, it will increase the complexity of classifier to achieve better results. Therefore, feature selection technique called Pigeon Inspired based Optimization (PIO) is used in this research study. In the testing process, when the user gives a query images, the process will be same as training process and finally, to extract the relevant images, the training images are taken from the dataset and compared with the query images using Artificial Neural Network (ANN) classifier. The experiments are carried out using publicly available dataset called WANG and compared with existing techniques in terms of accuracy, precision, recall and F-measure. Keywords: Artificial Neural Network; Content-Based Image Retrieval; Haralick Features; Pigeon Inspired based Optimization; Local Binary Pattern.

1. Introduction

The evolutions of recent techniques enhance the utilization of Internet, camera, and mobile phones. The received and shared multi-media data are increasing, and retrieving related images from a database is a difficult one [1]. The primary requirement of any image retrieval method are searching and sorting the images which are in visual semantic relation using the query image (QI) offered by the clients. Generally, the search engines work on the network recovery of images based on text approach that need caption as input data [2]. The clients submit queries by enrolling some keywords that are corresponding with the texts which are placed in the archive. The final outcomes are produced on the basis of similarities in keywords, and it is applied for dissimilar image content [3]. The dissimilarity in the views of individual's awareness and physical labelling is the important cause for creating outcomes which is not relevant. It is nearly not possible to relate the idea of manually labelling previous size image archives with massive amount images [4]. The alternate model for retrieval of images and testing applies an automated image annotation mechanism that tags images according to the image content. This model depends upon automated image annotation accuracy which identifies shape related details like texture, layout, spatial edges, and colour [5-6]. Major researches were carried out for enhancing the efficiency of automated image annotation, but the variation in visual point provides inferior impression about image retrieval (IR) process [7-8]. Content based image retrieval (CBIR) is a platform that surmounts the issues since they are depend upon visual examination of data which is considered as an element of QI. The QI can be fed as an input data under the mapping of images placed inside an archive, and resemblance in visual closeness with respect to image feature vector that offers a base for the identification of image with same content [9]. Here, low-level visual features are calculated from the query and comparing the features is validated to arrange the outcomes [10]. Query-By-Image Content (QBIC) and simplicity are the instance of IR technique depends upon the filtration of low-level visual semantics. Once the predefined models are successfully executed, CBIR as well as feature extraction models are implemented in different software's such as textile industry, remote sensing, armed forces, video realization, crime detection, and clinical image analysis [11]. Fig. 1 gives an outline of the fundamental concept and process of IR.
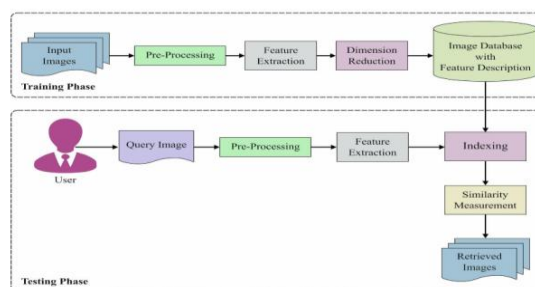
Figure 1. General Block of CBIR Framework

The fundamental requirement for some IR classification is to explore and arrange related image from the archive with least amount of individual interface with the device. This paper discussed election of visual features for a structure which depends upon the necessities of the customer [12]. The distinctive feature demonstration is highly required for any IR model. Moreover, the features can become highly effective and robust under the combination of low-level visual features, and maximum processing cost is necessary to accomplish better outcome. Unfortunately, the imbalanced feature selection limits the efficiency of IR process. Machine Learning (ML) model gets the feature vector as input data for training as well as testing methodologies that maximizes the performance efficiency [13]. Recently, the IR process is highly dependent on Deep Neural Networks (DNN) which is capable of providing optimal results at expensive platform.

Colour, texture, shape is the visual features of an image that are used to retrieve the visually similar content in CBIR. The traditional method which was based on keywords of an image has been demonstrated not worthy with respect to space and time complexities and that triggered the evolution of a new technique i.e. search by query instead of the keyword. It is a two-step practice where in the first step the image features are extracted to distinguishable term and the second step deals with the extracted features matching and the results are retrieved as per the visually similar images. CBIR mainly includes two steps: extraction of features and matching those features. Extraction of image features is done in the first step and the second step includes matching of those features.The rest of the paper is organized as follows: Section 2 presents the study of existing techniques; the explanation of proposed methodology is given in Section 3; the validation of proposed method using WANG database is presented in Section 4; the conclusion of the research study with future work is given in Section 5.

2. **Literature Review**

In this section, a study of existing CBIR techniques were presented that was used to retrieve the similar images from the database. In addition, the benefits of existing techniques along with limitations were presented as follows:

Li, et al., [14] aimed to achieve secure search for the encrypted image retrieval system, a new privacy preserving image retrieval system was developed which has the combination of asymmetric scalar-product-preserving encryption (ASPE) and homomorphic encryption (HE) schemes. Furthermore, the proposed scheme also utilized HE scheme to keep the secret key of ASPE confidential. In this scheme, trapdoor verification was also applied in searching phase to confirm the validation of trapdoor. Hence, through combination of ASPE and HE, this scheme provide a more secure image retrieval in cloud. In our scheme, each image was represented by the single vector. However, for the high dimensional descriptor, it leads to huge computation overheads, especially in executing the encrypted function.

