Improved Feature Selection Techniques for Image Retrieval based on Metaheuristic Optimization

Punit Kumar Johari[1] and Rajendra Kumar Gupta[2],
[1,2]Madhav Institute of Technology and Science, Gwalior, MP, INDIA

Summary

Content-Based Image Retrieval (CBIR) system plays a vital role to retrieve the relevant images as per the user perception from the huge database is a challenging task. Images are represented is to employ a combination of low-level features as per their visual content to form a feature vector. To reduce the search time of a large database while retrieving images, a novel image retrieval technique based on feature dimensionality reduction is being proposed with the exploit of metaheuristic optimization techniques based on Genetic Algorithm (GA), Extended Binary Cuckoo Search (EBCS) and Whale Optimization Algorithm (WOA). Each image in the database is indexed using a feature vector comprising of fuzzified based color histogram descriptor for color and Median binary pattern were derived in the color space from HSI for texture feature variants respectively. Finally, results are being compared in terms of Precision, Recall, F-measure, Accuracy, and error rate with benchmark classification algorithms (Linear discriminant analysis, CatBoost, Extra Trees, Random Forest, Naïve Bayes, light gradient boosting, Extreme gradient boosting, k-NN, and Ridge) to validate the efficiency of the proposed approach. Finally, a ranking of the techniques using TOPSIS has been considered choosing the best feature selection technique based on different model parameters.

Keywords:

CBIR, GA, EBCS, WOA, Feature Selection, Machine learning.

1. Introduction

In diverse image retrieval systems, machine learning and data mining techniques have been applied successfully [1][2]. In the selection of features for high dimensional databases, the significance of data mining is especially important[3]. The accuracy of classification may be decreasing due to the high number of feature dimensions. It may decrease the performance, even if it does not affect the accuracy. Therefore, the combination of the most important feature in the presence of labels has always been a promising study and insignificant have been discarded. The feature selection methods are classified into three major approaches Filter, Wrapper and Embedded[4][5]. Individual feature evaluation and subset feature evaluation can split into two major groups in the filter method. The heuristic and metaheuristic techniques or even the combination of heuristic and metaheuristic approaches, used in individual feature evaluation filters to rank the features. Next, based on a predefined threshold, for selection, the number of the top features are selected for classification. However, based on a certain approach using a certain measure, subset candidates try to find by subset feature evaluation. In order to find the best subset candidate, each time the previous best subset should be compared to the current subset. Similar ranking features are removed by the subset feature evaluation, however, in accordance with their relevance individual feature evaluation keeps redundant features in their final subset. The methods of filtering are low in complexity and consistent with the various data sets. In comparison, wrapper methods will determine the best subset of features by measuring the accuracy of classifiers for any sub-set of features from the widespread feature set. In general, in contrast to filter techniques, wrapper methods are very time-consuming. Wrapper methods depend on the classifier. Sequential selection and recursive removal of features involve in some common wrapper methods. Embedded approaches are the hybridization of filter and wrapper approaches. This paper concentrates on the filter method based on its negligible time and quantitative scalability, with individual feature evaluation.

2. Related Work

In the area of feature selection, there are several researchers, a significant gap is being attempted to cover by them. The study focused on the related analysis that makes use of meta-heuristic algorithms for the selection process. This section reviews few methods of feature selection and related work on the filter, wrapper and embedded methods.

Bolon-Canedo et al. suggested micro-array data in a distributed fashion classification using a filter approach. The framework allocates the data by features i.e. by the vertical distribution and then updates the features subset to boost the classification accuracy by performing a merging algorithm. To test the proposed method Naïve Bayes, SVM, C4.5 and K-NN classifier were used. The findings showed that the solution proposed could minimize the number of chosen features in comparison with traditional approaches such as ReliefF10%, ReliefF25%, IG10% and IG25% and retain accuracy [6].

Manuscript received January 5, 2021
Manuscript revised January 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.1.7

Moayedikia et al. suggested a wrapper approach based on two-stage called SYMON. In the first step, based on an emphasis on class labels, the degree of significance for each feature was measured. It is important to note that SYMON may differentiate it from other similar works in this area with its simplicity in dealing with features of equal significance. Next, based on their significance, the top k features have been chosen. The explanation behind SYMON for similar outcomes with other approaches could be that the significance of features may be calculated by symmetrical variance. SYMON was constrained by two. First, it really took time because this is the essence of the wrapper approach. Secondly, the feature selection is limited to the appropriate subset size(d). The constraint will harm precision [7].

