

# An Efficient Medical Content Based Image Retrieval with Lenient Relevance Feedback

R.I. Heaven Rose<sup>1\*</sup>, A.C. Subajimi<sup>2</sup>

<sup>1</sup>Research Scholar, Department Of Computer Application, Noorul Islam Centre For Higher Education, Kumaracoil, Thuckalay, Kanyakumari District, Tamilnadu, India.

<sup>2</sup>Professor, Department Of Computer Application, Noorul Islam Centre For Higher Education, Kumaracoil, Thuckalay, Kanyakumari District, Tamilnadu, India.

\*Corresponding Author E-Mail: [Heajeysa@Yahoo.Com](mailto:Heajeysa@Yahoo.Com)

## Abstract

Content Based Image Retrieval (CBIR) for medical imageries is still in its early stage. There are many challenging research issues. Retrieve similar images only is the current problem in medical CBIR. One idea to solve this difficult is minimizing the gap among two descriptions i.e. low level extracted features of image and high level human perception of image. There are various Relevance Feedback (WF) methods have been considered to minimize the semantic gap in medical CBIR system. But most of them were deals with hard Feedback. In Hard Feedback system user can interact with the system in one query session. We recommend to aid the usage of lenient relevance response to better capture the intention of users. The meta-knowledge mined from multiple user's experience be able to increase the precision of subsequent image recovery results. Here we suggest an algorithm to mine lenient association rules from the group of suggestion i.e. image weight value given by the user. To reduce the amount of strong rules we offer two rule lessening techniques related to redundancy detection and confidence quantization. Best first search and Binary search methods are similarly applied to advance the procedure of weight interface. The effectiveness of the offered system is assessed regarding precision and average retrieval time. The experimental results on medical images display that the proposed method is able to improve the accuracy of medical CBIR system and reduces the retrieval time than other usual methods.

**Keywords:** CBIR, CBMIR, CAD, association rule, euclidean distance, query vector modification, relevance association rule mining, hard feedback, lenient feedback.

## 1. Introduction

In medicinal field two kinds of computer systems are commonly used; the Content Based Image retrieval (CBIR) and Computer Aided Diagnosis (CAD). CAD [1] is aiming at providing second opinions to radiologist to aid in diagnosis. The aim of CBIR in medicine is to assist physicians in diagnosis, retrieving cases or images to increase the physician's self-confidence on explaining test image details. At present, image mining is one of the demanding field which combines both data mining and image processing techniques. The important task of mining image is successfully connecting pixel based information to object based information. It is difficult to relate the pixel information of image to object information. This is semantic gap. The aim of image retrieval is to reduce the gap of semantic and progress the precision of image retrieval. There are many technologies have been developed to reduce semantic gap such as object ontology, machine learning, relevance feedback, web image retrieval.

In Medicine, the radiologist uses content based image retrieval system to retrieve images from large image database which are similar to the query image. The CBIR method is not suitable for various types of medical images since, it is established for particular features of images. An interesting method in this attempt is called Content Based Medical Image Retrieval or simply (CBMIR) [3], which reclaim images from medical image databases. The CBMIR method focusses the important features of

medical images like haralick features, zernike moments, histogram intensity features and run-length features. There is increasing attention in the use of medical image retrieval systems to help diagnosis using similar past cases. But, selecting appropriate images is one of the challenging problems. One idea to solve this problem is minimizing semantic gap between the low level features of the images and the human interpretation of the image contents. In order to bridge the semantic gap, one should include the user knowledge in to the retrieval system. The most widely used approach for this purpose during the last few years is relevance feedback (RF) [4]. While using the RF approach in CBMIR, the user is queried every time for a proper result so that the images can be classified relevant and irrelevant accordingly [5]. There are many RF methods [6] are proposed and used by various CBIR applications. However, most of the current RF methodologies focusses on strong feedback (either applicable or non-applicable) and the strong feedback motivate specific practice (user interact in the structure in a single request period). Because of the image database is very large and the user may give incorrect feedback, it is essential to bring out an alternative feedback interface to user. Soft feedback interface provides multi-level annotations to user better way to give their suggestions about the query image.