Banharnsakun, et al., [15] a new efficient method was developed that was based on a combination of the gray-level co-occurrence matrix (GLCM) with the artificial bee colony (ABC), referred to as "GLCM-ABC," for CBIR. The objective of this work was to improve the accuracy rate of the image retrieval over other recently developed techniques. The results obtained from the proposed method showed that the hybrid GLCM-ABC approach offers good performance than other conventional methods. However, there was still room for improvement in the capability of the proposed method.

Alsmadi, et al., [16] an effective CBIR system with the application of genetic algorithm with simulating annealing was proposed in this study for the purpose of retrieving images from databases. Following the input of a query image from user, image features will be extracted from the image using the proposed CBIR system. In particular, YCbCr color with discrete wavelet transform and Canny edge histogram were used to extract color features, RGB color with neutrosophic clustering algorithm and Canny edge method were used to extract shape features, and GLCM was used to extract texture features. After that, images associated with query image were efficiently retrieved with the metaheuristic algorithm based similarity measure. The proposed CBIR system performed better than the existing methods and showed promising retrieval image results in terms of precision and recall rates in many groups of Corel image datasets as well as the proposed system has the highest overall performance results in terms of precision and recall rates compared with other existing systems. For future work, filtering techniques are proposed so that more accurate outcomes could be retrieved by the CBIR.

Garg, and Dhiman, [17] performed the multi-extraction in this analysis, which used PSO optimizer to remove most differentiating features from the extracted data. Classification tests were conducted on a COREL dataset consisting of ten categories and presented using four presentation parameters, i.e., precision, recall, Fmeasure, and accuracy. For validation purpose, three well-known classifiers were compared, i.e., support vector machines (SVM), decision tree (DT), and K-nearest neighbor (KNN). The proposed method consists of four steps. The first was decomposition, in which multi-scale decomposition was performed separately using DWT for channels R, G, and B. The second was concatenation of all three channels R, G, and B achieved from the set of functions. The third was reduction in features using the PSO algorithm to pick the most differentiating features. The last was classification where three classifiers were used to assess the category of images evaluated. Experimental results showed that SVM was the best optimizer which shows high parameter values of all the performance metrics. However; the feature dimension was small and needs a prominent calculation charge.

Chhabra, et al., [18] a new technique, namely Oriented Fast and Rotated BRIEF (ORB), has been proposed for CBIR. ORB and scale-invariant feature transform (SIFT) features were considered for effectively retrieval of content-based images from bulky dataset. Size of SIFT and ORB descriptor required a high memory space for storing features and high complexity, therefore, to reduce the space and complexity problem our system uses a K-means clustering algorithm and LPP over both descriptors. K-means reduce the descriptor into 32 clusters and LPP reduce into 4 and 8 components. Using 4- and 8-dimensional feature vector, we measure the precision, RMSE and time taken by the proposed CBIR system. Maximum precision rate of 86.20% and 99.53% has been accomplished for Wang dataset and corel dataset, respectively. We have also concluded that the proposed CBIR system performs better than already existing CBIR systems. The training time for

retrieving the data from bulky dataset by using a decision tree, random forest and MLP classifiers were higher than existing techniques due to more number of features, which requires optimal solution on features.

3. Proposed Methodology

CBIR consists of retrieving the most visually similar images to a given query image from a database of images. Retrieval of images is based not on keywords or annotations but on features extracted directly from the image data. Therefore, in this research study, the final output is retrieved from the database using feature extraction from the input query images. The research study has two process namely training process and testing process. In the training process, initially, the input data are taken from the WANG database and background are removed by using effective segmentation technique. Based on the texture and colour vector, the segmented images are converted into HSV images as well as gray-scale images. Then, these images are compressed by using Discrete Wavelet Transform (DWT) and given as input for extracting the features. In this research study, three different features are extracted. When more number of features are used, it will increase the complexity of classifier to achieve better results. Therefore, feature selection technique called Pigeon Inspired based Optimization (PIO) is used in this research proposal and selected the optimal features from the combined color and texture features. These optimal features are stored in the vector dataset during training process. In the testing process, when the user gives an query images, the process will be same as training process and finally, to extract the relevant images, the training images are taken from the dataset and compared with the query images using Artificial Neural Network (ANN) classifier. From the classification results, the most similar results are retrieved to the end-user. The workflow of the proposed methodology is given in the following Figure 2.
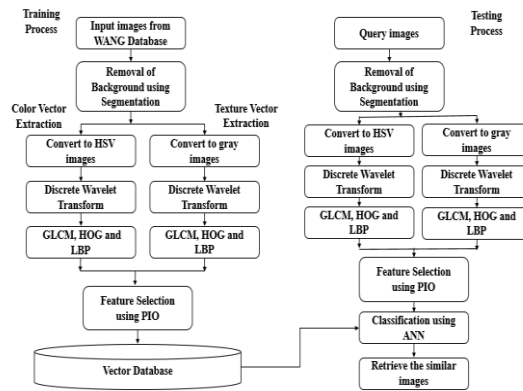
Figure 2. Workflow of the Proposed Methodology

Background Segmentation: To find edges of various objects present in an image using Canny, Sobel and fuzzy C means (FCM) techniques, we implemented an edge detection method that identifies all the edges in an input image by gradient magnitude approximation of the image. In case the objects' boundary is created with edges, we fill it in order to detect location of the object. If there are two objects that touch each other then we find the edges and use that information to separate the object. We can also use edges to find objects based on texture in certain situations where segmentation based on colour does not work very well. The edge detection method that we implemented convolves the input matrix with the three segmentation techniques and it outputs two gradient components of the image. On the other hand, the method can perform a thresholding operation on the gradient magnitudes and output a binary image which is a Boolean matrix with 1's being edges and 0's being other areas of image.