3. Proposed method

In this section, the proposed work has been suggested based on the fusion of color and texture features. The author has already proposed the feature extraction method in the earlier literature [8][9]. The brief details are given in section 4.1. It results in that motivation to enhance the performance of the image retrieval. The proposed work is mainly divided into two-stage: in the first stage, the detailed discussion based on the three existing algorithms that have been used for the feature selection in section 3.1. The next stage given a brief overview of the machine-learning algorithms (for classification accuracy) is being used for evaluating the performance of the feature selection methods in section 3.2.

3.1 Algorithms used for feature selection

The section describes the three metaheuristic-based feature selection algorithms (Whale, EBCS and GA) are briefly discuss. The core idea of these feature selection is identical. Each algorithm calculated and selected the important features from the feature datasets. In our approach, we use two feature datasets. The following is the overview of each algorithm for the feature selection.

3.1.1 Whale Algorithm:

Meta-heuristic algorithms today play an important part in the development of the best option available for complex issues due to their performance and their effectiveness in achieving the best possible solutions. An analysis of the evolutionary behaviors of animals (e.g. birds, insects, humans) was carried out and simulates metaheuristic optimization algorithms to a computer science algorithm. Whale algorithm [10], is a modern meta-heuristic algorithm that simulates hunting strategy for humpback whales with three lead operators, and a few random parameters. The directed section of the whale algorithm imitates how humpback whales, in addition to bubble nets, hunt and encircle the prey. The first step of the whale algorithm is to assess the issue with the number of whales. The next step is to establish the initial population and to define the fitness function. The only approach is to use the fitness function to find the leading whale. In addition, other solutions should update their location to the leader to look for prey, surround the beast and attack the bubble net. The algorithm functions so that the prey is randomly chosen in the first section of the algorithm with Equations (1)-(2) or in the second section with Equations (5)-(6) for a bubbles-net attack.

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{position}}(t) - \vec{X}(t)\right| \quad (1)$$

$$\vec{X}(t+1) = \overrightarrow{X_{position}}(t) - \vec{A}.\vec{D} \quad (2)$$

Where t shows current iteration, coefficients vectors are A and C, $\overrightarrow{X_{position}}$ is the best solution obtained from position vector, X is the position vector, | | is the absolute value and indicates multiplication of element-by-element. Here it is important to note that in any iteration $\overrightarrow{X_{position}}$ should be modified is a better solution exists.
The following vectors $\vec{A}$ and $\vec{C}$ are calculated:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (3)$$
$$\vec{c} = 2.\vec{r} \quad (4)$$

P is the random number in [0,1] and the coefficient of vector A in Eq. (3) to control the option of the search or encircling the prey in the first section of the whaling algorithm. Thus the prey seeking option if P < 0.5 and |a|>=1, while P < 0.5 and |A|<1 encircles the prey. The coefficient vectors are $\vec{A}$ and $\vec{c}$, and the positions of prey and the whale are $X_{position}$ and X respectively. In Eqs. (3)-(4) $\vec{A}$ and $\vec{c}$ are initialized so that through iterations a(vector) should be decreased from 2 to 0 constantly and r1 and r2 are random vectors in range [0,1]. However if a P > 0.5 Whale Algorithm selects the second section of the algorithm for a bubble-net attack.

D shows the distance between the optimal shape of the prey as well as Eq (6).

$$\vec{X}(t+1) = \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X_{position}}(t) \quad (5)$$
$$\overrightarrow{D'} = \left|\overrightarrow{X^{position}}(t) - \vec{X}(t)\right| \quad (6)$$

Where $\overrightarrow{D'}$ and shows the distance of the ith whale to the prey (obtained best solution), constant value of b for defining the shape of the logarithmic spiral, l is a random number in [-1, 1].
This paper proposed a new technique using whale algorithms to remove one-half of all irrelevant features as part of the hybrid filter selection process. There is also a satisfactory performance of the proposed system. The use

of whale algorithms has been shown to improve the accuracy of the proposed high-dimensional benchmarking approach relative to the state of the art.

