International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Fusion of Colour, Texture, and Shape Features with Supervised Learning Model for Content Based Image Retrieval

[1]Shilpa Marathe, [2]Sirshendu Arosh

Submitted:21/03/2023 **Revised:**25/05/2023 **Accepted:**09/06/2023

Abstract*:* To overcome a challenge in the field of imaging, Content-based Image Retrieval (CBIR) is used to find digital images in large datasets. When distinct functionalities are employed separately, the majority of present imaging systems provide less accuracy. Shape, texture, and colour are examples of low-level characteristics that are used to store various sets of models in the database. Based on the query images, related categories of images are then fetched. This paper proposes the hybrid approach of different shape, texture (cartoon feature) and colour feature. Further fuse features will be selected by neighbourhood Component Analysis (NCA) for machine learning i.e. SVM training. Validation of simulation results is achieved by using several databases. Experiments have shown that the accuracy of a NCA selected features in Corel dataset is up to 96%. The simulation results show strong performance based on recall, precision, accuracy, and F-score.

Keywords*: Content-based Image Retrieval; Cartoon texture; F-Score; Accuracy; Neighbourhood Component Analysis; Support Vector Machine.*

1. Introduction

The process of searching image libraries for comparable photos of a certain sort is known as image retrieval. Recent years have seen an increase in interest among scholars in practical applications of image retrieval. In retrieval systems, CBIR is a frequently used approach [1]. Visual attributes—often referred to as features—that are taken from the database are used by CBIR to obtain images. The chosen visual characteristic has a considerable impact on the retrieval accuracy and efficiency of the CBIR system. Shape [2], colour [3, 4], spatial connectivity [5, 6], and texture [7, 8] are a few examples of low-level properties. Previous CBIR research has placed a strong focus on capturing the spatial arrangement of colours because of the challenge of differentiating a large amount of colors. Effective image retrieval cannot be achieved using the local attributes of a picture, such as its texture, color, shape, and spatial arrangement, as employed by CBIR algorithms. Because when two objects with similar colours but different semantic categories are involved, the CBIR approach based on colour characteristics performs differently. Additionally, colour attributes disregard the spatial properties of the objects in an image. Additionally, they become noisy throughout the picture-taking process, which results in the return of irrelevant photos during the image retrieval step and lowers CBIR performance. Although shape-based qualities are frequently corrupted by noise, flaws, arbitrary distortion, and occlusion, they are more in line with human high-level perception [7]. Direction, coarseness, contrast regularity, and roughness are among adjectives that may be used to represent texture-based features, which show spatial variations across collections of pixels [8]. Segmenting the texture is still a challenge when trying to construct high-level semantic concepts [9]. Visual similarity is the cornerstone of CBIR. When feature vector values from several semantic categories are similar, CBIR performance deteriorates because pictures with no semantic link are retrieved [10].

An image is classified using a feature descriptor to enable comparison of visual similarity between images by quantifying the degree of visual content similarity. The similarity index is used to determine which images should be retrieved by comparing the feature vector of an inquiry image to the feature vectors of images stored in an image repository [7]. There are descriptions of regional and global visual aspects available. In contrast to global features, which separate an image's visual components based on local patches, local features for an image gather information on semantic similarity at an abstract level from the entire image.

This research provides a novel CBIR approach for retrieving similar images that combines three features colour, shape and texture. Further all features are combined together with NCA (neighbourhood component

1School of Interdisciplinary Studies and Resarch DY Patil International University, Akurdi Pune-411044, Maharashtra Email: shilpa.marathe@dypiu.ac.in
2Department of Electrical Engineering Asansol Engineering College (AEC) Asansol-713305, West Bengal Email: sirshenduarosh@gmail.com

analysis). The following are listed as the main offerings of this work:

- A novel CBIR method is presented, while finding related images, takes into consideration the visual characteristics contents of the original image.
- The suggested method classifies hybrid feature set images and extracts the features that are involved.
- Cartoon feature-based texture features are extracted and combined the shape feature and colour features with NCA for training and testing purpose.

2. Related Works

The problem of information retrieval, which might occur in the following ways, is resolved by the CBIR approach. Utilising just the traits or traits taken from the image, take or get an image similar to a vast collection from an attractive image [4]. Local characteristics, such as geographical domains, were assessed in [5] to demonstrate the impact of comparing photos with image catalogues. For the extraction of global characteristics, the segmentation approach is critically necessary.

Several studies [8] suggested using low-level visual features and descriptors to extract visual cues for obtaining related images. Local visual attention aspects were proposed by Yang et al. [9]. Visual attention features are extracted around the strongest salient locations using SURF features. Authors described a method for retrieving images in [7] that uses features from local binary patterns and colour information. Color information features can explain the colour information of important characteristics, whereas local binary patterns can extract textural features. The authors of [9] created a novel visual feature that transforms the input image into a normalised image using the discrete wavelet transform. It is transformed by using a colour space model. The normalised picture is divided into non-overlapping pieces using the edge image transformation and a colour space model. The authors of [10] suggested a unique framework that takes into consideration visual important components within a complicated context in order to improve accuracy. The authors of [11] proposed a method for generating a binary signature based on the colour and shape of objects of interest in an image. The retrieval results may be affected by rotation and scale. Multidimensional scaling can help overcome this issue by preventing image rotation [12, 13]. Furthermore, embedded and scene text recognition has recently received a lot of attention [14–16]. However, for detected text-based image retrieval, there is a very rare work [18, 19].

In order to discover similar images, authors of [17] devised a technique for identifying and recognising scene text. Using text in textual images as a query to locate further textual images that are comparable, the technique of [18] was put out. The authors of [19] provide a technique for text detection that combines edge cure with a variety of characteristics. CNN-based deep features and low level features are combined to achieve feature discrimination of textual portions. Only text from literary pictures may be discovered and recognised using these methods. It is essential to use recognisable text in CBIR applications in order to produce similar textual pictures. Other approaches that use semantic learning have also been reported.

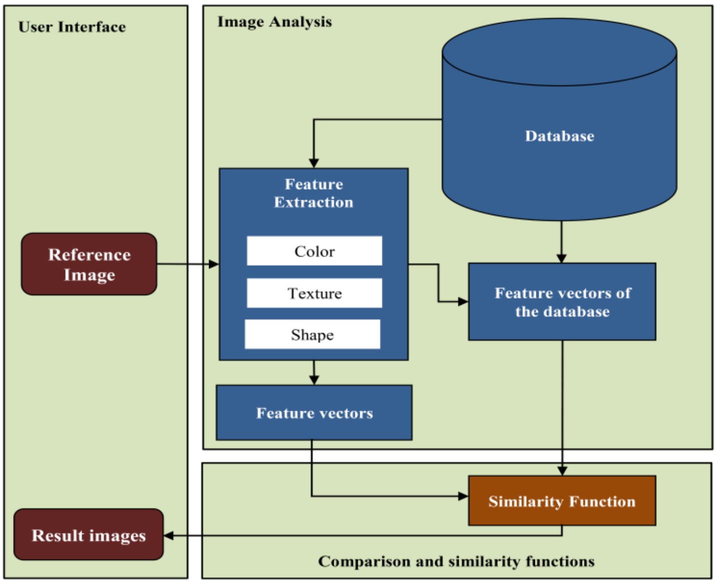


Fig 1: Schematic of a query in a CBIR System.

A unique method that takes into account hybrid textual visual relevance learning was presented by the authors of [20]. Social comments and tagging are combined with the extraction of visual characteristics. Authors of [21] unveiled a technique for assessing the usefulness of object and scene identifiers. Utilizing the annotated human description, the tags are determined. A technique for learning the feature paradigm from massive images and information was put out by authors of [22]. Deep learning will be primarily used to close the semantic gap between the social tags and the image. These systems have up to now depended on manually added, occasionally difficult to find descriptions and tags. The recommended research is distinctive in that pictures are retrieved using a classification-based method and that it is based on colour, texture, and form qualities. Based on an image categorization framework (training-testing model), the study models described in [7, 9, 23, 24].

3. Proposed Methodology

We will now describe the methods used to extract primitive functions from images, which are a key aspect of the CBIR system. Low-level image attributes include things like colour, form, and texture. Through the use of several feature extraction algorithms, the CBIR system retrieves features from an image [21]. All of the images from the repository are stored in a discrete database called a repository feature or database feature. The hybrid feature vector of the database image matches the hybrid feature vector of the query image, which is automatically extracted from both images.