Sobel's edge detection,

The part of the derivatives of the Sobel operator are measured as:

$$G_x = (a_2 + 2a_3 + a_4) - (a_0 + 2a_7 + a_6) G_x = (a_2 + 2a_3 + a_4) - (a_0 + 2a_7 + a_6) \tag{1}$$

$$G_y = (a_6 + 2a_5 + a_4) - (a_0 + 2a_1 + a_2) G_y = (a_6 + 2a_5 + a_4) - (a_0 + 2a_1 + a_2) \tag{2}$$

The gradient magnitude is measured in the following equation (3):

$$|G| = \sqrt[2]{G_x^2 + G_y^2} |G| = \sqrt[2]{G_x^2 + G_y^2} \tag{3}$$

The orientation angle is measured as follows in the Eq. (4):

$$\theta = \arctan\left(\frac{G_x}{G_y}\right) - \frac{3\pi}{4} \theta = \arctan\left(\frac{G_x}{G_y}\right) - \frac{3\pi}{4} \tag{4}$$

Canny Edge Detection: The first step of the canny algorithm is to smooth image. Canny deduced the first derivative of Gaussian function, which is the best approximation of the optimal edge detection operator. Choose appropriate 1-d Gaussian function to smooth the image according to the row and column respectively, that is, execute convolution operation to image

matrix. Since the convolution operation satisfies commutative law and associative law, canny algorithm generally uses two dimensional Gaussian function (as shown in (5)) to smooth image and get rid of the noise.

$$G(x,y) = exp[-(x^2 + y^2)/2\sigma^2]/2\pi\sigma^2 G(x,y) = exp[-(x^2 + y^2)/2\sigma^2]/2\pi\sigma^2 \tag{5}$$

Where $\sigma\sigma$ stands for the parameter of Gauss filter, and it controls the extend of smoothing image.

Image Gradient Calculation: The second step is to calculate the magnitude and direction of image gradient. The Canny algorithm adopts limited difference of 2×2 neighbouring area to calculate the value and direction of image gradient. The first order partial derivative's approximation on the X and Y X and Y directions can be got from these following formulas:

$$E_x[i,j] = (I[i+1,j] - I[i,j] + I[i+1,j+1] - I[i,j+1]/2$$
$$E_x[i,j] = (I[i+1,j] - I[i,j] + I[i+1,j+1] - I[i,j+1]/2 \tag{6}$$

$$E_y[i,j] = (I[i+1,j] - I[i,j] + I[i+1,j+1] - I[i+1,j]/2$$
$$E_y[i,j] = (I[i+1,j] - I[i,j] + I[i+1,j+1] - I[i+1,j]/2 \tag{7}$$

Therefore, the templates of the image gradient calculation operator are:

$$G_x = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} G_x = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \tag{8}$$

$$G_y = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} G_y = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \tag{9}$$

The magnitude and direction of gradient can be calculated, where the image gradient magnitude is given in Eq. (10):

$$||M(i,j)|| = \sqrt{E_x[i,j]^2} + \sqrt{E_y[i,j]^2} ||M(i,j)|| = \sqrt{E_x[i,j]^2} + \sqrt{E_y[i,j]^2} \tag{10}$$

The azimuth of the image gradient is given in Eq. (11):

$$\theta(i,j) = arctan(E_y[i,j]/E_x[i,j]) \; \theta(i,j) = arctan(E_y[i,j]/E_x[i,j]) \tag{11}$$

Non-maximum Suppression (NMS): After acquired the gradient magnitude image $M[i,j], M[i,j],$ it's needed to perform non-maximum suppression on the image to accurately position edges. The process of NMS can help guarantee that each edge is one-pixel width. Canny algorithm uses 3×3 neighboring area which consists of eight directions to execute interpolation to the gradient magnitude along gradient's direction. If the magnitude $M[i,j] M[i,j]$ is bigger than the two interpolation results on the gradient direction, it will be marked as candidate-edge point, otherwise it will be marked as non-edge point. Therefore, the candidate edge image is acquired through the process.

Checking and Connecting Edges: The Canny algorithm adopts double-threshold method to select edge points after carrying on non-maximum suppression. The pixels whose gradient magnitude is above the high-threshold will be marked as edge points, and those whose gradient magnitude is under the low-threshold will be marked as non-edge points, and the rest will be marked as candidate edge points. Those candidate edge points who are connect with edge points will be marked as edge points. This method reduces the influence of noise on the edge of the final edge image.

Fuzzy C-Means (FCM): The FCM clustering algorithm was first introduced by Dunn and later was extended by Bezdek. Here, FCM segmentation technique is used for first, second and fourth objectives. The algorithm is an iterative clustering method that produces an optimal c (number of clusters) partition by minimizing the weighted within group sum of squared error objective function JFCM:

$$J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q d^2(x_k, v_i) J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q d^2(x_k, v_i) \tag{12}$$

Where $X = \{X_1, X_2, \dots, X_n\} \subseteq R_p X = \{X_1, X_2, \dots, X_n\} \subseteq R_p$ is the data set in the p-dimensional vector space, nn is the number of data items, cc is the number of clusters with $2 \leq c < n$ $2 \leq c < n$, $u_{ik} u_{ik}$ is the degree of membership of $x_k x_k$

in the ith cluster, q is a weighting exponent on each fuzzy membership, vi is the prototype of the centre of cluster ii, $d^2(x_k, v_i) d^2(x_k, v_i)$ is a distance measure between object x_k and cluster centre $v_i v_i$. After this background segmentation, the output images are converted into HSV format for extracting the color vectors and also converted into gray format for extracting the texture vectors. DWT is applied on the images that are described as follows.