3.1.2 Extended Binary Cuckoo Search Algorithm

The Cuckoo Search (CS) is a stochastic swarm-intelligence population-based algorithm, which was developed by Yang and Deb [11] in 2009. The algorithms were influenced by the parasitic behavior of cuckoo birds, where female cuckoos placed their eggs inside the nest of other birds. The two potential circumstances on eggs may occur in this case: The host bird cannot differentiate a cuckoo egg from its own egg and hence the cuckoo egg will hatch into a new cuckoo generation, or the host bird will recognize a cuckoo egg and instinctively attempt to throw the egg away or abandon the nest to create a new nest. On the other hand, the search approach for Levy flights by CS investigates the search field with a flat flight route with unexpected 90-degree turns[12]. The CS also relies heavily on a random walk search mechanism and can rapidly skip from region to region without thoroughly investigating each cuckoo nest. These two behaviors have motivated CS method of investigation and created new solutions for the future generation using Levy flights and substitution of eggs.

CS use three idealized rules:
1) Every cuckoo kays one egg at a time and dump its egg in a randomly selected nests.
2) High quality of eggs of best nests are transported to feature generation.
3) Fixed number of available host nests and the egg that a cuckoo is laid is found by the host bird with a probability $p_a \in [0, 1]$.

For problems of optimization, each host nest is a solution and depending on the problem size, may contain one or more eggs. The nests are initialized by random means in the first step of the algorithm and modified using random walk Levy flights in each iteration of the algorithm as represented in the following equations (7) and (8):

$$s_i^{(t+1)}(t) = s_i^{(t)} + \alpha \oplus Levy(\lambda) \qquad (7)$$

$$Léay \sim u = t^{-\lambda}, (1 < \lambda \leq 3), \qquad (8)$$

3.1.3 Genetic Algorithm (GA)

A genetic algorithm [20] is a randomized search algorithm based on principles of the natural competition between natural selection and natural genetics. Biologically inspired operators such as selection, mutation and crossover use it to produce high-quality solutions to search problems and optimization.

3.2 A brief introduction of the machine learning algorithms

In this section, a brief overview of each machine learning algorithms [21] is given as follows:

Naïve Bayes: Naïve Bayes classifiers are in statistics a family of basic "probabilistic classifiers" that apply the Bayes theorem, with solid assumptions of independence between features.

LDA: the generalization of Fisher's linear discriminant approach, used in statistics and other fields to find a linear combination of features that characterizes or distinguishes two or more groups of objects or events are known as Linear Discriminant Analysis (LDA) [22]. As a linear classifier, the resultant combination may be used, more often, before the corresponding classification for dimensional reduction. When more than two classes exist, the analysis used in the derivation of the Fisher discriminant can be applied to identifying a subspace that seems to include all class variability [23].

CatBoost: It is based on gradient boosted decision trees. A set of decision-making trees are constructed consecutively during training. Compared to the previous trees, each successively constructed tree has lower losses [13].

LightGBM: It is open-source software distributed gradients booster platform originally created by Microsoft[14]. It is designed based on a decision tree and is used to classify, rank and other machine learning activities. Performance and scalability are the focus of development.

Random forest: Random forest is a powerful machine-learning algorithm, which can be, used easily that produces most of the time batter result, even without hyper-parameter tuning. Due to its simplicity and diversity (it can be used for both classification and regression tasks), it is one of the machine-learning algorithm used in many applications [15].

ExtraTree: Adding another randomization stage generates an extremely randomize tree or ExtraTree. Although they are an aggregate of individual trees, there are two key distinctions, comparable to traditional random forests. In particular, the first one includes the completely learning sample to train each tree (as compared to bootstrap) and second the top-down splitting in the tree learner is randomized. Rather than determine the locally optimal cut-point for each considered feature, a random cut-point is chosen (based on information gain or Gini impurity) [16].

4 Experimental design

4.1 Dataset used and Feature extraction:

In the proposed study, the experiments based on two datasets and the features dataset are being taken from the previous work are available in the literature [17] and motivation behind the work is to reduce the feature space relevant for the classify the images accurately. In the previous, study two variants of the feature vector were proposed by the author namely MBP+MBPH+CH and MBP+MBPH+FCH extracted from using well-known image datasets COREL 5K [18] and COREL 10K [19]. In this paper, the study proposed only one feature vector i.e. MBP+MBPH+FCH. Following is the dataset information:
Feature dataset1 = MBP+MBPH+FCH extracted from COREL 10K image dataset.
Feature dataset2 = MBP+MBPH+FCH extracted from COREL 5K image dataset.

