

# AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL REACTION OPTIMIZATION

Turkish Online Journal of Qualitative Inquiry (TOJQI)  
Volume 12, Issue 6, July, 2021: 7870-7880

Research Article

## An Improved Framework For Content Based Image Retrieval Based Chemical Reaction Optimization

Dr K. Sankar<sup>1</sup>, Dr A. Syed Musthafa<sup>2</sup>, Dr P. Chandra Kanth<sup>3</sup>, Mr. Rajesh Kumar Bhavani<sup>4</sup>

### Abstract

With regards to the areas pertaining to image processing and machine learning, Content Based Image Retrieval (CBIR) system has become the main area of research in the recent years. Low-level attributes such as shape, color and texture are in general analyzed through CBIR systems. While describing a data collection, combination of features are selected among a given larger set through feature selection. While classifying and categorizing, valuations of feature performance and selection techniques are carried out. The Chemical Reaction Optimization technique has a better search capacity and can solve NP hard optimization problems. To be more specific, many problems with high efficiency are addressed through this technique. This work involves a novel classifier called modified AdaBoost and in this proposed technique, the number of classification trees and maximum depth per tree is optimized with the help of CRO.

**Keywords:** Content-Based Image indexing and Retrieval (CBIR), Feature selection, modified Adaboost and Chemical Reaction Optimization (CRO).

### 1. INTRODUCTION

Of late, large volume of digital images are more often used in areas such as architecture, crime prevention, medical and military. Appropriate information is accessed from the image and the images are retrieved from large databases. Thus IR [1] from large database is an important issue. This is a computer system generated for the purpose of browsing, searching and retrieval of images from a large database of digital images. Certain methods of addition of metadata such as captioning, keywords or description are the most common conventional techniques of Image

---

<sup>1</sup>Associate Professor, Department of Computer Science and Engineering, Audisankara College of Engineering & Technology, Gudur, Andhra Pradesh

<sup>2</sup>Associate Professor, Department of Information Technology, M.Kumarasamy College of Engineering, Karur, Tamil Nadu

<sup>3</sup>Assistant Professor, Department of Computer Science and Engineering, Audisankara College of Engineering & Technology, Gudur, Andhra

<sup>4</sup>Assistant Professor, Department of Computer Science and Engineering, Audisankara College of Engineering & Technology, Gudur, Andhra

E-Mail Address: sankarkaliappan@gmail.com

Retrieval, so that recovery take place through annotation words. It is quite costly, difficult and time consuming to use manual image annotation.

For retrieval of similar images, CBIR is used to retrieve similar images from image database based on image features that are automatically derived. The basis for retrieval system is its quality. With regards to information management, in general Image Retrieval and in particular content-based image retrieval are potent areas of research. There are various types of visual information in an image which are hard to extract and manually combine by humans. Instead of Tags and Metadata, on the basis of visual contents of image, search and retrieval take place in CBIR. Significant visual contents are colour, shape features and texture [2]. For CBIR, color is the more commonly used visual attribute and they play a primary role in perception of mankind.

Another attribute of importance in an image which is extracted for IR is texture. This refers to the patterns of surface which show granular details of an image [3]. Information on arrangement of various colors are given here. For example, patterns seen in grass fields and block walls vary from one another.

One of the important stages in CBIR system is feature selection and it helps in improving semantic image retrieval results through mitigation the complexity of the process of recovery and enhancing overall system efficiency. In CBIR systems, the main objective is to reduce complexity, improve precision and retrieval results and specifically to improve semantic performance.

Based on various criteria, metaheuristics are classified, of which the most important can be nature-inspired such as GA, artificial immune systems, ACO, and PSO; and non-nature inspired such as Tabu Search, iterated local search, and so on; and based on the number of solutions that are processed as population based and trajectory based metaheuristics. As every metaheuristic implements diversification and intensification strategies differently, there are both advantages and also disadvantages. It is thus shown that a successful solution is to create hybrid metaheuristics which can combine just the advantages of classical metaheuristics.