3.1. Colour

By contrasting the colour histograms of the images in the database and the images in the query, it is possible to determine how similar the two submitted images are to one another. The likelihood that the randomly selected pixel p from image I has the colour that is, in accordance with the specification of a colour histogram for an image I, is [17]:

$$h_I(c_i) = P(color\ (p) = c_i | p \in I) \tag{1}$$

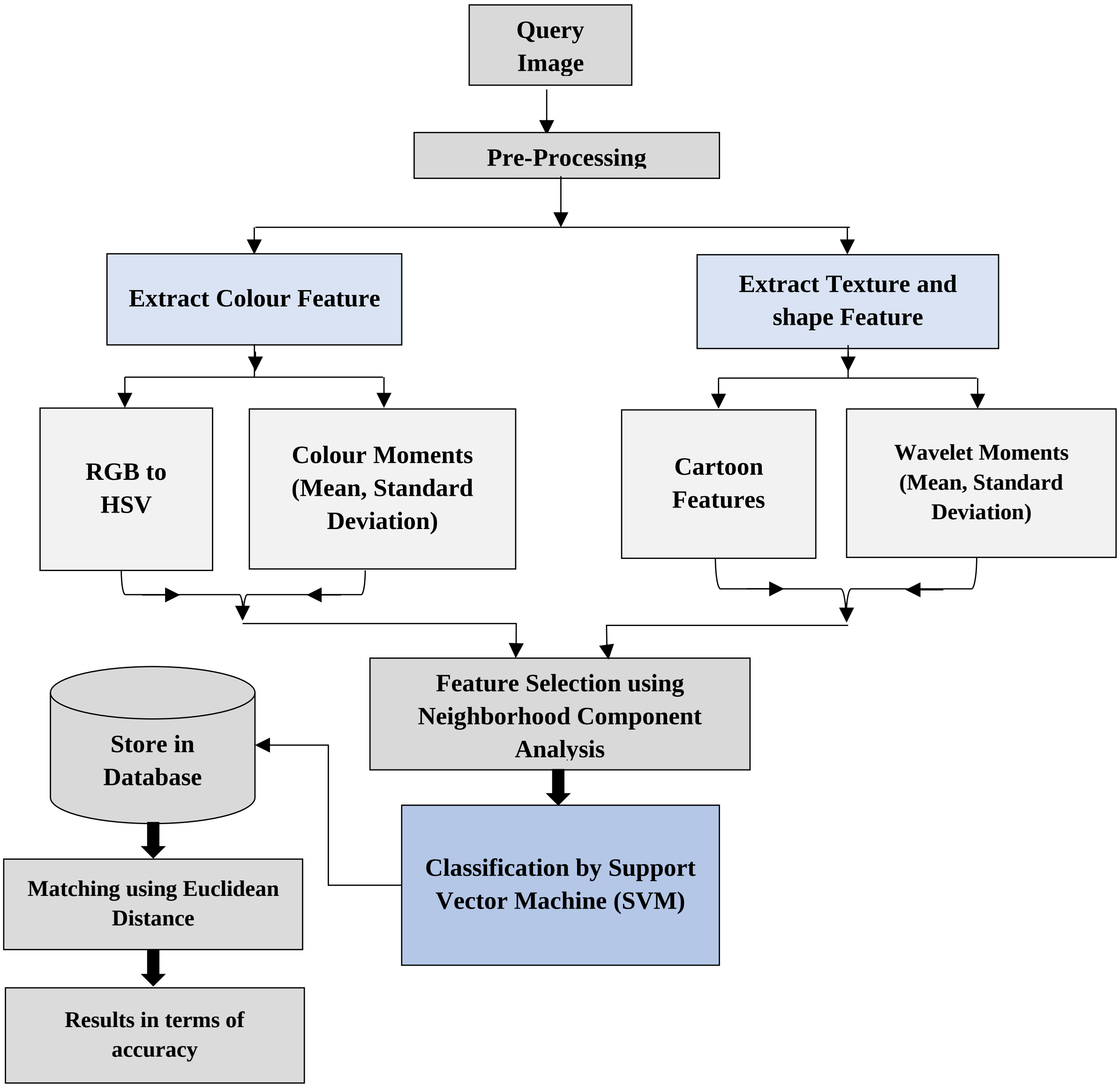

Query Image
Pre-Processing
Extract Colour Feature
Extract Texture and shape Feature
RGB to HSV
Colour Moments (Mean, Standard Deviation)
Cartoon Features
Wavelet Moments (Mean, Standard Deviation)
Feature Selection using Neighborhood Component Analysis
Store in Database
Classification by Support Vector Machine (SVM)
Matching using Euclidean Distance
Results in terms of accuracy

Fig 2: System architecture of proposed CBIR system

3.2. Texture

The texture of the image corresponds to the phenomenon of perception, easily recognizable by a person, but difficult to describe mathematically. For example, grass and rose petals vary in texture due to their softness and pattern. Texture refers to a visual pattern that displays a uniform nature that is not the result of a single color or intensity [18].

Characteristics of Image: The characteristics of image are total six that are contrast, coarseness, roughness, line-likeness, directionality, and regularity, and their authors found that there are three of them that strongly correlate with human perception: coarseness, contrast and directionality. In [19], authors have proposed, how to calculate these characteristics to obtain a scalar value for each processed image.

Cartoon features and Gabor Wavelet are extracted after the gray scale conversion of the colon biopsy images. RGB colon biopsy image is converted to a gray scale image using rgb2gray function. These features are then unified to form a texture feature set.

3.3. Texture Features Set

Gabor Wavelet: For an input image $I(x, y)$ of size $P \times Q$, the discrete Gabor wavelet transform is defined as the following convolution [20]:

$$G_{mn}(x,y) = \sum_{s}\sum_{t} I\,(x-s, y-t)\psi_{mn}^{*}(s,t)$$

(2)

Where, s and t are symbolized for the variables of filter mask size, and ψ_{mn}^{*} is symbolized for the complex conjugate of ψ_{mn}.

$$\psi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right].exp(j2\pi Wx)$$

(3)

Where W denotes the modulation frequency. Now the generating function:

$$\psi_{mn}(x,y) = a^{-m}\psi(\tilde{x},\tilde{y})$$

(4)

Here, the values n and m denote the orientation and wavelet's scale respectively.

And "$m = 0,1, \dots M-1, n = 0,1,\dots,N-1$"

$$\tilde{x} = a^{-m}(x\cos\theta + y\sin\theta)$$

(5)

$$\tilde{y} = a^{-m}(-x\sin\theta + y\cos\theta)$$

(6)

Where $a > 1$ and θ =nπ/N.

And,

$$a = (U_h/U_l)^{\frac{1}{M-1}}$$

(7)

$$W_{m,n} = a^m U_l$$

(8)

$$\sigma_{x,m,n} = \frac{(a+1)\sqrt{2\ln 2}}{2\pi a^m (a-1)U_l}$$

(9)

$$\sigma_{y,m,n} = \frac{1}{2\pi\tan\left(\frac{\pi}{2N}\right)\sqrt{\frac{U_h^2}{2\ln 2}-\left(\frac{1}{2\pi\sigma_{x,m,n}}\right)^2}}$$

(10)

Wavelet Moments: Following the application of Gabor filters to a photo of a certain orientation at a particular size, the following array is obtained [21]:

$$E(m,n) = \sum_{x}\sum_{y}|G_{mn}(x,y)|$$

(11)

Where, "$m = 0,1,\dots,M-1;\ n = 0,1,\dots,N-1$"

In the event where uniformly textured images or regions are what we are interested in, the mean and standard deviation are represented as follows:

$$\mu_{mn} = \frac{E(m,n)}{P \times Q}$$

(12)

$$\sigma_{mn} = \frac{\sqrt{\sum_x\sum_y(|G_{mn}(x,y)| - \mu_{mn})^2}}{P \times Q}$$

(13)

A feature vector f_g (texture representation) is created using μ_{mn} and σ_{mn} as the feature components [21]:

$$f_g = (\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01} \dots \dots .\mu_{45}, \sigma_{45})$$

(14)

3.4. Color Features Set

RGB to HSV Conversion Using the following formula, image pixel values are transformed from RGB representation:

$$H = \cos^{-1}\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}$$

(15)

$$S = 1 - \frac{3[\min{(R,G,B)}]}{R+G+B}$$

(16)

$$V = \left[\frac{R+G+B}{3}\right]$$

(17)

Color Moments

a. Mean

b. Standard Deviation

If there are N pixels in the image and the i^{th} colour channel value is the j^{th} image pixel equal to I_{ij}, then the area 'r' and the index entries for this colour channel exactly match the time suggested by the formula [22]:

Mean:

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

(18)

a. Standard Deviation:

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

(19)

Where $E_{r,i}(1 \le i \le 3)$ is the color mean of the r region, and $\sigma_{r,i}$ is the standard deviation of the r region. And the extracted color characteristics are specified by the following feature vector:

$$f_c = \{E_{1,1}, \sigma_{1,1} E_{2,2}, \sigma_{2,2} E_{3,3}, \sigma_{3,3} \dots\dots\dots\dots\dots . E_{r,i}, \sigma_{r,i}\}$$

(20)

Color Correlogram: A three-dimensional histogram that depicts the colour distribution and spatial correlation between colour pairings is called a colour correlogram. The histogram's first and second dimensions stand in for each pair of pixels' colour, while its third dimension measures their spatial separation [23].