Discrete Wavelet Transform: Wavelet Transform exhibit gradually changing oscillations punctuated with transients in the original signals. On the other way, images have plain area restricted by edges also can be called as abrupt changes. These abrupt changes consist of information. A powerful tool for data analysis is Fourier transform but it doesn't view abrupt changes efficiently as it views data as a sum of sine waves which are not localized in space or time. In order to perfectly analyse the abrupt changes in images and signals, Wavelets are used which can localize in time as well as frequency. Mainly there exists two class of wavelet transform, here Discrete Wavelet Transform is used in compression as well as denoising the images and signals. Wavelet transform of a 2D signal can be computed by recursive filtering and sub-sampling. Low frequency is represented by L and H denotes high frequency. In signal processing, wavelets make it possible to recover weak signals from noise also provides valuable advancement in quality of the image at higher compression ratios as compared with the conventional techniques.

Feature Extraction: The feature is defined as a function of one or more measurements, each of which specifies some quantifiable property of an object, and is so computed that it quantifies some significant characteristics of the object.

Colour Feature Extraction: Extracting the color feature of the images includes HSV histogram, Color moments, and Color auto-correlogram. HSV histogram process consist the conversion of RGB image into HSV color space, quantize the image to 8x2x2 and the unit sum is described to find normalized HSV histogram. In feature extraction of color moments process includes image analyzation and extract RGB color channels and compute the color moments for each channel using standard deviation (stdR, stdG, stdB), mean (meanR, meanG, meanB) and skewness (skeR, skeG, SkeB). The process of color auto-correlogram might include to integrate color histogram along with the spatial information.The feature extraction for proposed CBIR system have been considered first three color moments. Such as mean, standard deviation and skewness values are first, second and third order moment respectively. 9-D feature vector generated by this method. Mean consists average of the color value, standard deviation contain variance square root and the skewness asymmetric distribution. The color distribution values has been calculated by

$$Mean = A_l = \frac{1}{R}\sum_{m=1}^{R} S_{l,m} \quad Mean = A_l = \frac{1}{R}\sum_{m=1}^{R} S_{l,m} \tag{13}$$

$$Standard\ deviation = \partial_l = \sqrt{\left(\frac{1}{R}\sum_{m=1}^{R}(S_{l,m} - A_l)^2\right)}$$

$$Standard\ deviation = \partial_l = \sqrt{\left(\frac{1}{R}\sum_{m=1}^{R}(S_{l,m} - A_l)^2\right)} \tag{14}$$

$$Skewness = W_l = \sqrt[s]{\left(\frac{1}{R}\sum_{m=1}^{R}(S_{l,m} - A_l)^3\right)}$$

$$Skewness = W_l = \sqrt[s]{\left(\frac{1}{R}\sum_{m=1}^{R}(S_{l,m} - A_l)^3\right)} \tag{15}$$

Texture Feature Extraction: Texture is one of the most important defining characteristics of an image. The image texture depends on the scale or resolution at which it is displayed. A texture with specific characteristics in a sufficiently small scale could become a uniform texture if it is displayed at a larger scale. The Gray-Level Co-occurrence Matrix (GLCM) seems to be a well-known statistical technique for feature extraction [19]. The GLCM is a tabulation of how often different combinations of pixel gray levels could occur in an image. The goal is to assign an unknown sample image to one of a set of known texture classes. Textural features can be scalar numbers, discrete histograms or empirical distributions. They characterize the textural properties of the images, such as spatial structure, contrast, roughness, orientation, etc and have certain correlation with the desired output.

GLCM proposed by Haralick has become one of the most well-known and widely used texture measures. Haralick Features describe the correlation in intensity of pixels that are next to each other in space. Haralick proposed fourteen measures of textural features which are derived from the co-occurrence matrix a well-known statistical technique for texture feature extraction. It contains information about how image intensities in pixels with a certain position in relation to each

other occur together. For each matrix, the fourteen features like Angular Second Moment, Contrast, Correlation, Sum of Squares or Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation and Cluster Tendency are obtained. The homogeneity, contrast, entropy and energy are sensitive to the choice of the direction. The homogeneity and entropy supplies the indication on the dominancy values of the main diagonal on the basis of the frequencies of the problem. The energy supplies the information on the randomness of the spatial distribution. Additionally, sum of HOG features and sum of LBP features are extracted from the segmented image. The HOG, LBP features are cascaded with the GLCM features. The combined features are used for training and testing for the machine learning network. However, the more number of features results poor classification results of CBIR framework. Therefore, optimal features are required for better overall accuracy, where the feature selection technique is proposed in this research study that are explained along with comparative algorithms as follows:

Feature Selection Technique: In this study, PIO feature selection technique is used as proposed technique in this research study, where the comparative algorithms namely GA, and WOA are also presented here:

Pigeon Inspired Optimization Algorithm:PIO algorithms have recently been exposed to be effective in solving various optimization issues, including aerial robot trajectory planning, three-dimensional trajectory planning, an automatic landing system, and a PID development controller. In this article, we adopt the Learning rate selection on ANN network based on the fresh binary version of PIO. This unit offers two versions of the PIO. The first version or algorithm uses a sigmoid function to sample the speed of the doves, the second forms offers an updated improved binary version of the basic PIO, which uses cosine similarity to determine the speed of the doves. Both versions use an equal fitness function, another each form has methods that represent a pigeon or a solution.