Many existing evolutionary algorithms are outperformed by the chemical reaction optimization. This has been applied with success to quadratic assignment and channel assignment problem in wireless mesh network, traveling salesman problem, knapsack problem, heterogeneous computing environments, scheduling scheme on heterogeneous computing systems, network coding optimization, modular neural networks applied in emotion classification, fuzzy controller design for mobile robots [4], producing a hybrid method for optimization, and many other problems.

Tabu Search (TS) is a high level local search technique which utilizes relevant techniques in guiding the search targeting to intelligent exploration of search space, simultaneously avoiding getting trapped in local optima. An adaptive intelligent memory is used by TS under the name of

## AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL REACTION OPTIMIZATION

“list Tabu” in order to avoid searching towards paths of solutions that are already explored and to direct the search towards non-visited area, so as to save the already visited solutions.

Adaptive Boosting algorithm, otherwise called AdaBoost adaptively adjusts to errors which are returned by classifiers through previous iterations. Equal weight is assigned by the algorithm to every training instance at the beginning. Then a classifier is built by the application of learning algorithm to training data. There is increase in weights of misclassified instances, where there is decrease in weights of correctly classified instances. Thus, there is concentration of new classifier that is incorrectly classified in every iteration.

Feature selection based CRO-TS is proposed in this work. Literatures related to research are explained briefly in section. Techniques and methods proposed are elaborated in section 3 and section analyses the results in detail and setion 5 concludes the study.

### 2. LITERATURE SURVEY

The issue of imbalanced dataset is suffered by most of the relevance feedback based algorithms where the numbers of irrelevant images are larger to a considerable extent by the number of relevant images to train the classifier. Degradation in retrieval results in imbalanced dataset problem. To handle this problem of imbalanced dataset, Bhosle &Kokare [5] proposed a long-term learning approach with its basis on Radom Forest Classifier ensemble. The user feedback information is collected by long-term learning relevance feedback approach to gain scientific knowledge on database images. This knowledge is gotten by random forest classifier, in order to enhance the outcome of retrieval. The results of the study showed that there is very good improvement in classification accuracy when contrasted with the existing technique that is available in literature. An accuracy of 93% has been observed in nine iterations of the relevance feedback.

Accuracy of CBIR systems is enhanced by feature selection with the help of binary gravitational search algorithm. In the area of pattern recognition, CBIR is the most challenging issue, where CBIR does not depend on the features which are extracted from images. Thus to select the most relevant features, semantic gap between high and low level features lead to higher accuracy. One of the most recent heuristic search algorithms is gravitational search algorithm which was introduced by Rashedi&Nezamabadi-pour [6], with GA and binary PSO in feature selection. Examination of the proposed technique is done in Corel database. Outcome of the study show that the efficiency of BGSA increased the precision of CBIR systems.

A signature in calculating final similarity among query and database images was proposed by Benloucif&Boucheham [7]. For every given query, search is personalized through this approach. Based on metaheuristics in a learning step, Feature Selection (FS) is used before searching. For every learning image, best signature weightings are established upon this step. In the query phase, a communication is established between signatures and weightings which are developed from learning phase weightings. “Greedy Heuristics” is the first developed learning

method in the learning phase which is made up of “Tabu Search” and “Genetic Algorithm.” The basis for the evaluation of used approach is “Corel-1K image database” (Wang image database). The results were listed with regards to “Weighted Precision.” Outcome showed that in the context of CBIR, FS technique is very powerful. Moreover, outcome of the study showed that the comparable results were gotten by the three metaheuristics compared to the published works of the same class.

A new approach which combines chemical reaction optimization framework was proposed by Dam et al., [8] with the Unified Tabu Search (UTS) heuristic to solve the Capacitated Vehicle Routing Problem (CVRP). CVRP is considered as the best algorithm as the efficiency of new algorithm could be evaluated and optimization methods. Nature of chemical reactions is imitated through CRO which is a new optimization framework. Solutions in these operations are educated through the design of elementary chemical reaction operations and the adaptation of UTS algorithm was presented. Finally, an in depth research has been conducted through well-known benchmark problems. Results of the experiments show that the proposed algorithm is effective highly competitive in comparison with several efficient algorithms for this problems. Similar algorithms are presented to address other routing variants in the presented methodology.