The probability that a pixel of colour j is situated k pixels away from a pixel of colour I for the image is characterized by the k^{th} entry for (i,j). A colour correlogram may be thought of as an indexed table of colour pairings.

If H is a collection of color-coded pixels $c(j)$ and $H_{c(j)}$ is a set of image-related pixels, the correlogram of this image is defined as:

$$\gamma_{i,j}^{k} = p_r\left[p_2 \epsilon H_{c(j)}, |p_1 - p_2| = k\right]$$

(21)

Where, $i,j \in \{1,2,3,\dots,N\}$

$k \in \{1,2,3,\dots,d\}$

$|p_1 - p_2|$ symbolizes the distance between pixels p_1 and p_2, and

p_r represents the probability function.

3.5. Shape

This is a crucial characteristic of the segmented image region, and its effectiveness and toughness are crucial to its recovery. Depending on your application, you might need to express the form using displacement, rotation, or scale invariance.

The two categories of shape are border-based and region-based. The first is based on the shape's perimeter, while the second is based on the shape's overall area. Fourier descriptors, which use the Fourier transform's limits, and invariance moments, which use moments based on transform-invariant fields, are the two representations for these two categories that work best [24] [25].

Geometric moment and central moment: Define the two-dimensional moment of the geomorphic order $(p+q)$ of the density distribution function $f(x,y)$ as follows:

$$m_{pq} = \sum\sum x^p y^q f(x,y)$$

(22)

Moments with translation invariance are called central moments and are defined as:

$$\mu_{pq} = \sum\sum (x-\bar{x})^p (y-\bar{y})^q f(x,y)$$

(23)

Where x and y are symbols for the coordinates of the centroid of the image function $f(x,y)$.

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$$

(24)

Related invariant moments implemented:

$$Inv_1 = \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2)$$

(25)

$$Inv_2 = \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{13}\mu_{31} + 3\mu_{22}^2)$$

(26)

$$Inv_3 = \frac{1}{\mu_{00}^{7}}(\mu_{20}\mu_{21}\mu_{30} - 4\mu_{20}\mu_{12}^{2} - \mu_{11}\mu_{03}\mu_{30} + \mu_{11}\mu_{21}\mu_{12} + \mu_{02}\mu_{30}\mu_{12} - \mu_{02}\mu_{21}^{2})$$

(27)

$$Inv_4 = \frac{1}{\mu_{00}^{10}}(\mu_{30}^{2}\mu_{02}^{2} - 6\mu_{21}\mu_{12}\mu_{30}\mu_{03} - 4\mu_{30}\mu_{12}^{3} + 4\mu_{03}\mu_{21}^{3} - 3\mu_{21}^{2}\mu_{12}^{2})$$

(28)

3.6. Feature Extraction using Cartoon Descriptors

A pair of low-pass and high-pass filters applied on image I by the following minimization is the traditional method for obtaining the cartoon feature in images:

$$Cartoon\ Feature = \min_{u}\{\sigma^4 \int |Du|^2 + \|I - u\|_{H^{-1}}^{2}\}$$

(29)

The cartoon part of the image I is represented by u.

3.7. Color, Texture and Shape Features Fusion

Retrieval using multiple sources can be much more efficient, so this task uses a combination of the three primary sources.

Apart from the disparate nature of the information, the structure of the primitives is an added difficulty when it comes to combining all the information available.

The best of each will be obtained from the most basic data sets using this weighted sum approach, which is extensively used in the scientific world.

For every T-image in the dataset, given N evaluation primitives for the images $P_1, P_2, \dots\dots\dots, P_N$ and the query image Q, the difference between the two according to the primitive $\mathrm{P_K}$ is symbolised as $D_k(Q,T) \in [0,1]$. The following is the definition for the total distance D^c:

$$D^c(Q,T) = \sum_{i=1}^{N} w_i D_i(Q,T)$$

(30)

Each primitive $\mathrm{P_i}$ is given a weight $\mathrm{w_k}$ such that $0 < w_k \leq 1$ and $\sum_{i=1}^{N} w_i = 1$ are both true.

3.8. Selection of Features utilizing the Neighborhood Component Analysis (NCA)

The NCA is a classifier whose goal is to apply a kNN in a new space after finding a linear transformation for it from a dataset. The goal of the linear transformation is to have the kNN perform better in the new space than it did in the old space [26].

Due to the kNN's characteristic of partitioning space, the value of the function that calculates the kNN LOO classification error for a set, $\{x_1, \cdots, x_n\} - \{x_i\}$ is different from the value obtained for the set $\{x_1, \cdots, x_n\} - \{x_j\}$. This difference highlights a discontinuity in the LOO error function. To address the discontinuity problem of LOO classification error, NCA smooth the assignment of neighbors in the new space. Smoothing is done in such a way that the selection of which observation is next to each observation is no longer deterministic, as it was originally done in kNN, and becomes probabilistic. So, let p_{ij} be the probability of point i choose point j as its neighbor, the value of p_{ij} is given by,

$$p_{ij} = \begin{cases} \dfrac{\kappa\left(D(x_i, x_j)\right)}{\sum_{k \neq i} \kappa\left(D(x_i, x_j)\right)}, & i \neq j \\ 0, & i = j \end{cases}$$

(31)

In equation (31), $D(x_i, x_j) = \|Ax_i - Ax_j\|^2$ is the Euclidean distance in space transformed by A and $\kappa(z) = e^{-\frac{z}{\sigma}}$. The function $\kappa(z)$ is a kernel function with width defined by σ [27]. Let $C_i = \{j | c_i = c_j\}$ is the set of points with the same class as point i, the probability of this point being correctly classified by the algorithm is given by Equation (32),

$$p_i = \sum_{j \in C_i} p_{ij}$$

(32)

The A transformation is obtained by maximizing the probability of each point in the set being correctly classified, given by Equation (33):

$$f(A) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i$$

(33)

Differentiating this function from A, the following expression is obtained:

$$\frac{\partial f}{\partial A} = 2A \sum_i \left(p_i \sum_k p_{ik} x_{ik} x_{ik}^T - \sum_{j \in C_i} p_{ij} x_{ij} x_{ij}^T \right)$$

(34)

The two functions presented can be modified to include a regularization parameter [27]. The inclusion of a regularization parameter is done to minimize the classification error of new unseen observations, avoiding overfitting [26]. The new functions with regularization are defined by equations (35) and (36) where N is the number of elements of A, as:

$$f(A) = \sum_i p_i - \frac{\lambda}{N} \|A\|_F^2 \tag{35}$$

$$\frac{\partial f}{\partial A} = 2A \sum_i \left(p_i \sum_k p_{ik} x_{ik} x_{ik}^T - \sum_{j \in C_i} p_{ij} x_{ij} x_{ij}^T \right) - \frac{\lambda}{N} A \tag{36}$$