Fitness Function:It is the terms of a process for evaluating the sufficiency of solutions. The fitness function evaluates the solution, which is a subset of the functions selected according to the true positive speed (TPS), the false positive speed (FPS), and the sum of functions. The sum of functions is involved in the adaptation function, so, if there is a few function that does not disturb the TPS or FPS, we want to avoid it. Eq. 16 represents the formula used to calculate the taste of a dove or

solution. Here is the sum of objects chooses, the total number of objects in SF and NF is $w1 + w2 + w3 = 1.$
$w1 + w2 + w3 = 1.$ The weight is set as follows: $w1 = 0.1, w2 = w3 = 0.45$
$w1 = 0.1, w2 = w3 = 0.45$, because TPS and FPS are equal.

$$FF = w_1 * \frac{SF}{NF} + w_2 * FPS + w_3 * \frac{1}{TPS} FF = w_1 * \frac{SF}{NF} + w_2 * FPS + w_3 * \frac{1}{TPS} \quad (16)$$

Sigmoid PIO for FS: Defines a solution or a pigeon vector of length equal to the sum of training data. Since the basic PIO procedure continuously processes the dove's position, the specific PIO solution for the learning rate is defined as a vector whose values of velocity and position vectors are fixed randomly among initial [0, 1]. The traditional method is used to measure the rapidity of every pigeon according to Equ. (16), and then the sigmoid function is used to translate the velocity into a binary version according to Equation 17.

    For the binary files of the cluster intelligence algorithm, the location of each dove is updated based on the value of the sigmoid function and the probability of a uniform random numeral between [0, 1] according to Equation 18. The algorithm will act as an old PIO, except for updating the position of the ground operator. Additionally, the sigmoid function will be used to transferal the speed, and then the locations will be informed accordingly.

$$S(V_i(t)) = \frac{1}{1+e^{\frac{-\pi_j}{2}}} S(V_i(t)) = \frac{1}{1+e^{\frac{-\pi_j}{2}}} \quad (17)$$

$$X(t)_{(i,p)}[i] = \begin{cases} 1, & if(S(V_i(t)) > r) \\ 0, & otherwish \end{cases} X(t)_{(i,p)}[i] = \begin{cases} 1, & if(S(V_i(t)) > r) \\ 0, & otherwish \end{cases} \quad (18)$$

Classification: An inspired by the biological nervous system, an ANN is an information processing system which contains numerous processing neurons which are densely interconnected. Neural Networks are a form of multiprocessor computer system with a high degree of interconnection, simple processing elements, adaptive interaction between elements and simple scalar messages. These neurons work together in a distributed manner:

    ❖ To be trained from the input information
    ❖ To manage inner processing
    ❖ To optimize its concluding outcome

The main advantage of using ANN is that it does not necessitate a priori detail of the image.By introducing ANN, algorithms have been developed for processing the CBIR analysis often become more intelligent than conventional techniques. The

purpose of a neural network is to map an input into a desired output. To resolve highly complex dilemmas, neurons can be combined in layers in artificial neural networks. There are many kinds of neural networks available.

Models of Neural Network Used in the Proposed Research:

(i) Multilayer Perceptrons (MLP) / Feedforward Neural Network (FNN)

The most popular neural network type which belongs to basic types of neural network called feed forward neural networks comprises of a sequence of layers namely input, hidden and output layers. Input and output layers show inputs and outputs on the whole networks. Hidden layers may be more than one in a network between these two layers. Every consequent layer has a connection with the preceding layer. In general, all neurons in a layer are associated to all neurons in the adjacent layers through unidirectional links which are represented by connection weights. The information moves in only one direction-forwards in this type of network. Figure 3 presents the diagram of proposed ANN method.
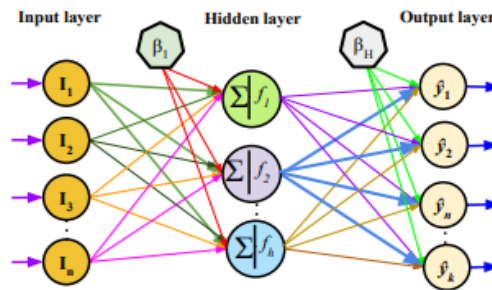

Figure 3. Structure of MLP Network

Feed forward neural networks (FFNNs) are known as a well-regarded class of ANN based neural models which are capable of realizing and approximating complex models based on their next-level, parallel, layered structure. The fundamental processing elements of FFNNs are a series of neurons. These neurons are disseminated over a number of fully-linked loaded layers. MLP is one of the widespread examples of FFNNs. In MLP, the initial processing elements are prearranged according to a one-directional manner. In these networks, evolution of information happens based on the communications among three types of matching layers: input, hidden, and output layers. Figure. 3 shows a MLP network that has a single hidden layer. The networks between these layers are associated with some weighting values varied inside $[-1,1]$ $[-1,1]$. Two functions can be carried out on every node of MLP, which are called summation and activation functions. The product of input values, weight values, and bias values can be attained based on the summation function described in Eq. (27).

$$S_j = \sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \qquad S_j = \sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \qquad (27)$$

Where nn denotes the total number of inputs, I_i is the input variable i, β_j is a bias value, and w_{ij} reveals the connection weight. In next step, an activation function is activated based on the outcome of the Eq. (27). Various activation approaches can be utilized in the MLP, which, according to literature, the most utilized one is S-shaped sigmoid function. This function can be calculated based on Eq. (28).

$$f_j(x) = \frac{1}{1+e^{-S_j}} \qquad f_j(x) = \frac{1}{1+e^{-S_j}} \qquad (28)$$

Therefore, the final output of the neuron jj is attained using Eq. (29)

$$y_i = f_j\left(\sum_{i=1}^{n} \omega_{ij} I_i + \beta_j\right) \qquad y_i = f_j\left(\sum_{i=1}^{n} \omega_{ij} I_i + \beta_j\right) \qquad (29)$$

After building the final structure of ANN, the learning process is instigated to fine tune and evolve the weighting vectors of network. These weighting vectors should be updated to approximate the results and optimize the total error of the network.