4.2. Measurement criteria

The criteria used to measure the efficiency of the system proposed in Equation (9)-(12) are precision, Recall, Accuracy, and F-Measure. Precision indicates the ratio of negative and positive samples accurately identified. The proportion of negative and positive samples estimated also indicates the specificity and sensitivity. Positive and negative samples are properly classified in the following True Positive (TP) and True Negative (TN). False positives (FP) and false negatives (FN) are both positive and misleading samples, which have been incorrectly categorized.

- Precision

Reflect classifier discrimination ability among the different class images.

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

- Recall:

Shows how well system recognized the different images.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

- Accuracy:

It is defined as the proportion of instances that are correctly classified. It is calculated as:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (11)$$

- F-measure: is combined measure for precision and recall

$$F-measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (12)$$

- AUC: Area under ROC curve of a classifier is the probability that the classifier ranks a randomly chosen negative image lesser than a randomly chosen positive.

5 Result and Discussion.

Table 1 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms. The analysis based on the proposed Whale feature selection algorithm.

Table1: The measurement criteria on feature dataset1 using the Whale feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
Linear Discriminant Analysis	0.8846	0.9924	0.8800	0.8860	0.8701
CatBoost Classifier	0.8275	0.9788	0.8350	0.8280	0.8070
Extra Trees Classifier	0.8198	0.9706	0.8300	0.8123	0.7938

Where the classification performance of linear discriminant analysis (LDA) is outperforming as compared with the other machine-learning model with accuracy (0.8846), area under ROC curve (0.9924), recall (0.8800), precision (0.8860) and F1 measure (0.8701) in the CBIR scheme. Fig. 1(b) shows the confusion matrix that shows class 3 is less classify as compared to other classes. Fig. 1(c) shoes the class report based on confusion matrix. The model incorrectly classified 2 cases as class 0, 1 case as class 2, 1 case as class 3, 1 case as class 5, 2 cases as class 6 and one case as class 7. The model incorrectly classified 3, 1, 2, 1, 1 cases as not belonging to class 1, 3, 4, 6 and 8 respectively. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 1(d) the learning curve for Linear Discriminant Analysis the training score is still around the maximum and the validation score could be increased with more training samples. Fig. 1(a) shows the area under the ROC curve (AUC) is a useful tool for evaluating the quality of class separation for the soft classifier. The mean AUC of .97 indicates that the model separates all classes very well.

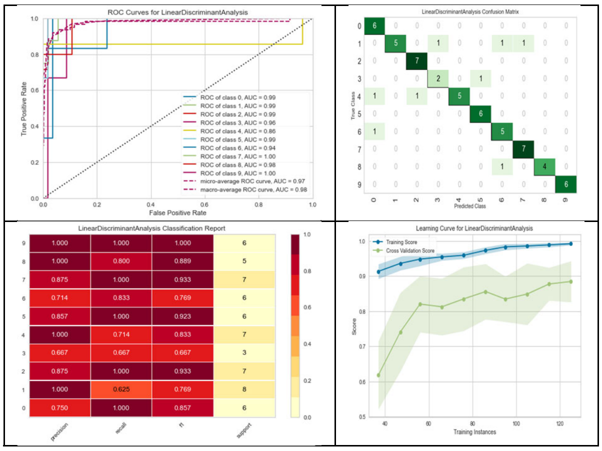

Fig. 1 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 2)

Table 2 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms. The analysis based on the proposed EBCS feature selection algorithm.

Table 2 The measurement criteria on feature dataset1 using the EBCS feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
CatBoost Classifier	0.8396	0.9685	0.8550	0.8496	0.8227
Extra Trees Classifier	0.7758	0.9587	0.7800	0.7848	0.7537
Random Forest Classifier	0.7687	0.9614	0.7800	0.7687	0.7430

Where the classification performance of the CatBoost model is also performing well as compared with the other machine-learning model with accuracy (0.8396), the area under ROC curve (0.9685), recall (0.8550), precision (0.8496) and F1 measure (0.8227) in the CBIR scheme. Fig. 2(a) shows the area under the ROC curve (AUC), the plot indicates that most of the classes can be predicted very well, while class 3 are difficult to predict. The mean AUC of .98 indicates that the model separates all classes very well. Fig. 2(b) shows the confusion matrix that shows class 1, 3, 4 and 8 is less classify as compared to other classes. The model incorrectly classified 1 case as class 2, 4 cases as class 3, 1 case as class 4, 3 cases as class 5, 2 cases as class 7 and one case as class 8. The model incorrectly classified 4, 2, 1, 2, 1, 2 cases as not belonging to class 1, 3, 4, 6, 7 and 8 respectively. Fig. 2(c) shows the class report based on the confusion matrix. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 2(d) the learning curve for CatBoost the training score is still around the maximum and the validation score could be increased with more training samples.