3.9. Support Vector Machine (SVM) based Classification

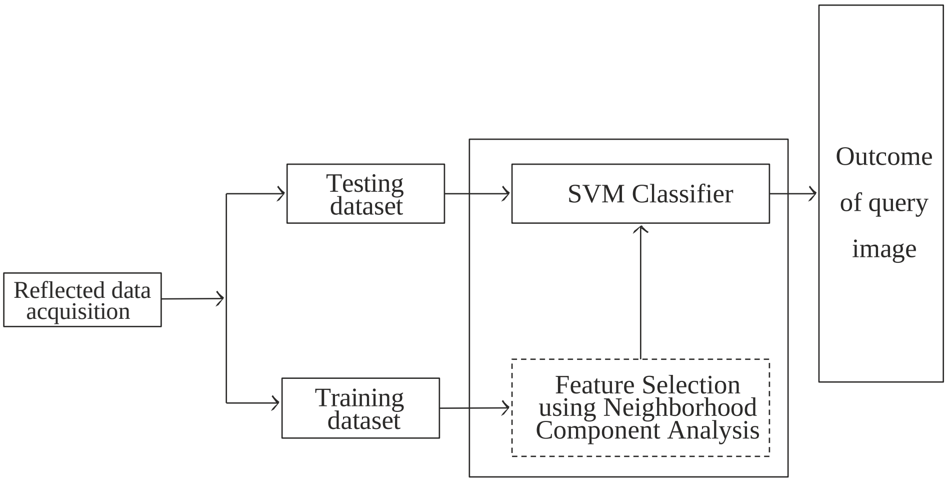

Reflected data acquisition
Testing dataset
Training dataset
SVM Classifier
Feature Selection using Neighborhood Component Analysis
Outcome of query image

Fig. 3: Working of support vector machine classifier

Classification is present in many real world problems, initially SVM were designed to handle binary (+/- 1) problems. Now let's see how to solve this problem. Following equation presents the objective function [15]:

$$w_r \in H, \in^r \in R^m, b_r \in R \frac{1}{2} \sum_{r=1}^{M} \|w_r\|^2 + \frac{c}{m} \sum_{i=1}^{m} \sum_{r \neq y_1} \varepsilon_i^r \tag{37}$$

Subject to:

$$\langle W_{y_i}, X_i \rangle + b_{y_i} \geq \langle W_r, X_i \rangle + b_r + 2 - \varepsilon_i^r, \quad \varepsilon_i^r \geq 0 \tag{38}$$

Where, $m \in \{1, \dots, M\} \backslash Y_i$ and $Y_i \in [1, \dots, M]$ is the multi-class label of the X_i pattern.

4. Simulation and Result Analysis

4.1. Description of Dataset

To evaluate our approach, Caltech 101 [28], Caltech 256 [28] databases were chosen for the experiment because of its large size and inter-class variability. Most of the images in this database are medium resolution. They are used by many researchers to evaluate their work.

4.1.1. Caltech 256

We used a different Caltech 256 database with 30,607 images in 256 categories (rings, brains, diamonds, camels, etc.), with a scale of 80 images as the minimum amount of images for each category [28].

Fig 4: Retrieved images of Caltech Dataset [28]

4.1.2. Corel Dataset

The dataset consists of 500 images from the Corel dataset, divided into 5 classes, with 100 images in each class. This setup allows for experimentation and evaluation of algorithms or techniques in the context of real-life images.

Having a dataset with a sufficient number of images per class (100 in this case) can provide a representative sample for each class and allow for robust evaluations. It is important to ensure that the dataset covers a diverse range of images within each class to capture the variations and characteristics of real-life images accurately.

The images are either 256×384 or 384×256 in size. Africa, Beach, Transportation, Architecture, and Dinosaur are the first five photo categories that are accessible. Each class's images have a unique number (Class 1: 0-99, Class 2: 100-199 etc.). In the literature, it is known as the Corel database and which is utilised for experimenting with the simplicity structure put out by the authors of [29]. This database is publicly available for experimentation [30] [31].

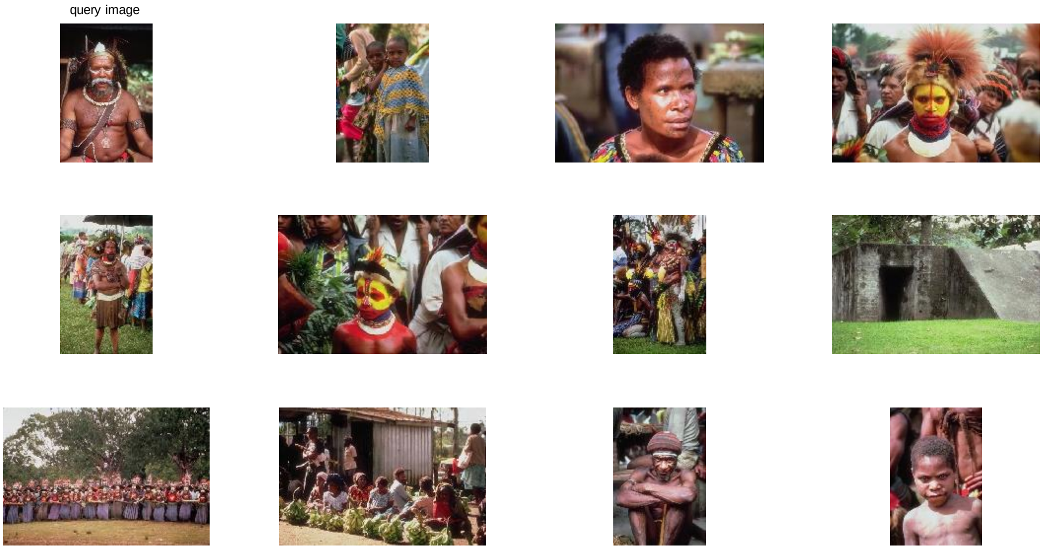


Fig. 5: Retrieved images of Africa Category from Corel Dataset [29]

When comparing Content-Based Image Retrieval (CBIR) systems, it is indeed important to evaluate their quality using appropriate metrics. However, due to the varied applications and contexts in which CBIR systems are used, it can be challenging to find a single metric that is universally applicable. There are a few key considerations to keep in mind when selecting metrics for evaluating CBIR systems.

Ease of Interpretation: The chosen metric should be easy to understand and interpret. It should provide a clear indication of the system's performance without requiring complex calculations or specialized knowledge. This is important to ensure that the results of the evaluation are accessible and meaningful to both technical and non-technical stakeholders.

User-Centric Quality: CBIR systems are developed to serve users' needs, so it is crucial to capture the quality notion as perceived by the users. Different user groups may have diverse requirements and expectations from a CBIR system. It is essential to consider the specific domain and the users' preferences when selecting metrics.

4.2. Evaluation Parameters

Table 1: Parameters of evaluation

"TP (True Positive)"	"Indicated the number of desired image that were classified as correctly classified"
"TN (True Negative)"	"Indicated the number of desired image that were classified as not classified correctly"
"FP (False Positive)"	"Indicated the number of desired image that were classified as incorrectly classified"
"FN (False Negative)"	"Indicated the number of desired image that were classified as not classified incorrectly"

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (39)$$

$$Precision = \frac{TP}{TP+FP} \quad (40)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (41)$$