A hybrid algorithm CROTS (Chemical Reaction Optimization combined with Tabu Search) was proposed by Yan et al [9] to address the problem. At first, one of the four basic reaction is performed and afterwards Tabu search was employed in searching for neighbors of optimum solution in the population. Results of the study show that CROTS can perform better in comparison with GA and original CRO.

BP is replaced in training neural networks through a newly developed global optimization technique proposed by James et al., [10]. This is a population-based metaheuristics which imitates the transition of molecules and their interaction in chemical reaction. Results of study showed that CRO outperforms many EA strategies commonly used in training neural networks.

### 3. METHODOLOGY

The problem of feature selection depends on the selection of feature subset, within a larger set. It can be used (i) to simplify and understand the model, so as to make it user-friendly; (ii) in reducing the time of computation of algorithms which exploit the data; (iii) to mitigate overfitting and to reduce specialization of the model to known observations. In classification, the issue of feature selection can be prototyped as a combinatorial problem, primarily because it works in selecting a subset of features among  $N$  ( $2^N$  possible subsets are present) and secondly the quality of subset needs to be assessed (for instance, through the quality of classification model constructed with this subset). However, if an elaborate one is used, the use of a classifier in constructing a model might prove to be expensive. This can be considered as an issue in optimizing approaches in dealing with large datasets. This section details the feature selection

## AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL REACTION OPTIMIZATION

based Chemical Reaction Optimization (CRO), Tabu Search (TS), Adaboost and Radom forest classifier, hybrid TS-CRO.

### 3.1 Feature Selection based Chemical Reaction Optimization (CRO)

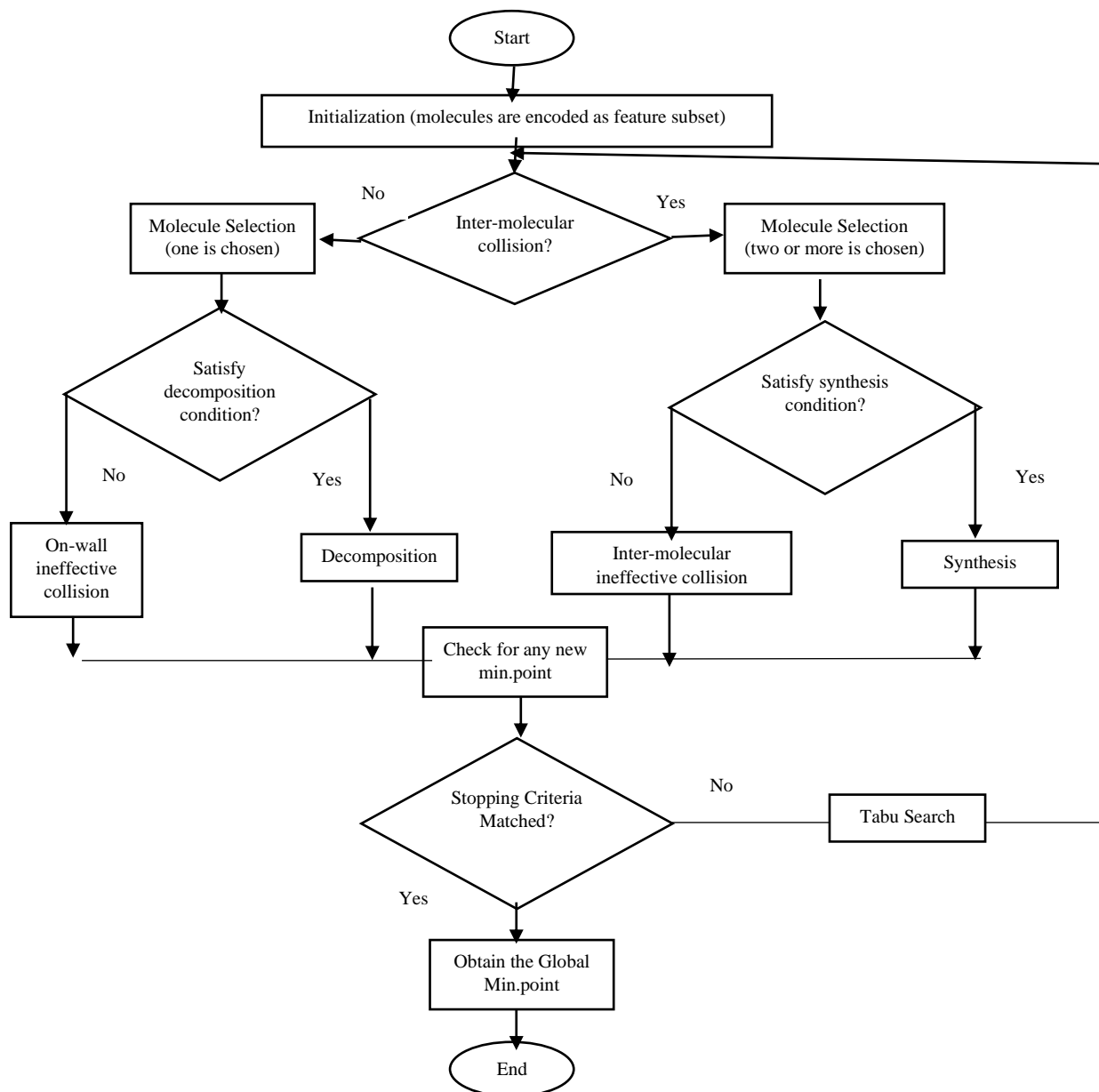
Data dimensions are decreased by selecting a subset of initial features through feature selection methods. Two candidates are selected through the best subset. In all these techniques, the subset that can evaluate the value of function will be selected as a solution. Best features are chosen despite each method with regards to the extent of possible answers and the answer sets increase by  $N$ , it is hard to identify optimal solution and is expensive in medium and large. With an idea of solving this issue, it has been referred to one of the intelligent technique named as Chemical Reaction Optimization (CRO).