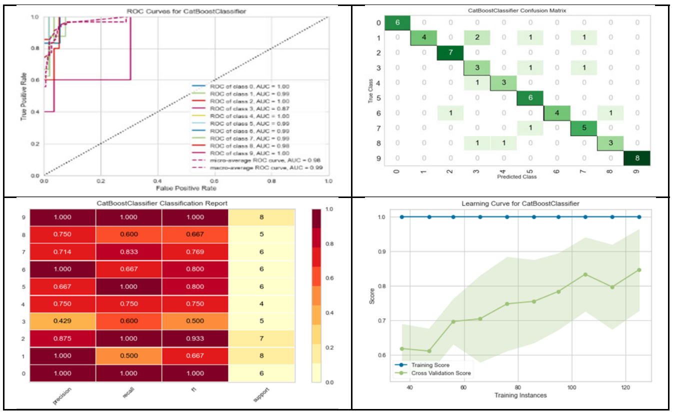

Fig. 2 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 2)

Table 3 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms. The analysis based on the proposed EBCS feature selection algorithm.

Table 3: Table1: The measurement criteria on feature dataset1 using the GA feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
CatBoost Classifier	0.8137	0.9775	0.8300	0.8045	0.7897
Random Forest Classifier	0.7995	0.9533	0.8100	0.7920	0.7766
Linear Discriminant Analysis	0.7995	0.9740	0.8100	0.7735	0.7714

Where the classification performance of the CatBoost model is also performing well as compared with the other machine-learning model with accuracy (0.8137), area under ROC curve (0.9775), recall (0.8300), precision (0.8045), F1 measure (0.7897), kappa (0.7913) and MCC (0.8028) in the CBIR scheme. Fig. 3(a) shows the area under the ROC curve (AUC), the plot indicates that most of the classes can be predicted very well, while classes 1 and 4 are difficult to predict. The mean AUC of .98 indicates that the model separates all classes very well. Fig. 3(b) shows the confusion matrix that shows class 1, 3 and 8 is less classify as compared to other classes. The model incorrectly classified 2 cases as class 0, 3 cases as class 1, 1 case as class 2, 2 cases as class 6, 1 case as class 7 and 2 cases as class 8. The model incorrectly classified 1, 1, 3, 2, 3, 1 cases as not belonging to class 1, 2, 3, 4, 5 and 6 respectively. Fig.

3(c) shows the class report based on the confusion matrix. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 3(d) the learning curve for CatBoost the training score is still around the maximum and the validation score could be increased with more training samples. The mean AUC of .98 indicates that the model separates all classes very well.

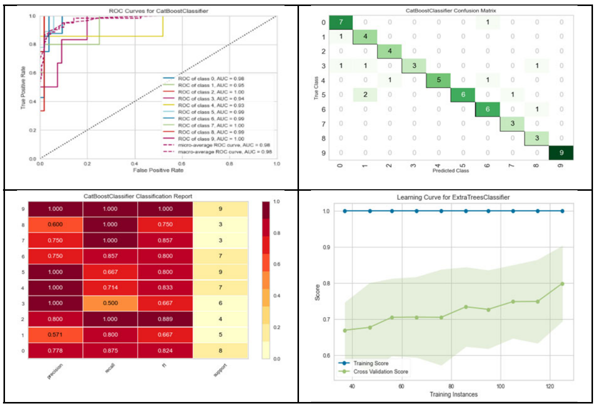

Fig.3 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 3)

Table 4 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms. The analysis based on the proposed Whale feature selection algorithm. Where the classification performance of the Linear Discriminant Analysis (LDA) model is also performing well as compared with the other machine-learning model with accuracy (0.7412), area under ROC curve (0.9505), recall (0.7400), precision (0.7340), F1 measure (0.7074), kappa (0.7090) and MCC (0.7219) in the CBIR scheme.