$$F-Score = \frac{2TP}{2TP+FP+FN} \quad (42)$$

4.3. Simulation Results

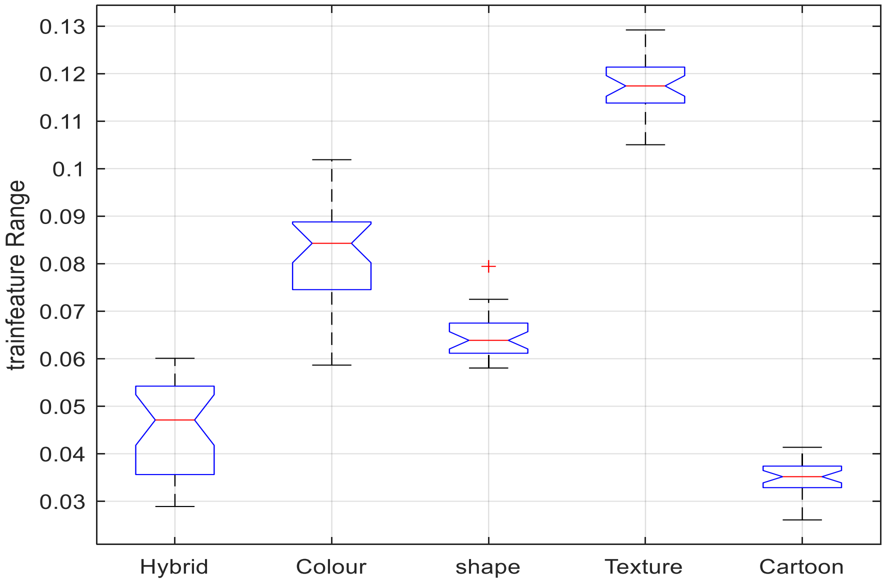


Fig. 6: Range of training features extracted by proposed hybrid method

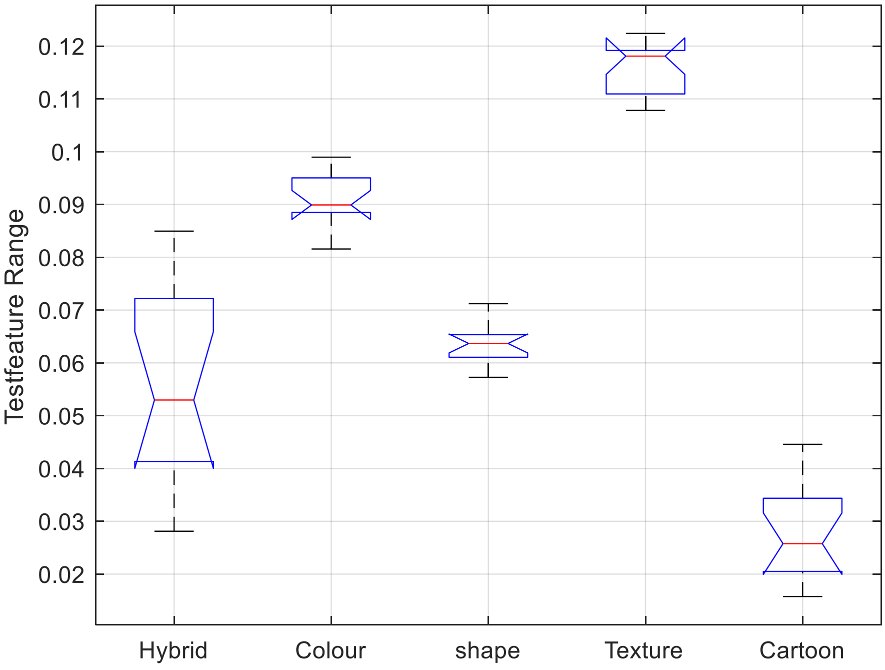


Fig. 7: Range of testing features extracted by proposed hybrid method

Figures 6 and 7's boxplots display the distribution of hybrid features across datasets, illustrate how well the system performed throughout training and testing. There are two main observations in this chart:

- The hybrid resource can provide a more representative and balanced training sample for the classifier. This is important because classifiers often perform better when trained on data that follows a more symmetrical distribution. Such data allows the classifier to learn patterns and make predictions more effectively.

- Hybrid features, shapes, textures, and boxes are less prone to graphic displacement based on hybrid features than color-A functions. The average range for hybrid features is 0.056 to 0.07 within this range. The distribution among all classes and the average weight of the tag table are almost the same.

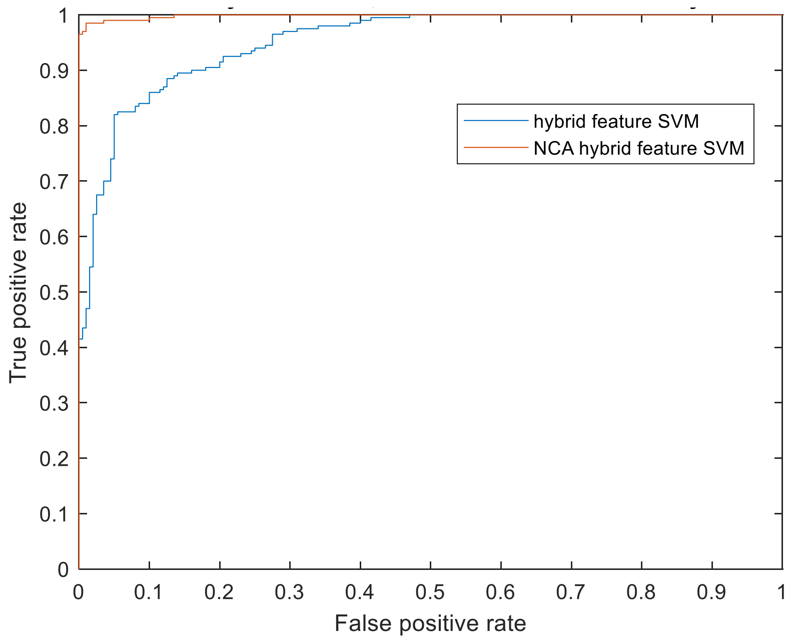
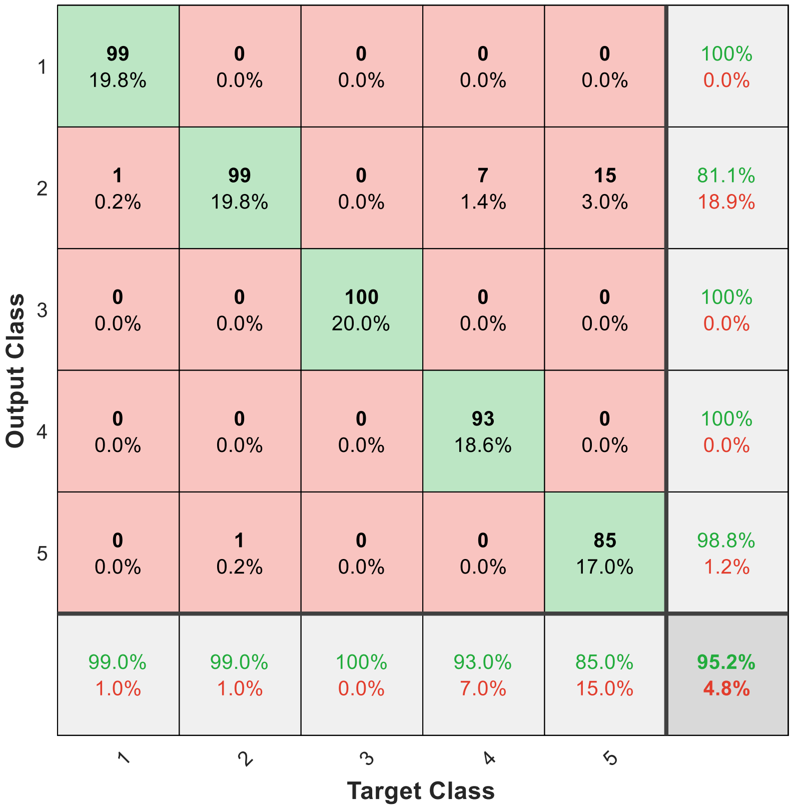


Fig. 8: ROC curve for hybrid features and NCA based hybrid features with SVM



	1	2	3	4	5	
1	**99** 19.8%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	100% 0.0%
2	**1** 0.2%	**99** 19.8%	**0** 0.0%	**7** 1.4%	**15** 3.0%	81.1% 18.9%
3	**0** 0.0%	**0** 0.0%	**100** 20.0%	**0** 0.0%	**0** 0.0%	100% 0.0%
4	**0** 0.0%	**0** 0.0%	**0** 0.0%	**93** 18.6%	**0** 0.0%	100% 0.0%
5	**0** 0.0%	**1** 0.2%	**0** 0.0%	**0** 0.0%	**85** 17.0%	98.8% 1.2%
	99.0% 1.0%	99.0% 1.0%	100% 0.0%	93.0% 7.0%	85.0% 15.0%	**95.2%** **4.8%**

Fig 9: Confusion matrix plot for Corel dataset with SVM approach

Figure 9 shows a confusion matrix plot for Corel dataset. As per the dataset, it exhibits five classes for the five image category. Calculation for each class is given as follows:

Table 2: Result for Corel dataset using SVM classifier

Parameter	1	2	3	4	5
TP	99	99	100	93	85
FP	1	1	0	7	15
FN	0	23	0	0	1
TN	400	377	400	400	399

For '1':

TP=99, TN=400, FP=1, FN=0

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{99+400}{99+400+1+0} = 99.8\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{99}{99+1} = 99\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{99}{99+0} = 100\%$$

$$F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 99}{2\times 99+1+0} = 99.5\%$$

For '2':

TP=99, TN=377, FP=1, FN=23

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{99+377}{99+377+1+23} = 95.2\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{99}{99+1} = 99\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{99}{99+23} = 81.15\%$$

$$F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 99}{2\times 99+1+23} = 89.19\%$$

For '3':

TP=100, TN=400, FP=0, FN=0

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{100+400}{100+400+0+0} = 100\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{100}{100+0} = 100\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{100}{100+0} = 100\%$$

$$F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 100}{2\times 100+0+0} = 100\%$$

For '4':