4. **Results and Discussion**

In this section, we have presented brief information about datasets and experimental results performed using the proposed framework for CBIR. The proposed algorithm is coded in MATLAB programming language. The system comprises of 4GB RAM and Intel i3 processor using Windows 10 operating system. All tests are carried out to evaluate the performance of the

overall system. In this proposed system, we used Wang database that comprises of 1000 images for 10 classes, including African people, food, buildings, elephant, beach, horse, flower, mountain, bus and dinosaurs. In Figure 4, we have shown a few samples of Wang database [21].
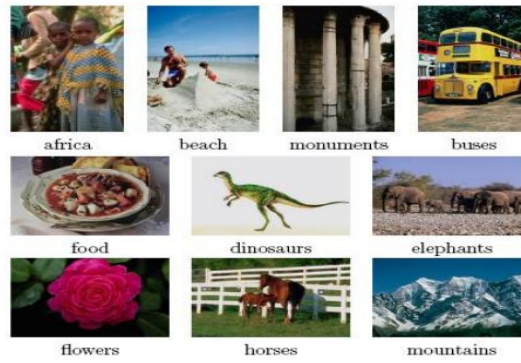

Figure 4. Sample Images of Wang Dataset [21]

A WANG dataset is available in size of 384 $\times$ 256 or 256 $\times$ 384 in JPEG format. This collection of data is also used to test other CBIR systems. It is commonly used because the dataset class information is strong in size and availability.

The classification of views can be calculated using four indexes considered based on the following calculations: precision, recall, F-measure, and accuracy. Some noted documents and most categories involve the utility of the solution offered based on the consistency of a particular class and the specifics of the rest of the documents. Therefore, detection of accuracy or recall value which determines the sensitivity parameter will effectively represent the accuracy of the text classifier. The equations to compute these measures are described as follows:

$$Precision = \frac{True\_Positive}{True\_Positive + False\_Positive} \quad (30)$$

$$Recall = \frac{True\_Positive}{True\_Positive + False\_Negative} \quad (31)$$

$$F - Measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (32)$$

$$Accuracy = \frac{True\_Positive + True\_Negative}{True_{Positive} + True_{Negative} + False_{Positive} + True\_Negative} \quad (33)$$

Where True Positive is suitably categorized subjects in a category, False Positive as appropriately classified subjects in a category and False Negative as correctly classified documents of rest categories.

In this research study, Root-mean-square error (RMSE) is also used to measure the error between training and testing data by comparing the expected value with the observed value.

Performance Analysis of Proposed Feature Selection Techniques: In this section, the proposed ANN classifier is validated with feature selection techniques namely GA, PIO and WOA algorithms in terms of accuracy, precision and recall that is described in Chart 1.
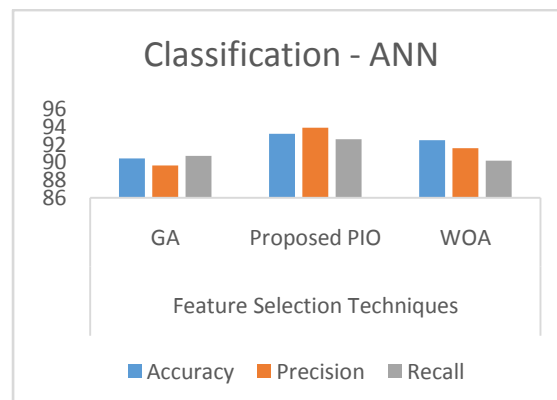
FIGURE 5. Performance Analysis of ANN classifier with Different Feature Selection techniques

From the chart 1, it is clearly stated that PIO has highest accuracy with ANN classifier, i.e. 93.21% of accuracy, where WOA achieved 92.49% of accuracy and GA achieved only 90.41% of accuracy with ANN classifier. While testing the proposed feature selection techniques with ANN classifier in terms of precision and recall, PIO has better performance than other techniques namely GA and WOA. For instance, PIO achieved 92.60% of recall and 93.89% of precision with ANN classifier, where GA and WOA achieved nearly 89% to 90% of recall and precision with ANN classifier. This validation results proved that ANN with PIO achieved better performance than ANN with GA as well as WOA algorithms.