Table 4: The measurement criteria on feature dataset2 using the Whale feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
CatBoost Classifier	0.7412	0.9505	0.7400	0.7340	0.7074
Linear Discriminant Analysis	0.7330	0.9548	0.7250	0.6923	0.6903
Random Forest Classifier	0.6912	0.9316	0.6750	0.6597	0.6459

Fig. 4(a) shows the area under the ROC curve (AUC), the plot indicates that most of the classes can be predicted very well, while class 2, 5, 6, 8 are difficult to predict. The mean AUC of .98 indicates that the model separates all classes very well. Fig. 4(b) shows the confusion matrix that shows class 1, 3, 4 and 8 is less classify as compared to other classes. The model incorrectly classified one case as class 2, 3 cases as class 3, 3 cases as class 5, 2 cases as class 6, 1 case as class 7 and four cases as class 8. The model incorrectly classified 1, 2, 2, 1, 1, 3, 3, 1 cases as not belonging to class 0, 1, 2, 4, 5, 6, 7 and 8 respectively. Fig. 4(c) show the class report based on confusion matrix. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 2.2 the learning curve for LDA the training score is still around the maximum and the validation score could be increased with more training samples. The mean AUC of .98 indicates that the model separates all classes very well.

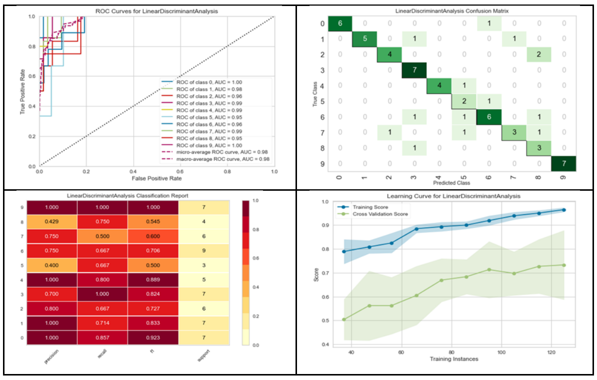

Fig.4 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 4)

Table 5 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms. The analysis based on the proposed EBCS feature selection algorithm. Where the classification performance of CatBoost model is also performing well as compared with the other machine-learning model with accuracy (0.6764), area under ROC curve (0.9574), recall (0.7000), precision (0.6231), F1 measure (0.6259), kappa (0.6396) and MCC (0.6539) in the CBIR scheme.

Table 5: The measurement criteria on feature dataset2 using the EBCS feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
CatBoost Classifier	0.6764	0.9574	0.7000	0.6231	0.6259
Random Forest Classifier	0.6264	0.9149	0.6500	0.5894	0.5776
Linear Discriminant Analysis	0.6253	0.9204	0.6300	0.5921	0.5805

Fig. 5(a) shows the area under the ROC curve (AUC), the plot indicates that most of the classes can be predicted very well, while class 2, 5, 6, 8 are difficult to predict. The mean AUC of .98 indicates that the model separates all classes very well. Fig. 5(b) shows the confusion matrix that shows class 1, 3, 4 and 8 is less classify as compared to other classes. The model incorrectly classified one case as class 2, 3 cases as class 3, 3 cases as class 5, 2 cases as class 6, 1 case as class 7 and four cases as class 8. The model incorrectly classified 1, 2, 2, 1, 1, 3, 3, 1 cases as not belonging to class 0, 1, 2, 4, 5, 6, 7 and 8 respectively. Fig. 5(c) shows the class report based on confusion matrix. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 5(d) the learning curve for CatBoost the training score is still around the maximum and the validation score could be increased with more training samples. The mean AUC of .98 indicates that the model separates all classes very well.

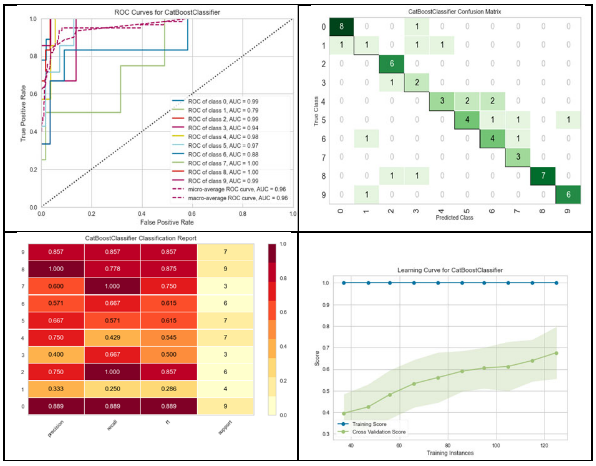

Fig.5 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 5)

Table 6 shows the performance analysis on the CBIR scheme that achieved the best classification accuracies based on machine learning algorithms.