TP=93, TN=400, FP=7, FN=0

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{93+400}{93+400+7+0} = 98.6\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{93}{93+7} = 93\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{93}{93+0} = 100\%$$

$$F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 93}{2\times 93+7+0} = 96.37\%$$

For '5':

TP=85, TN=399, FP=15, FN=1

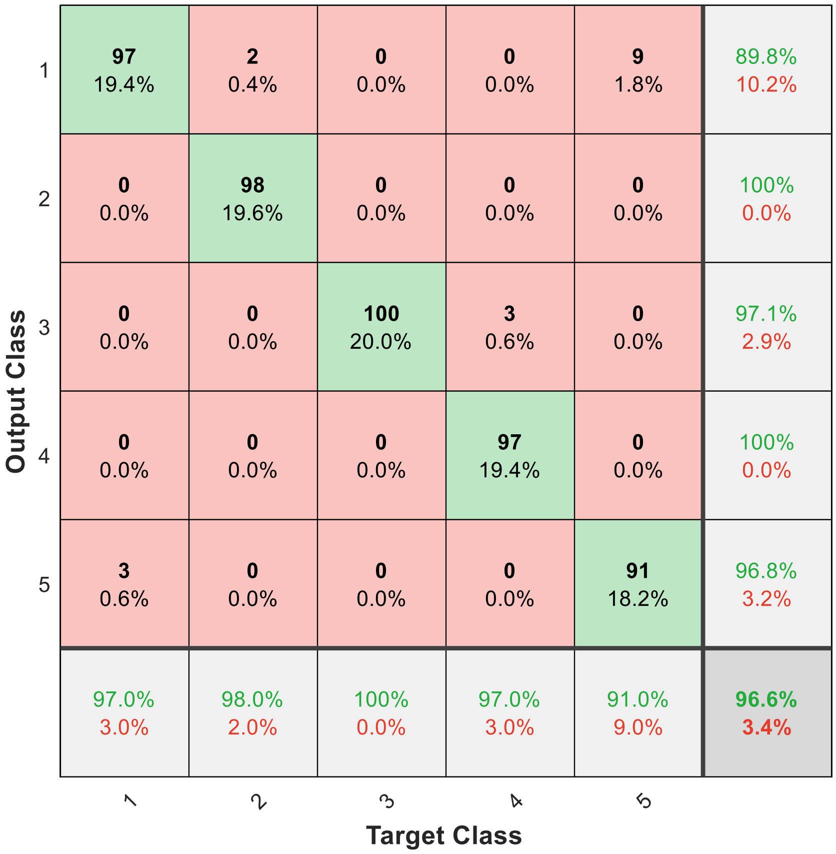
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{85+399}{85+399+15+1} = 96.8\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{85}{85+15} = 85\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{85}{85+1} = 98.83\%$$

$$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times85}{2\times85+15+1} = 91.4\%$$

Total accuracy of this approach will be the mean of individual accuracies of all the classes, i.e. 98.08%.



	1	2	3	4	5	
1	**97** 19.4%	**2** 0.4%	**0** 0.0%	**0** 0.0%	**9** 1.8%	89.8% 10.2%
2	**0** 0.0%	**98** 19.6%	**0** 0.0%	**0** 0.0%	**0** 0.0%	100% 0.0%
3	**0** 0.0%	**0** 0.0%	**100** 20.0%	**3** 0.6%	**0** 0.0%	97.1% 2.9%
4	**0** 0.0%	**0** 0.0%	**0** 0.0%	**97** 19.4%	**0** 0.0%	100% 0.0%
5	**3** 0.6%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**91** 18.2%	96.8% 3.2%
	97.0% 3.0%	98.0% 2.0%	100% 0.0%	97.0% 3.0%	91.0% 9.0%	**96.6%** **3.4%**

Fig. 10: Confusion matrix plot for Caltech dataset with SVM approach

Figure 10 shows a confusion matrix plot for Caltech dataset. As per the dataset, it exhibits five classes for the five categories of image. Calculation for each class is given as follows:

Table 3: Result for Caltech dataset using SVM classifier

Parameter	**1**	**2**	**3**	**4**	**5**
TP	97	98	100	97	91
FP	3	2	0	3	9
FN	11	0	3	0	3
TN	389	400	397	400	397

For '1':

$TP = 97, TN = 389, FP = 3, FN = 11$

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{97+389}{97+389+3+11} = 97.2\%$

$Precision = \frac{TP}{TP+FP} = \frac{97}{97+3} = 97\%$

$Recall = \frac{TP}{TP+FN} = \frac{97}{97+11} = 89.81\%$

$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 97}{2\times 97+3+11} = 93.26\%$

For '2':

TP=98, TN=400, FP=0, FN=0

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{98+400}{98+400+2+0} = 99.6\%$

$Precision = \frac{TP}{TP+FP} = \frac{98}{98+2} = 98\%$

$Recall = \frac{TP}{TP+FN} = \frac{98}{98+0} = 100\%$

$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 98}{2\times 98+2+0} = 98.98\%$

For '3':

TP=100, TN=397, FP=0, FN=3

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{100+397}{100+397+0+3} = 99.4\%$

$Precision = \frac{TP}{TP+FP} = \frac{100}{100+0} = 100\%$

$Recall = \frac{TP}{TP+FN} = \frac{100}{100+3} = 97.1\%$

$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 100}{2\times 100+0+3} = 98.52\%$

For '4':

TP=97, TN=400, FP=3, FN=0

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{97+400}{97+400+3+0} = 99.4\%$

$Precision = \frac{TP}{TP+FP} = \frac{97}{97+3} = 97\%$

$Recall = \frac{TP}{TP+FN} = \frac{97}{97+0} = 100\%$

$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 97}{2\times 97+3+0} = 98.47\%$

For '5':

TP=91, TN=397, FP=9, FN=3

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{91+397}{91+397+9+3} = 97.6\%$

$Precision = \frac{TP}{TP+FP} = \frac{91}{91+9} = 91\%$

$Recall = \frac{TP}{TP+FN} = \frac{91}{91+3} = 96.8\%$

$F - Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 91}{2\times 91+9+3} = 93.81\%$

Total accuracy of this approach will be the mean of individual accuracies of all the classes, i.e. 98.64%. Similarly, Table 4 calculates and displays the accuracy, recall, precision and F-score values.

Table 4: Comparative analysis

Dataset and Method	Precision	Accuracy	Recall	F-Score
Corel dataset with SVM	95.2%	98.08%	95.99%	95.29%
Caltech dataset with SVM	96.6%	98.64%	96.74%	96.6%

On observing the Table 4, it was found that both the datasets have nearly similar performance. It is clear that the results with Caltech outperforms the Corel database.

Table 5: Feature computation time for an image

Features	Computation time
Cartoon Features	34.308
Gabor wavelets	0.4900
Wavelet moments	0.345
Histogram	0.1121
Color autocorrelogram	0.0887
Color moments	0.0283
Hybrid	79.1245

Table 5 represents the computation time for different feature extraction methods. It can be seen that hybrid feature set are taking higher computational time as compared to other feature extraction. Color moments takes very less computation time. Even though hybrid feature set having higher computation time but best suitable for classification.

A general-purpose WANG dataset of 1000 Corel images in ten distinct topic categories is used to evaluate the proposed technique. It has a dimension of 384×256 pixels. JPEG images are 256×384 pixels. 100, as shown by table 6.

Photos from all 10 categories, including Africa, Beaches, Buses, Horses, Buildings, and more, are included in this collection. Mountain ranges, food, elephants, dinosaurs, and trees. Additionally, this information is gathered to evaluate other CBIR systems. Since the dataset class information is easily accessible, it is commonly used.