## 2. Related Works

Aswini Kumar Mohanty, Manas Ranjan et al., [7] in their effort, They put on image mining in the area such as breast mammograms to categorize and discover the cancerous tissue. A hybrid approach of feature selection is offered, which about reduces 75% of the features, and novel decision tree is used for classification. The greatest motivating algorithm is that branch and bound algorithm that is used for feature selection delivers the finest best features and no where it is functional or used for gray level Co-occurrence matrix feature selection from mammogram. It has the drawbacks of dimensionality curse problem. Bugatti, P. Ribeiro, et al., [8] suggest a supervised method for continuous feature selection. The suggested method applies statistical association rules to find patterns relating low-level image features to high-level knowledge about the images, and it uses the patterns extracted to determine the weight of the features. Furthermore, the proposed method achieves dimensionality decrease of image features ducking the "dimensionality curse" problem so this method only used for small databases. Baddeti Syam and Yarravarapu Srinivasa Rao, wide features are extracted from the database images and stored in the feature archive. The Squared Euclidean Distance (SED) [10] gives the similarity amount of the Genetic Algorithm. Using the Genetic Algorithm-based similarity measure the database images that are pertinent to the particular query image are recovered. Liang Lei, Jun Peng and Bo Yang [11] gave a transitory outline to image feature selection about the purposes, tasks, and commonly used algorithms. Then, the paper enhanced genetic operator of genetic algorithm, and parallel calculating was used to genetic algorithm, to improve the performance of image feature selection.

In the Query Vector Modification (QVM) [12] approach, the query vector was iteratively reformulated based on the feedback given by the user for moving the query towards the set of relevant images. The Feature Relevance Estimation (FRE) [13] approach explains as some features may have importance than other features based on user's feedback. In this approach, first the relevance weight of each feature is estimated then using Euclidean distance the top n most images can be found. In Bayesian Interface (BI) [14] approach, the Bayesian framework calculates the relevance probability that the retrieved image form the database is relevant to the query feedback given by the user. Due to the probability distribution on all images in the database updated on next individual loop, the method can increase the enactment of retrieved results. SVM-Active learning [15] proposed by Tong and Chang, allows the operator to tag the selected images that be situated nearer toward the SVM boundary to train the SVM. After all the feedback rounds completed, the SVM is trained to retrieve top n imageries that stand for away from the support vector machine boundary. Currently, many researchers focus on long term methods with multi sessions to increase the retrieval accuracy.

Here, we suggest a Lenient Relevance Feedback Association Rule Mining (LRFARM) method for Lenient Feedback image retrieval structure. In this system, the image relevance interface uses the already mined association rules and displays the most pertinent imageries on the screen. If the user is not satisfied with the retrieved images, then the user able to mark the weight of individual image and run the lenient relevance procedure to expand the enactment of the retrieved results.

The organization of the remaining part of this paper is as follows: Section 3 pronounces the proposed SFARM method. Section 4 provides the Experimental Results. The last Section 5 offerings the Conclusion of this work.

## 3. Proposed System

The contribution of the proposed LRFARM model for lenient feedback interface includes the following features: 1) To provide

an efficient medical CBIR, the existing relevance method QVM is revised and proposed a new soft QVM method. 2) Mining lenient image relevance association rules from the user's comment to improve the output of image retrieval. 3) We suggest a soft FP-Growth algorithm with fuzzy Support-Count to improve the retrieval result. 4) We offer two rule optimization methods to decrease the Association Rules size related to redundancy based reduction and Confidence quantization. 5) Generate various rules from Lenient image relevance interface to retrieve most relevant images. 6) To speedup the image retrieved process, best first and binary search algorithms are used. In this system, during the test phase the Lenient image relevance interface uses the already mined Association rules and display the best pertinent imageries on screen. If the retrieved images are not happy to user, then the user is able to classify the relevance level of individual image and run the lenient relevance method to increase the presentation of the retrieved results. The system has the following advantages:

1. Improve the performance of medical CBIR.
2. Decrease the total time required to generate image relevance rules.

### 3.1. Lenient Relevance Feedback Association Rule Mining (Lrfarm) Model

In the proposed system may users can use the current database which contains images, the useful information gathered through image relevance interface can significantly increase the accuracy of the image retrieved results. The proposed method is an efficient approach which includes Lenient association rule mining algorithm to find efficient rules using image relevance weight information gathered from user's feedback. To decrease the rule set size, two rule t optimization techniques are used. Finally, best first search and binary search algorithms are accomplished to speed up the computation of lenient image relevance interface.

#### 3.1.1. Lenient FP-Growth Algorithm

An interesting method to find frequent item set without candidate item set generation is called Frequent Pattern Growth or FP-Growth. This algorithm follows divide and conquer strategy. The strategy includes the following steps: 1) Compasses the data base representing frequent items into FP-tree. 2) Divide the compressed database into a set of conditional data bases 3) Mines each such database separately. In the medical CBIR system with relevance feedback, the transactional database is created by the retrieved images feedback given by the user. The following figure1 shows the model of the transactional database.