Table 6: The measurement criteria on feature dataset2 using the GA feature selection algorithm and different machine learning algorithms.

Model	Accuracy	AUC	Recall	Prec.	F1
CatBoost Classifier	0.6984	0.9487	0.6750	0.6707	0.6578
Extreme Gradient Boosting	0.6923	0.9261	0.6850	0.6873	0.6650
Linear Discriminant Analysis	0.6918	0.9526	0.6700	0.6769	0.6573

The analysis based on the proposed GA feature selection algorithm. Where the classification performance of the CatBoost model is also performing well as compared with the other machine-learning model with accuracy (0.6984), area under ROC curve (0.8549), recall (0.6861), precision (0.6618), F1 measure (0.6545), kappa (0.6529) and MCC (0.6659) in the CBIR scheme. Fig. 6(a) shows the area under the ROC curve (AUC), the plot indicates that most of the classes can be predicted very well, while class 2, 5, 9 are difficult to predict. The mean AUC of .97 indicates that the model separates all classes well. Fig. 6(b) shows the confusion matrix that shows class 0, 2, 4, 5 and 9 is less classify as compared to other classes. The model incorrectly classified 2 cases as class 0, 1 case as class 1, 2 cases as class 2, 2 cases as class 3, 1 case as class 4, 6 cases as class 5, 2 cases as class 7, 1 case as class 8 and one case as class 9. The model incorrectly classified 2, 2, 1, 5, 1, 3, 4 cases as not belonging to class 1, 2, 3, 4, 5, 6 and 8 respectively. Fig. 6(c) shows the class report based on confusion matrix. The analysis shows that most of the classes predicted well with less error rate. We can see clearly in Fig. 6(d) the learning curve for CatBoost the training score is still around the maximum and the validation score could be increased with more training samples. The mean AUC of .97 indicates that the model separates all classes very well.

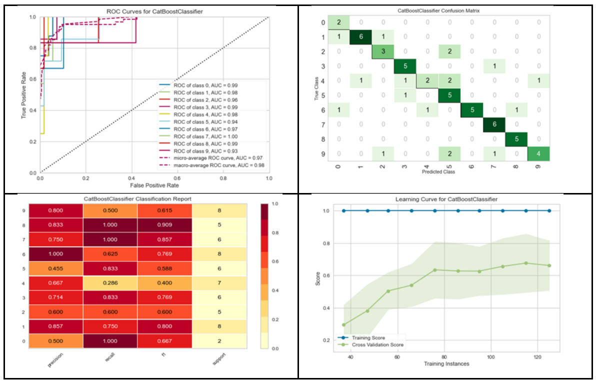

Fig.6 (a) ROC-AUC curve (b) Confusion-Metrix (c) Class report (d) Learning curve (from Table 6)

Although the AUC-ROC curve is mostly used in binary classification problems, it is extended to multiclass classification problems by using the One vs All technique. Therefore, in the case of multiple classes 0, 1, …, n, the ROC for class 0 will be generated as classifying 0 against not 0, i.e. 1 to n and so on.

5.1 Evaluation criteria using TOPSIS:

Hwang & Yoon [20] develop the “Technique for Order Preference by Similarity to Ideal Solution” (TOPSIS) is a technique to evaluate the performance of the alternatives through the similarity with the ideal solution. This technique will be a batter choice that is similar to the positive-ideal solution [21] and most far from the negative ideal solution. In this study, the TOPSIS compares performance among algorithms in terms of accuracy, AUC, Precision, Recall, F1 Measure and selected features. This approach helps to find the best, the second better and the worst algorithms to be found. It consists of a very basic computing method to enable scientists/ professionals to use it in various fields of knowledge. Tables 7 and 8 show the analysis based on feature dataset1 and feature dataset2 respectively. The analysis shows that the rank of the whale algorithm is higher among others.