General optimization framework is CRO, which is a dedicated problem specific heuristics which should be integrated into four elementary reactions. There will be low efficiency, if decomposition and synthesis operations which are not designed properly. A variable population-based metaheuristic is CRO [8]. It imitates the chemical reaction process where interactions of molecules which go toward the minimum state of free energy, which is similar to objective function in optimization problems. A molecule ( $w$ ) has potential energy ( $PE$ ), kinetic energy ( $KE$ ), hit numbers and other optional characteristics to represent a solution of the considered problem. The two key properties attached to a molecule are  $PE$  and  $KE$ . The fitness value of the solution corresponds to the former while the latter controls the toleration to new solutions with worse fitness. Four types of chemical reactions are implemented through CRO which include on-wall ineffective collision, decomposition, intermolecular ineffective collision and synthesis. Here single molecule reactions include on-wall ineffective collision and decomposition reactions where inter-molecular ineffective collisions and synthesis reactions are multiple molecule reactions. Exploitation is mainly produced by on-wall ineffective collision and inter-molecular ineffective collision, while the two remaining reactions gain exploration. After doing a number of reactions, the potential energy changes to the lowest state and the best solution is the molecule with the lowest  $PE$ .

There are three stages in CRO framework: initialization, iteration and final stage. In the first/initial population, initial values of parameters  $KE_{lossRate}$ ,  $InitialKE$ ,  $PopSize$ ,  $MoleColland$  buffer are generated. In the second stage, the process of reacting is simulated. Here, based on the condition, there are four types of elementary reactions that occur. They include on-wall ineffective collision, decomposition, inter-molecular ineffective collision and synthesis. The first two operators are uni-molecular collisions and the remainder inter-molecular collisions. Here, global search operators include decomposition and synthesis and local search operators are on-wall ineffective collision and inter-molecular ineffective collision.

As mentioned earlier, the issues involving feature selection include varying number of features which are chosen to decrease subset calculations. In terms of its performance, these subset should be the best possible subset, when compared to the initial subset. So, based on CRO, the components from feature selection problem is written and this initializes the encoded population as a subset of feature selection. There is creation in the number of molecules and there is random selection of one among them for decomposition process, or else synthesis process is undertaken. Once minimum point solution is found, the best output is obtained.