Table 6: Performance evaluation

Class	Category	“True Positive Rate”	“True Negative Rate”	“False Positive Rate”	“False Negative Rate”
1	“Africa”	0.9118	0.9824	0.0882	0.0176
2	“Beach”	0.9300	0.9825	0.0700	0.0175
3	“Transportation”	0.9043	0.9631	0.0957	0.0369
4	“Architecture”	0.9519	0.9975	0.0481	0.0025
5	“Dinosaur”	1.0000	1.0000	0	0

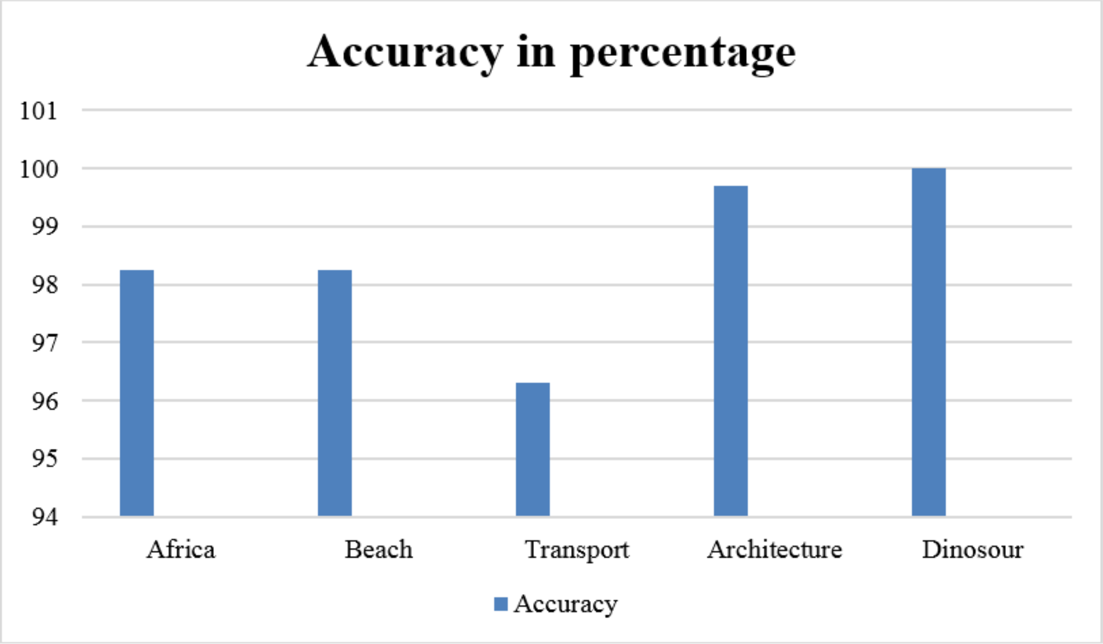


Fig. 11: Accuracy in percentage of different dataset

For a certain class of observations, recall is defined as the ratio of observations that were reasonably predicted to be positive to all observations. As shown in Fig. 11 and Table 7, NCA-Hybrid has a high recall ratio for all ten classes in the CORAL dataset, with recall values of 0.968, 0.98, 0.87, and 0.88 for the Create Beaches, Food, and Mountain groups respectively. Comparing hybrid feature set classifiers to NCA filtered hybrid feature SVM classifiers, the recall parameter value of the former is much lower.

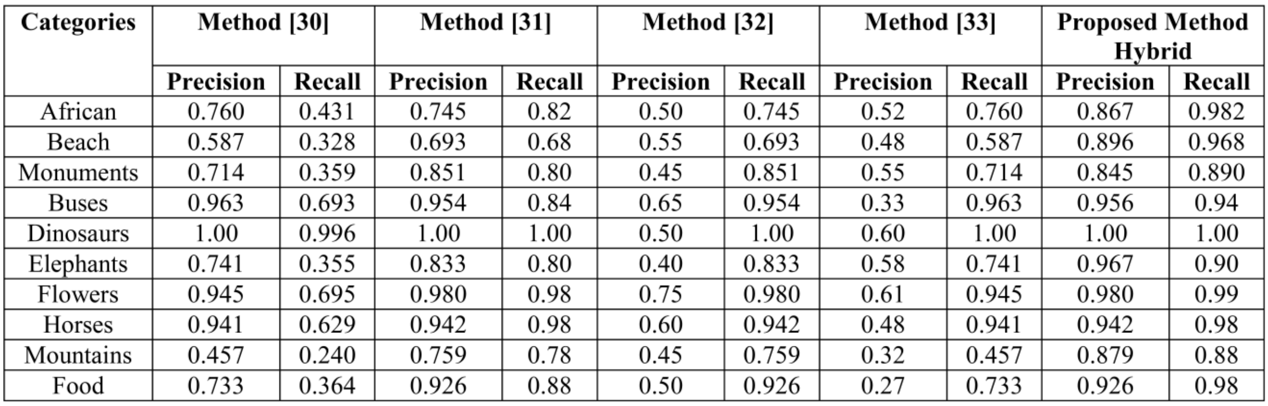
The estimated percentage of positive observations to all of the positive observations is how precision is determined. The SVM accuracy value was in the range of 0.86 and 1.

Table 7: Proposed Method Verses Individual Methods

Categories	Hybrid			NCA-Hybrid		
	“Precision”	**“Recall”**	**“Accuracy”**	**“Precision”**	**“Recall”**	**“Accuracy”**
“African”	0.680	0.76	0.822	0.867	0.982	0.968
“Beach”	0.638	0.60	0.836	0.896	0.968	0.962
“Monuments”	0.740	0.80	0.838	0.845	0.890	0.990
“Buses”	0.957	0.90	0.812	0.956	0.94	1.00
“Dinosaurs”	0.980	0.98	0.842	1.00	1.00	1.00
“Elephants”	0.700	0.70	0.846	0.967	0.90	0.987
“Flowers”	0.979	0.94	0.816	0.980	0.99	0.997
“Horses”	0.941	0.96	0.822	0.942	0.98	0.956
“Mountains’	0.733	0.66	0.816	0.879	0.88	0.986

“Food”	0.914	0.86	0.824	0.926	0.98	0.964
“Average”	0.826	0.816	0.827	0.998	0.958	0.976

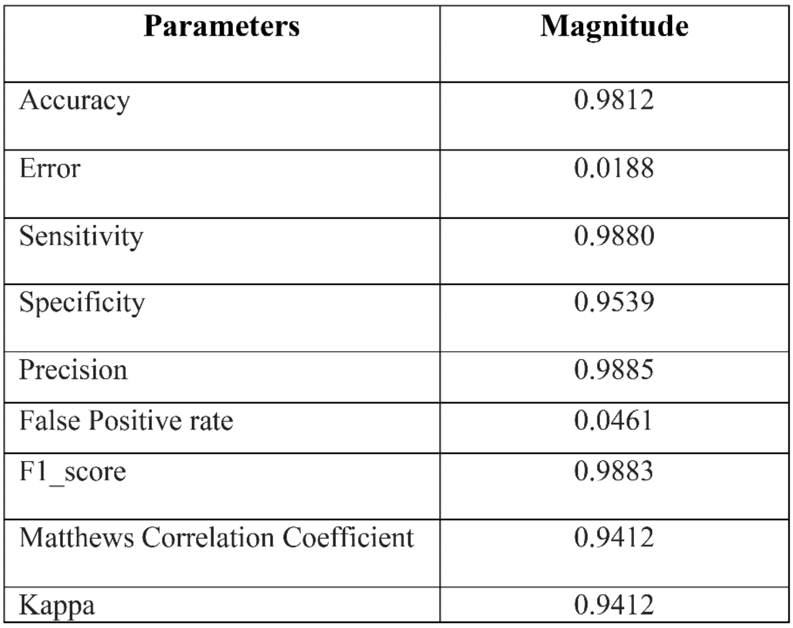
Table 8: Proposed Method vs. previously existing methods



Categories	Method [30]		Method [31]		Method [32]		Method [33]		Proposed Method Hybrid	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
African	0.760	0.431	0.745	0.82	0.50	0.745	0.52	0.760	0.867	0.982
Beach	0.587	0.328	0.693	0.68	0.55	0.693	0.48	0.587	0.896	0.968
Monuments	0.714	0.359	0.851	0.80	0.45	0.851	0.55	0.714	0.845	0.890
Buses	0.963	0.693	0.954	0.84	0.65	0.954	0.33	0.963	0.956	0.94
Dinosaurs	1.00	0.996	1.00	1.00	0.50	1.00	0.60	1.00	1.00	1.00
Elephants	0.741	0.355	0.833	0.80	0.40	0.833	0.58	0.741	0.967	0.90
Flowers	0.945	0.695	0.980	0.98	0.75	0.980	0.61	0.945	0.980	0.99
Horses	0.941	0.629	0.942	0.98	0.60	0.942	0.48	0.941	0.942	0.98
Mountains	0.457	0.240	0.759	0.78	0.45	0.759	0.32	0.457	0.879	0.88
Food	0.733	0.364	0.926	0.88	0.50	0.926	0.27	0.733	0.926	0.98

Recall for a particular class real is the ratio of observations that were appropriately predicted to be positive to all observations. Table 8 shows that NCA-Hybrid has a high recall ratio for all 10 classes in the CORAL dataset, with values ranging from 0.88 to 1., whereas [31] yields recall 0.68 to 1, [32] 0.69 to 1 and [33] gives recall values range from 0.58 to 1. The recall parameter value of the proposed hybrid feature set classifiers is significantly higher than that of the other methods.