Table 7: TOPSIS analysis for Dataset 1

	Whale	EBCS	GA
Si*	0	9.000373	23.00002417
Si'*	23.00015	14	0.052144366
Si* + Si'*	23.00015	23.00037	23.05216853
Si'*/(Si* + Si'*)	1	0.608686	0.002262016
	⇧ Best solution		

Table 8: TOPSIS analysis for Dataset 1

	Whale	EBCS	GA
Si*	3.000008	0.182005	25.00059
Si'*	22.00099	0.103438	0.052698
Si* + Si'*	25.001	0.285443	25.05329
Si'*/(Si* + Si'*)	0.880004	0.362378	0.002103
	⇧ Best solution		

6. Conclusion

This article is an experimental analysis for the evaluation of the results of three feature selection algorithms and five machine-learning algorithms in multiclass image classification for image retrieval. In terms of classification accuracy, the ranking of the three feature selection algorithms for conducting multiclass image classification is results that the whale algorithm has more discriminative power for classification as compared to the algorithms. In terms of classification accuracy of the multiclass image classifier CatBoost and Linear discriminant analysis perform the best on the dataset. Based on experiment results ideal solution has found using TOPSIS based method.

References

[1] F. Tang, L. Adam, and B. Si, "Group feature selection with multiclass support vector machine," *Neurocomputing*, vol. 317, pp. 42–49, Nov. 2018.

[2] S. Nagpal, S. Arora, S. Dey, and Shreya, "Feature Selection using Gravitational Search Algorithm for Biomedical Data," *Procedia Comput. Sci.*, vol. 115, no. October, pp. 258–265, 2017.

[3] T. R. Shultz and S. E. Fahlman, *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US, 2017.

[4] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[5] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *J. Biomed. Inform.*, vol. 43, no. 1, pp. 15–23, 2010.

[6] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Distributed feature selection: An application to microarray data classification," *Appl. Soft Comput.*, vol. 30, pp. 136–150, May 2015.

[7] A. Moayedikia, K.-L. Ong, Y. L. Boo, W. G. Yeoh, and R. Jensen, "Feature selection for high dimensional imbalanced class data using harmony search," *Eng. Appl. Artif. Intell.*, vol. 57, no. October 2016, pp. 38–49, Jan. 2017.

[8] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Syst. Appl.*, vol. 80, pp. 323–339, Sep. 2017.

[9] S. K.V and R. K.K, "Classification of Abnormalities in Medical Images Based on Feature Transformation- A Review," *Int. J. Sci. Eng. Res.*, vol. 10, no. 8, pp. 1304–1308, 2019.

[10] M. A. El Aziz, A. A. Ewees, and A. E. Hassanien, "Multi-objective whale optimization algorithm for content-based image retrieval," *Multimed. Tools Appl.*, vol. 77, no. 19, pp. 26135–26172, Oct. 2018.

[11] X. Yang and Suash Deb, "Cuckoo Search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214.

[12] A. M. Reynolds and M. A. Frye, "Free-Flight Odor Tracking in Drosophila Is Consistent with an Optimal Intermittent Scale-Free Search," *PLoS One*, vol. 2, no. 4, p. e354, Apr. 2007.

[13] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. Section 4, pp. 6638–6648, Jun. 2017.

[14] D. Wang, Y. Zhang, and Y. Zhao, "LightGBM," in *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics - ICCBB 2017*, 2017, pp. 7–11.

[15] V. P. Singh and R. Srivastava, "Improved content-based image classification using a random forest classifier," *Adv. Intell. Syst. Comput.*, vol. 554, pp. 365–376, 2018.

[16] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[17] P. K. Johari and R. Kumar, "An Improved Image Retrieval by Using Texture Color Descriptor with Novel Local Textural Patterns," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, 2020.

[18] J. Z. Wang, J. Li, and G. Wiederholdy, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1929, no. 9, pp. 360–371, 2000.

[19] W. Bian and D. Tao, "Biased Discriminant Euclidean Embedding for Content-Based Image Retrieval," *IEEE*

Trans. Image Process., vol. 19, no. 2, pp. 545–554, Feb. 2010.
[20] K. P. Yoon and C. L. Hwang, "Multiple attribute decision making: an introduction," vol. 1, 1995.
[21] R. A. Krohling and A. G. C. Pacheco, "A-TOPSIS – An Approach Based on TOPSIS for Ranking Evolutionary Algorithms," *Procedia Comput. Sci.*, vol. 55, no. Itqm, pp. 308–317, 2015.
[20] Mitchell, Melanie (1996). An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press.
[21] Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7.
[22] Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441
[23] Garson, G. D. (2008). Discriminant function analysis. "Archived copy". Archived from the original on 2008-03-12. Retrieved 2008-03-04.