Glover [12] invented Tabu Search which is used to solve a wide range of hard optimization problems. This is an iterative procedure which is created to resolve optimization problems. TS begins with random solution and the fitness solution is evaluated for the given solution, after which all possible neighbors are generated and evaluated. A neighbor can be defined as solution which is reached from current solution through simple basic transformation. If best of these neighbors is not in the list, then the new current solution is picked. Figure 2 shows the Tabu search representation.



**Figure 1 Flowchart for proposed Chemical Reaction Optimization (CRO)-Tabu Search (TS)**

## AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL REACTION OPTIMIZATION

Previously explored solutions are tracked by Tabu list which avoids TS from revisiting them again. TS will go uphill, if the best neighbor solution is worse than current design. Local minima can be overcome in this manner. Any reversal in these solutions is forbidden and is classified as Tabu. Some aspiration criteria is introduced which overrides Tabu's status and is found to lead a better fitness with regards to the fitness of current optimum. The algorithm comes to a halt, if no neighbors are present or when there is no improvement in the predetermined number of iterations, otherwise, algorithm continues TS procedures.

In local search algorithms such as Tabu Search, it can be time consuming to explore the neighborhood of a solution. After every iteration, non-Tabu neighbors are evaluated during exploration of neighborhood. In the problem of feature selection, solution is evaluated through the application of classification procedure (KNN, SVM....). This can be computationally expensive when number of observations and features become large. At each iteration, the evaluation of the whole neighborhood is considered. So, this issue is solved through AdaBoost classifier.

Local search performance is gotten better through Tabu Search through combination of CROTS. After the conduction of each reaction, optimum solution in the population bestSol is obtained and TS is employed to search its neighborhoods and update bestSO but structure with more profits.  $fitness(\omega)$  is used to calculate molecule  $\omega$ 's fitness value,  $neighborMol(\omega)$  changes an item of  $\omega$  in a random position and returns a new molecule structure;  $tabuTableUpdate(\omega)$  add molecule structure  $\omega$  to Tabu table, which is an FIFO data structure;  $judge(\omega)$  is used to judge whether solution  $\omega$  is in the Tabu list, if  $\omega$  is not in the Tabu list, then return 0; or return 1; Parameters  $tabuLength$  is the length of Tabu table and  $numNeighbors$  number of the bestSol's neighbors.

### 3.2 Random Forest (RF)

Breiman [13] proposed RF classifier as an ensemble of binary decision trees which is used for classification and regression problems. Ensemble can be defined as a collection of classifiers where decisions of individual classifiers are combined in a certain manner to decide the class of input data. Every tree in RF is trained with the help of random vector that is sampled independently from training set, but in the same distribution. But majority of votes determine the class of input vector from all trees in ensemble. Let T be the training set consisting of N labeled samples as  $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ . Vector  $X = (x^1, x^2, \dots, x^p)$  is a vector consisting of explanatory variables and  $X \in \mathcal{R}^p$ .  $Y \in (+1, -1)$  is a class label of samples. Then the objective is to design the classifier  $\hat{h}$  which performs the mapping  $\hat{h}: \mathcal{R}^p \rightarrow Y$ . Random forest R is a collection of such m classifier, defined by equation (2):

$$R = \{\hat{h}(x, \Omega_k), k = 1, \dots, m\} \quad (2)$$

Where  $x$  is an n-dimensional random sample of random vector X and  $\Omega_k$  are the random vectors chosen independently from the training set T. Thus every tree is grown with the help of randomly selected input variables set having dimension  $\check{n}$ , from the bootstrapped training set of T such that  $\check{n} < n$ . The random vector  $\Omega_k$  for the kth tree is generated with independent of previous

random vectors are the random vectors  $\Omega_1, \Omega_2, \dots, \Omega_{k-1}$ . The classifier R then can be used for the classification of the input data.