Performance Analysis of Proposed Method: In this section, the performance of proposed feature selection techniques with ANN is compared with the existing techniques namely Decision tree (DT) [18], Random Forest (RF) [18], MLP [17], SVM [17] and KNN [17] in terms of all parameters. Initially, Chart 2 describes the comparison analysis on the basis of precision, recall and F-measure for 10 images.
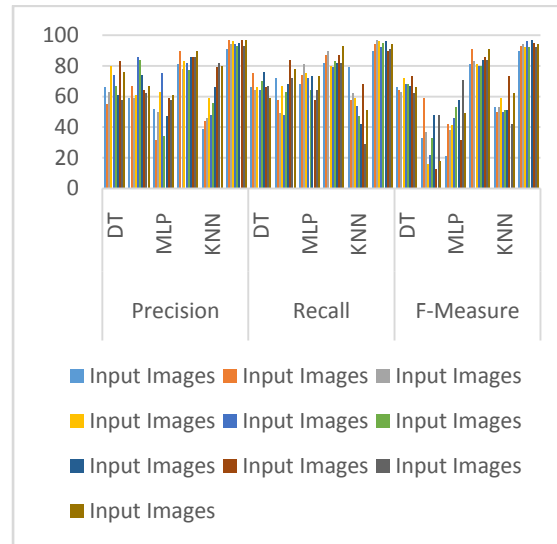

FIGURE 6. Comparative Analysis of proposed method

From the Chart 2, it is clearly stated that our proposed algorithm achieved better performance than various existing techniques. Among the existing techniques, MLP achieved poor performance, i.e. 53.1% of average precision and KNN achieved 59.9% of average precision. DT and RF achieved nearly 69% of average precision, where SVM achieved 83.9% of average precision. But, the proposed ANN achieved 94.7% of average precision and the reason is that proposed FS is included in this research study. The optimal features are selected by HFS techniques for better retrieval performance. For the recall analysis, the existing techniques such as DT, RF and MLP achieved nearly 67% to 70% of average performance, where KNN achieved very low recall value. The SVM achieved only 84.5% of average recall without feature selection techniques. By using proposed FS, the proposed ANN achieved 93.7% of average recall for all images. As like recall analysis, SVM achieved 84% of average F-measure and KNN achieved only 54.4% of average F-measure. However, RF and MLP achieved very low F-measure for all input images, where DT achieved 66.9% of average F-measure. The reason for poor performance is that the existing techniques didn't use the optimal features for retrieving the query images from the database. In this study, proposed FS techniques is used along with ANN and therefore, it is achieved 93.5% of average F-measure and proved that it achieved better performance than all other existing techniques.

Comparative Analysis of Proposed Classifier: In this section, the performance of proposed FS with ANN is compared with other existing techniques in terms of accuracy and RMSE for overall input WANG images, which is given in Chart 3.
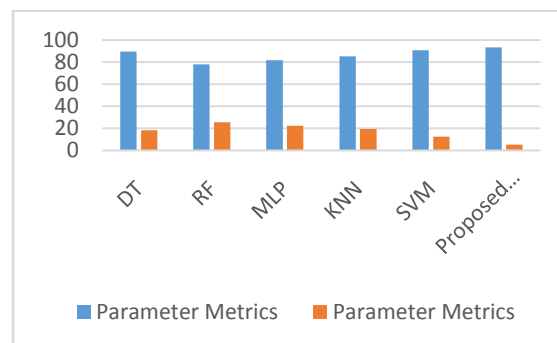

FIGURE 7. Comparative Analysis of Proposed Classifier

Accuracy is the ratio of true negatives and true positive ones to true negatives and true positive ones, and false negatives and false positives. This defines how much is rated correctly for a measure of the event. From the Chart 3, it is stated that DT,

SVM, KNN and MLP achieved nearly 89% to 90% of overall accuracy for CBIR query results. However, RF achieved very low classification accuracy and high RMSE value than any other existing techniques. In this research study, a proposed FS techniques are included with the ANN classifier to improve the query classification results and the results proved that proposed FS+ANN achieved 97.34% of overall accuracy and less RMSE i.e. 5.12. The next section will discuss the performance of proposed FS technique with various classifiers.

Comparative Analysis of Proposed Feature Selection Technique: Chart 4 presents the performance of proposed FS in terms of overall classification accuracy. Here, all the classifiers are implemented with proposed FS and results are taken for validating the performance of different classifiers.
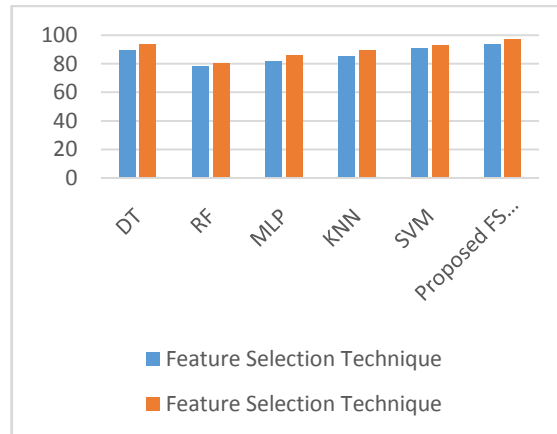

FIGURE 8. Comparative Analysis of Proposed HFS technique in terms of overall classification accuracy (%)

The existing and proposed classifier achieved better performance, while incorporating with proposed FS technique for better retrieval query images. The reason is that existing techniques and ANN requires optimal solution for better performance and also they achieved less performance without proposed FS technique. For instance, ANN and SVM achieved only 90% to 93% of overall classification accuracy. While proposed FS is included with these classifiers, every classifier achieved better performance. For example, ANN achieved 97.34% of accuracy and SVM achieved 93% of accuracy. When compared with every techniques, ANN achieved better performance this is due to learning of ANN is better than DT, RF, MLP, KNN and SVM. From the simulations, the implementation results proved that the proposed FS along with ANN achieved better retrieval accuracy.