Table 9: Overall Parameters of Proposed Method of NCA-Hybrid



Parameters	Magnitude
Accuracy	0.9812
Error	0.0188
Sensitivity	0.9880
Specificity	0.9539
Precision	0.9885
False Positive rate	0.0461
F1_score	0.9883
Matthews Correlation Coefficient	0.9412
Kappa	0.9412

5. Conclusion and Future Research Challenges

The proposed method integrates shape, texture, and colour data in order to construct an image retrieval framework. Additionally, this study focuses on a field where equivalent technologies are not yet accessible to professionals in the field, giving us insight on adaptability. The study effort in this article involved experiments on 10 different categories, and it is determined that the end result of this work includes the design of an effective intelligent computer vision system that can retrieve images with more accuracy. Finally, all approaches under comparison provided good results of 98.64 percent with the Caltech dataset and claimed in accordance with their performance assessment variables. It is also suggested that the fused SVM method for the betterment of the CBIR system.

References

[1] Wan, Ji, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. "Deep learning for content-based image retrieval: A comprehensive study." In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 157-166. 2014.

[2] Lande, Milind V., Praveen Bhanodiya, and Pritesh Jain. "An effective content-based image retrieval using color, texture and shape feature." In *Intelligent Computing, Networking, and Informatics*, pp. 1163-1170. Springer, New Delhi, 2014.

[3] Agarwal, Swati, Anil Kumar Verma, and Preetvanti Singh. "Content based image retrieval using discrete wavelet transform and edge histogram descriptor." In *2013 International Conference on Information Systems and Computer Networks*, pp. 19-23. IEEE, 2013.

[4] Tunga, Satish, D. Jayadevappa, and C. Gururaj. "A comparative study of content based image retrieval trends and approaches." *International Journal of Image Processing (IJIP)* 9, no. 3 (2015): 127-155.

[5] Nazir, A., Ashraf, R., Hamdani, T. and Ali, N., 2018, March. Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. In *2018 international conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-6). IEEE.

[6] Wang, X.Y., Liang, L.L., Li, Y.W. and Yang, H.Y., 2017. Image retrieval based on exponent moments descriptor and localized angular phase histogram. *Multimedia Tools and Applications*, *76*(6), pp.7633-7659.

[7] Liu, Guang-Hai, and Jing-Yu Yang. "Content-based image retrieval using color difference histogram." *Pattern recognition* 46, no. 1 (2017): 188-198.

[8] Lasmar, Nour-Eddine, and Yannick Berthoumieu. "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms." *IEEE Transactions on Image Processing* 23, no. 5 (2018): 2246-2261.

[9] Wang, Xinjian, Guangchun Luo, and Ke Qin. "A composite descriptor for shape image retrieval." In *2018 International Conference on Automation, Mechanical Control and Computational Engineering*. Atlantis Press, 2018.

[10] Wang, Xiang-Yang, Bei-Bei Zhang, and Hong-Ying Yang. "Content-based image retrieval by integrating color and texture features." *Multimedia tools and applications* 68, no. 3 (2019): 545-569.

[11] Basu, S., Karki, M., Ganguly, S., DiBiano, R., Mukhopadhyay, S., Gayaka, S., Kannan, R. and Nemani, R., 2017. Learning sparse feature representations using probabilistic quadtrees and deep belief nets. *Neural Processing Letters*, *45*(3), pp.855-867.

[12] Chandrasekhar, Vijay, Jie Lin, Qianli Liao, Olivier Morere, Antoine Veillard, Lingyu Duan, and Tomaso Poggio. "Compression of deep neural networks for image instance retrieval." In *2017 Data Compression Conference (DCC)*, pp. 300-309. IEEE, 2017.

[13] Yu, Wei, Kuiyuan Yang, Hongxun Yao, Xiaoshuai Sun, and Pengfei Xu. "Exploiting the complementary strengths of multi-layer CNN features for image retrieval." *Neurocomputing* 237 (2017): 235-241.

[14] Tzelepi, Maria, and Anastasios Tefas. "Deep convolutional learning for content based image retrieval." *Neurocomputing* 275 (2018): 2467-2478.

[15] Sugamya, Katta, Suresh Pabboju, and A. Vinaya Babu. "A CBIR classification using support vector machines." In *2016 International Conference on Advances in Human Machine Interaction (HMI)*, pp. 1-6. IEEE, 2016.

[16] Sarwar, Amna, Zahid Mehmood, Tanzila Saba, Khurram Ashfaq Qazi, Ahmed Adnan, and Habibullah Jamal. "A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine." *Journal of Information Science* 45, no. 1 (2019): 117-135.

[17] Wang, Xiang-Yang, Hong-Ying Yang, and Dong-Ming Li. "A new content-based image retrieval technique using color and texture information." *Computers & Electrical Engineering* 39, no. 3 (2013): 746-761.

[18] Manoharan, S., and S. Sathappan. "A novel approach for content based image retrieval using hybrid filter techniques." In *2013 8th International Conference on Computer Science & Education*, pp. 518-524. IEEE, 2013.

[19] Chen, Heng, Zhicheng Zhao, Anni Cai, and Xiaohui Xie. "An effective relevance feedback algorithm for image retrieval." In *2010 2nd IEEE International Conference on Network Infrastructure and Digital Content*, pp. 251-255. IEEE, 2010.

[20] Yalavarthi, A., Veeraswamy, K. and Sheela, K.A., 2017, July. Content based image retrieval using enhanced Gabor wavelet transform. In *2017 International Conference on Computer,*

Communications and Electronics (Comptelix) (pp. 339-343). IEEE.

[21] 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*.

[22] Mahajan, R. ., Patil, P. R. ., Potgantwar, A. ., & Bhaladhare, P. R. . (2023). Novel Load Balancing Optimization Algorithm to Improve Quality-of-Service in Cloud Environment. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 57–64. https://doi.org/10.17762/ijritcc.v11i2.6110

[23] 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), pp.1-12.

[24] Bhunia, A.K., Bhattacharyya, A., Banerjee, P., Roy, P.P. and Murala, S., 2020. A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern. *Pattern Analysis and Applications*, *23*(2), pp.703-723.

[25] Mistry, Y., Ingole, D.T. and Ingole, M.D., 2018. Content based image retrieval using hybrid features and various distance metric. *Journal of Electrical Systems and Information Technology*, *5*(3), pp.874-888.

[26] Ahmed, K.T., Ummesafi, S. and Iqbal, A., 2019. Content based image retrieval using image features information fusion. *Information Fusion*, *51*, pp.76-99.

[27] Cai, J., Luo, J., Wang, S. and Yang, S., 2018. Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, pp.70-79.

[28] Yang, W., Wang, K. and Zuo, W., 2012. Neighborhood component feature selection for high-dimensional data. *J. Comput.*, *7*(1), pp.161-168.

[29] Griffin, Gregory, Alex Holub, and Pietro Perona. "Caltech-256 object category dataset." (2007).

[30] Wang, J.Z., Li, J. and Wiederhold, G., 2001. SIMPLIcity: Semantics-sensitive integrated matching for image libraries. *IEEE Transactions on pattern analysis and machine intelligence*, *23*(9), pp.947-963.

[31] Pise, D. P. . (2021). Bot Net Detection for Social Media Using Segmentation with Classification Using Deep Learning Architecture. Research Journal of Computer Systems and Engineering, 2(1), 11:15. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/13

[32] Patil, Priyadarshini, and Bhagya Sunag. "Analysis of image retrieval techniques based on content." In *2015 IEEE International Advance Computing Conference (IACC)*, pp. 958-962. IEEE, 2015.

[33] Gali, Raghupathi, M. L. Dewal, and R. S. Anand. "Genetic algorithm for content based image retrieval." In *2012 fourth international conference on computational intelligence, Communication Systems and Networks*, pp. 243-247. IEEE, 2012.

[34] Vikhar, Pradnya, and Pravin Karde. "Improved CBIR system using edge histogram descriptor (EHD) and support vector machine (SVM)." In *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, pp. 1-5. IEEE, 2016.

[35] Ansari, Mohd Aquib, Manish Dixit, Diksha Kurchaniya, and Punit Kumar Johari. "An Effective Approach to an Image Retrieval using SVM Classifier." *database* 1 (2017): 2.