### 3.3 Proposed AdaBoost Classifier

Base learner’s prediction accuracy is improved this most popular ensemble method. This is a learning algorithm which can generate multiple classifiers and use them in building the best classifier. This algorithm’s advantage is that less input parameters are required and little prior knowledge is needed about weak learner. Moreover, there is higher flexibility that is suited to combine with other methods to find weak hypotheses. AdaBoost help in predicting classification tasks and also in presenting self-rated confidence score that can estimate the reliability of predictions. Less knowledge of computing is required by this algorithm, to enhance accuracy of models over datasets. AdaBoost classifier is defined by equation (3):

$$H(z) = \text{sgn} \left( \sum_{t=1}^L \gamma_t c_t(z) \right)$$

where L is the quantity of weak learners

$c_t$  is the learner

$\gamma_t$  is the weight

(3)

At each iteration, classifier chooses a new hypothesis  $c_t$  to mitigate the error that is encountered in earlier rounds. AdaBoost’s performance is dependent on weak learner c. While final classifier H improve the performance in selecting weak learners, classifiers cannot be used if feature space is high. Choosing h is NP hard, CRO is used to optimize H.

To characterize a particular class, a large set of features is assumed to be available. This technique starts by training AdaBoost classifier for every feature to be used in fitness function. Then feature selection is applied in finding an optimal subset of features. Thus, the algorithm that is proposed is depicted in two steps:

Step 1: Train Adaboost classifier for each feature.

Step 2: Use the CRO-TS to select the best classifier or feature subset.

### 3.2 RESULTS AND DISCUSSION

For experiments, four classes and 750 images were considered for the proposed feature selection based CRO-TS with random forest, Adaboost, CRO-Modified Adaboost and CRO-Tabu Modified Adaboost. Table 1 and figure 2 to 4 shows the Recall, Precision and F Measure respectively.

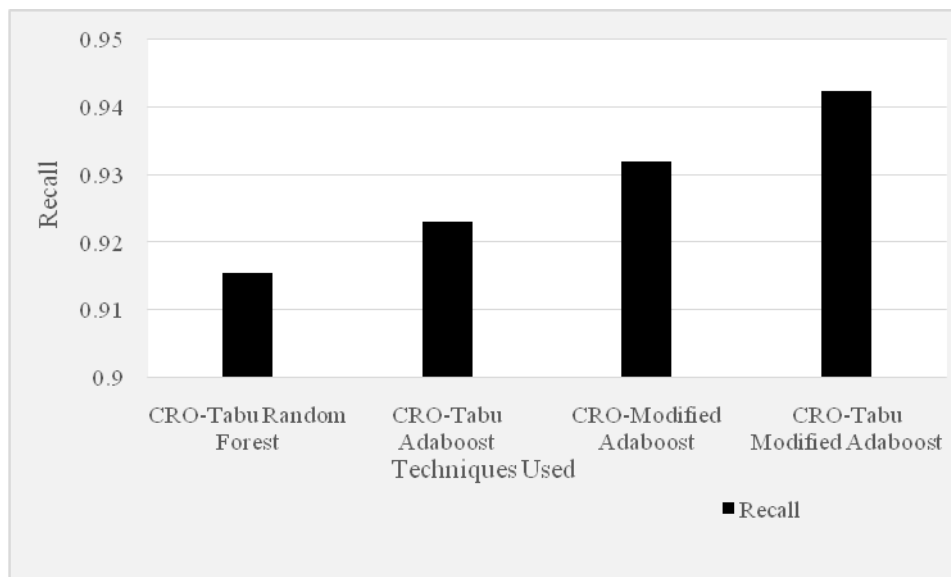
**Table 1 Results**

	<b>CRO-Tabu Random Forest</b>	<b>CRO-Tabu Adaboost</b>	<b>CRO-Modified Adaboost</b>	<b>CRO-Tabu Modified Adaboost</b>
--	-----------------------------------	------------------------------	----------------------------------	---