5. Conclusion

CBIR uses the features that are, contents of an image like color, shape, texture instead of keywords of an image. Image retrieval systems basically are of two types namely, text based systems and content based systems. The text based method is a tedious task to do the annotation of a huge number of images on the basis of the keyword. There no need of doing manual annotation in CBIR system's and it retrieves the result in the form of visually alike images as per user's interest. The technique of efficiently retrieving similar images from the database is called as CBIR. In addition to contents of the image i.e. color, shape, texture, CBIR requires querying, matching, indexing and searching. In this research study, the final output is retrieved from the database using feature extraction from the input query images. The research study has two process namely training process and testing process. Based on the texture and colour vector, the segmented images are converted into HSV images as well as gray-scale images. Then, these images are compressed by using DWT and given as input for extracting the features. In this research study, three different features are extracted and when more number of features are used, it will increase the complexity of classifier to achieve better results. Therefore, proposed FS techniques are used for the optimal features from the combined color and texture features. In the testing process, when the user gives a query images, the process will be same as training process and finally, to extract the relevant images, the training images are taken from the dataset and compared with the query images using ANN classifier. From the classification results, the most similar results are retrieved to the end-user. The experimental results shown that the proposed ANN achieved only 93.34% of overall classification accuracy without FS techniques, where the same technique achieved 97.34% of overall classification accuracy, while incorporating with proposed FS techniques. In future, an ensemble classifier is required to implement for effective image retrieval by using huge amount of query images.

References

[1]. Rehman, M., Iqbal, M., Sharif, M. and Raza, M., 2012. Content based image retrieval: survey. *World Applied Sciences Journal*, *19*(3), pp.404-412.

[2]. Varma, N.M. and Mathur, A., 2020, September. A Survey on Evaluation of Similarity Measures for Content-based Image Retrieval Using Hybrid Features. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 557-562). IEEE.

[3]. Da Silva Torres, R. and Falcao, A.X., 2006. Content-based image retrieval: theory and applications. *RITA*, *13*(2), pp.161-185.

[4]. Zhang, H. and Su, Z., 2002, May. Relevance feedback in CBIR. In *Working Conference on Visual Database Systems* (pp. 21-35). Springer, Boston, MA.

[5]. Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S. and Jeon, G., 2018. Content based image retrieval by using color descriptor and discrete wavelet transform. *Journal of medical systems*, *42*(3), p.44.

[6]. Saadatmand-Tarzjan, M. and Moghaddam, H.A., 2007. A novel evolutionary approach for optimizing content-based image indexing algorithms. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *37*(1), pp.139-153.

[7]. Zhao, T., Lu, J., Zhang, Y. and Xiao, Q., 2008, May. Feature selection based on genetic algorithm for cbir. In *2008 Congress on Image and Signal Processing* (Vol. 2, pp. 495-499). IEEE.

[8]. Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N.I., Zafar, B., Dar, S.H., Sajid, M. and Khalil, T., 2019. Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, *2019*.

[9]. Hiremath, P.S., Shivashankar, S. and Pujari, J., 2006. Wavelet based features for color texture classification with application to CBIR. *International Journal of Computer Science and Network Security*, *6*(9A), pp.124-133.

[10]. Muneesawang, P. and Guan, L., 2001, October. A neural network approach for learning image similarity in adaptive CBIR. In *2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No. 01TH8564)* (pp. 257-262). IEEE.

[11]. Baig, F., Mehmood, Z., Rashid, M., Javid, M.A., Rehman, A., Saba, T. and Adnan, A., 2020. Boosting the performance of the BoVW model using SURF–CoHOG-based sparse features with relevance feedback for CBIR. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, *44*(1), pp.99-118.

[12]. Arjunan, R.V. and Kumar, V.V., 2009, October. Image Classification in CBIR systems with color histogram features. In *2009 International Conference on Advances in Recent Technologies in Communication and Computing* (pp. 593-595). IEEE.

[13]. Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y. and Li, J., 2014, November. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166).

[14]. Li, J.S., Liu, I.H., Tsai, C.J., Su, Z.Y., Li, C.F. and Liu, C.G., 2020. Secure content-based image retrieval in the cloud with key confidentiality. *IEEE Access*, *8*, pp.114940-114952.

[15]. Banharnsakun, A., 2020. Artificial bee colony algorithm for content-based image retrieval. *Computational Intelligence*, *36*(1), pp.351-367.

[16]. Alsmadi, M.K., 2020. Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features. *Arabian Journal for Science and Engineering*, pp.1-14.

[17]. Garg, M. and Dhiman, G., 2020. A novel content based image retrieval approach for classification using glcm features and texture fused lbp variants. *Neural Comput Appl*.

[18]. Chhabra, P., Garg, N.K. and Kumar, M., 2020. Content-based image retrieval system using ORB and SIFT features. *Neural Computing and Applications*, *32*(7), pp.2725-2733.

[19]. Porebski, A., Vandenbroucke, N. and Macaire, L., 2008, November. Haralick feature extraction from LBP images for color texture classification. In *2008 First Workshops on Image Processing Theory, Tools and Applications* (pp. 1-8). IEEE.

[20]. Hamamoto, A.H., Carvalho, L.F., Sampaio, L.D.H., Abrão, T. and Proença Jr, M.L., 2018. Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Systems with Applications*, *92*, pp.390-402.

[21]. Al-Rawi, S.S., Sadiq, A.T. and Shafeeq, A.F., 2013. Content based Image Retrieval using Combination between Moment and DCT Methods. *International Journal of Computer Applications*, *83*(17).