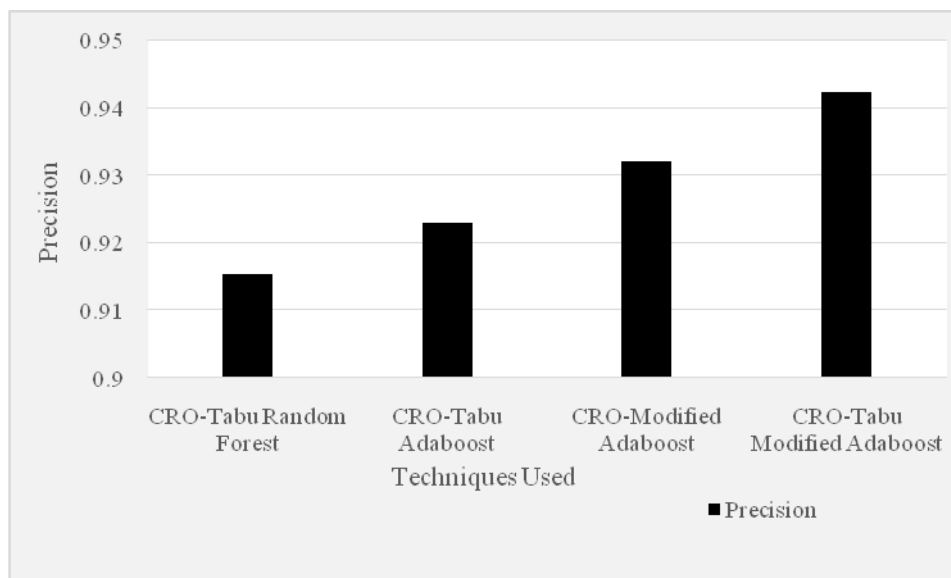


**AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL REACTION OPTIMIZATION**

Recall	0.915325	0.923	0.932	0.942325
Precision	0.9153	0.922975	0.932025	0.94235
F Measure	0.9153	0.92295	0.93195	0.9423



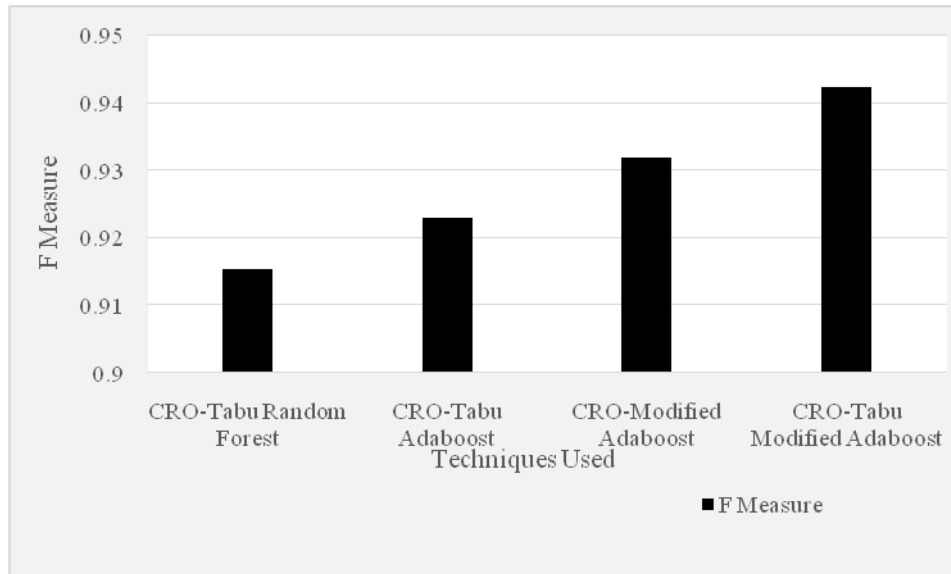
**Figure 2 Recall for CRO-Tabu Modified AdaBoost**



**Figure 3 Precision for CRO-Tabu Modified AdaBoost**

Table 1 and figure 2 shows that the recall for CRO-Tabu Modified Adaboost performs better than CRO-Tabu Random forest by 2.9%, than CRO-Tabu AdaBoost by 2.07% and better than CRO-Modified AdaBoost by 1.1%.

Table 1 and figure 3 shows that the precision for CRO-Tabu Modified Adaboost performs better than CRO-Tabu Random forest by 2.91%, than CRO-Tabu AdaBoost by 2.08% and better than CRO-Modified AdaBoost by 1.1%.



**Figure 4 F Measure for CRO-Tabu Modified AdaBoost**

Table 1 and figure 4 shows that the F Measure for CRO-Tabu Modified Adaboost performs better than CRO-Tabu Random forest by 2.91%, than CRO-Tabu AdaBoost by 2.07% and better than CRO-Modified AdaBoost by 1.1%.

#### 4. CONCLUSION

CBIR is considered as one of the most effective way of retrieving visual data. FS is basically an optimization problem, where weightings assigned to the different images signatures shall yield the best possible similarity between the query and the retrieved images. The Tabu Search is a local search that uses a memory to escape from local optima. This work combines the CRO based feature selection with TS to produce an optimal solution. Results show that the recall for CRO-Tabu Modified Adaboost performs better than CRO-Tabu Random forest by 2.9%, than CRO-Tabu AdaBoost by 2.07% and better than CRO-Modified AdaBoost by 1.1%..

#### References

- [1] Nandhini, V. G. K. Content Based Image Retrieval using Feature Selection Techniques.
- [2] Gavves, E., Snoek, C. G. M., & Smeulders, A. W. M. (2012). Visual synonyms for landmark image retrieval. *Computer Vision and Image Understanding*, 116(2), 238-249.
- [3] Krishnamoorthi, R., & Sathiyadevi, S. (2012). A multi-resolution approach for rotation invariant texture image retrieval with orthogonal polynomials model. *Journal of Visual Communication and Image Representation*, 23(1), 18-30.
- [4] de la O D, Castillo O, Meléndez A, Melin P, Astudillo L, Sánchez C, (2015) Optimization of reactive fuzzy controllers for mobile robots based on the chemical reactions algorithm. In: Melin P, Castillo O, Kacprzyk J (eds) Design of intelligent systems based on fuzzy logic, neural networks and nature-inspired optimization, studies in computational intelligence, vol 601. Springer, New York, pp 253–266.

**AN IMPROVED FRAMEWORK FOR CONTENT BASED IMAGE RETRIEVAL BASED CHEMICAL  
REACTION OPTIMIZATION**

- [5] Bhosle, N., &Kokare, M. (2016, October). Random forest based long-term learning for content based image retrieval. In *Signal and Information Processing (IconSIP), International Conference on* (pp. 1-4). IEEE.
- [6] Rashedi, E., &Nezamabadi-pour, H. (2012, May). Improving the precision of CBIR systems by feature selection using binary gravitational search algorithm. In *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on* (pp. 039-042). IEEE.
- [7] Benloucif, S., &Boucheham, B. (2014, November). Impact of feature selection on the performance of content-based image retrieval (CBIR). In *ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb), 2014 4th International Symposium* (pp. 1-7). IEEE.
- [8] Dam, T. L., Li, K., & Fournier-Viger, P. (2017). Chemical reaction optimization with unified tabu search for the vehicle routing problem. *Soft Computing*, 21(21), 6421-6433.
- [9] Yan, C., Gao, S., Luo, H., & Hu, Z. (2015, June). A hybrid algorithm based on tabu search and chemical reaction optimization for 0-1 knapsack problem. In *International Conference in Swarm Intelligence* (pp. 229-237). Springer, Cham.
- [10] James, J. Q., Lam, A. Y., & Li, V. O. (2011, June). Evolutionary artificial neural network based on chemical reaction optimization. In *Evolutionary Computation (CEC), 2011 IEEE Congress on* (pp. 2083-2090). IEEE.
- [11] Astudillo L, Melin P, Castillo O (2015) Introduction to an optimization algorithm based on the chemical reactions. *Inf Sci* 291:85–95 Lam AY, Li VO (2012) Chemical reaction optimization: a tutorial. *MemetComput* 4(1):3–17.
- [12] Glover, F., 1986. Future paths for Integer Programming and Links to Artificial Intelligence”. *Comput. Oper. Res.* 5, 533–549.
- [13] Breiman Leo, "Random Forests", *Machine Learning*, volume 45, no.1, pp. 5-32.