TID	Account of image retrievals and relevance weight
1	Img <sub>1</sub> (0.1), Img <sub>2</sub> (0.5)
2	Img <sub>3</sub> (0.5), Img <sub>4</sub> (0.1), Img <sub>5</sub> (0.5), Img <sub>6</sub> (0.5)
3	Img <sub>1</sub> (0.1), Iimg <sub>2</sub> (0.1), Img <sub>3</sub> (0.5)
4	Img <sub>1</sub> (0.1), Img <sub>4</sub> (0.1), Img <sub>5</sub> (0.5)

Figure 1: Transactional database

Here four records are stored and each record contains Transaction Identification (TID), displayed image's ID and the lenient relevance weight of the displayed image. Since the proposed system supports soft feedback, a novel algorithm named soft FP-Growth is proposed to produce lenient image relevance association rules from the set of lenient feedback given by users. The item set's support-count is calculated by adding the weight of lenient image relevance specified in the resultant transaction record. Let us consider the support-count of {Img1} is calculated by support-count {Img1} = 0.1+0.5+0.1=0.7. After finding the set of frequent items, generate the strong association rules which satisfy the minimum confidence threshold. For example, X => Y be a lenient Association rule where X and Y are frequent items then the confidence of the rule can be defined as  $conf(X \Rightarrow Y) = \frac{sup(X \cap Y)}{sup(X)}$  where, X and Y are rule precedence and rule

subsequent. The high confidence value implies that there is a high similar relationship between the image X and image Y. The strong associations rules are generated from the retrieved images are further used to collect most relevant images.

### 3.1.2. Association Rule Reduction

The large size association rule set increase the time of searching process. To reduce the ruleset and also reduce the search time two rule reduction techniques are used related to rule confidence and redundant rule detection.

One of the useful techniques for deriving highly pertinent images through optimized rules is: Collect all the rules whose predecessor is same then Merge these rules those have the same confidence value. To do this, first collect all the confidence values and create cut points with fixed distance breaks. Then the rules are merged since, they have the similar predecessor and the confidence value is in the similar break. For example, let us consider N rules. All rules having the same predecessor X and same confidence value that can be computed as  $X \Rightarrow Y_k$ ,  $k = 1, \dots, n$ . We can merge the above rules into one rule as  $X \Rightarrow A$ ,  $A = \bigcup_{k=1 \dots n} Y_k$  and  $Conf(X \Rightarrow A) = \frac{1}{n} \sum_{k=1 \dots n} Conf(X \Rightarrow Y_k)$ . The length of the interval I should be small in size to improve the retrieval accuracy.

Another important rule set lessening technique is finding redundancy of rule set. If two rules have same predecessor X for example  $X \Rightarrow Y$  and  $X \Rightarrow Z$ , the rule redundancy occurs in the rule set when two rules satisfy any one of the following two conditions.

- When  $Y \subset Z$  and  $conf(X \Rightarrow Y)$  is less than  $conf(X \Rightarrow Z)$ , the rule  $X \Rightarrow Y$  is out of work. So, it is detached from the set.
- When  $P = Y \cap Z$  and  $conf(X \Rightarrow Y)$  is less than  $conf(X \Rightarrow Z)$ , the rule  $X \Rightarrow Y$  is reduced as  $X \Rightarrow Y \setminus P$ . Here,  $Y \setminus P$  means the difference between the rule set Y and the rule set P.

The second reduction technique only removes the redundant rules and will not aim any loss in retrieval accuracy.

### 3.1.3. Image Relevance Inference

Set of similar images to a specific query can be found using the predefined rules mined from the user's feedback through lenient relevance interface. For a given query image p, the following relevance interface rules can be derived:

- Search the first order lenient inference rule. If all the rules  $X \Rightarrow Y$  and  $X = \{p\}$  then assume the relevance of image i,  $i \in Y$ , for the query image p is  $\varphi_p(i) = conf(X \Rightarrow Y)$ .
- Search the first order lenient inference rule. If all the rules  $X \Rightarrow Y$  and  $P \Rightarrow Q$  with  $X = \{p\}$  and  $p \subset Y$  then conclude that the relevance of image i,  $i \in Q$ , for the query image p is  $\varphi_p(i) = conf(X \Rightarrow Y) \times conf(P \Rightarrow Q)$ .

### 3.1.4. Binary Search and Best-First Search

The best first search and binary search algorithms are offered to the proposed method for increasing the speed of the image retrieval task. Before execute these algorithms all the lenient relevance rules are sorted in rule confidence descending order. Next, searching all the first order and second order rules using binary search algorithm. Also, best first algorithm is executed on image retrieval time and select the top n most relevant images. This process continues until the top n most images retrieved or the retrieved images relevance is below the user specified minimum confidence. If more than one images take equal relevance value, then the images will be retrieved related to Euclidean Distance measure value. The innovative procedure for the LRFARM method is defined below:

1. Discover query image key i, and all iteration key j
2. Assume  $X_i^{(j)}$  keep on the present query construction

3. For the query  $X_i^{(j)}$ , regain the nbest pertinent images from the earlier association rules with image relevance interface.
4. Else calculate close images rendering to Euclidean matching score.
5. Repeat step 4 until the user satisfied with the present retrieval outcome
  - a. Identify each retrieved image's the lenient weight level.
  - b. store the present lenient feedback information in the transaction database as a set of items
  - c. Execute the approved soft RF method and recover the top nbest relevant images
6. End

## 4. Experimental Results and Comparative Performances

We have tested our proposed system with the MRI brain images obtained from NAMIC: Public Data (<http://www.insight-journal.org/midas>). The database contains three different classes of brain images like Normal, Benign and Malignant are taken for experiments. To improve the result of image retrieval, many relevance feedback iterations can be executed for every query. Retrieval score is a new measure used to evaluate the standard of soft feedback based image retrieval result. The Retrieval Score is computed by

$$\text{Ret-Score} = \sum_{k=1}^n w_k$$

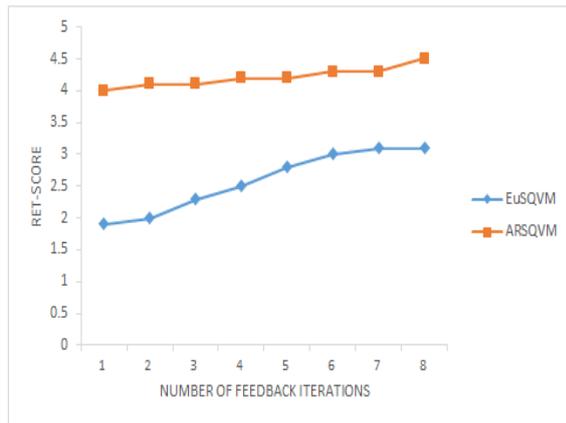
Where, n represents the total number of retrieval images in single iteration.  $w_k$  is the lenient relevance weight of  $k^{\text{th}}$  image. Higher Ret-Score value indicates that the retrieved images are more relevant to the query.

**Table 1:** Execution Time and Ret-Score for the First Result Using Two Matching Techniques

	Euclidean Matching	SFARM(800,820 rules)	SFARM(10,100 rules)	SFARM(4235 rules)
Ret-Score	1.9	3.89	3.9	4
Second/ Query	1.95	1.28	1.25	1.1

The experiment is conducted and the retrieved images of first round are searched by using SFARM and Euclidean Distance matching. The above table 1 shows that the average Ret-Score for Euclidean matching at first retrieved results is 1.9 at the time Ret-Score of SFARM is 4. Hence, the SFARM method expressively progress the performance of first display. These two methods are further experienced by applying SQVM for 8 iterations. Since both these systems uses SQVM, they are referred to as ARSQVM and EuSQVM. We assess the performance of the above two methods regarding Ret-Score value. The average Ret-Score value obtained by using ARSQVM and EuSQVM method at different amount of iterations is shown in the below Figure 2.

While we look at the figure the Ret-Score curve of ARSQVM is higher than the EuSQVM curve at all iterations. Henceforth, the proposed SFARM method can significantly increase the performance of the retrieval system with soft relevance feedback.



**Figure 2:** The average ret-score value gained using ARSQVM and EuSQVM at different number of relevance feedback iterations

## 5. Conclusion

All the traditional feedback methods deal with hard feedback. In the hard feedback method, user can give relevant and non-relevant labels as the feedback to refine the system. In this paper, we propose a novel SFARM model to improve the accuracy of the image retrieval result. The soft FP-Growth algorithm is proposed to generate strong association rules from the lenient relevance of the retrieval results. The strong association rules are generated by using two techniques related to redundancy detection and confidence. The best first and binary search algorithms can increase the speed of the lenient relevance interface computation. The conducted experiments prove that the offered technique greatly progresses the accuracy of the retrieval results than other traditional methods.

## References

- [1] Lin TW, Huang PY & Cheng CWC, "Computer-aided diagnosis in medical imaging: Review of legal barriers to entry for the commercial systems", *IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, (2016), pp.1-5.
- [2] Raghu K & Swarup M, "Content-Based Image Retrieval Based on a Fuzzy Approach", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.10, (2004).
- [3] Prachi B & Cheeran AN, "Content Based Image Retrieval in Medical Imaging", *International Journal of Computational Engineering Research*, Vol.3 No.8, (2013).
- [4] Sunitha J & Sihem M, "Towards Effective Relevance Feedback Methods in Content-Based Image Retrieval Systems", *International Journal of Innovation, Management and Technology*, Vol.5, No.1, (2012), pp. 201435-38.
- [5] Rajalakshmi T & Minu RI, "Improving relevance feedback for content based medical image retrieval", *IEEE International Conference on Information Communication and Embedded Systems (ICICES)*, (2014), pp.1-5.
- [6] Patil PB & Kokare MB, "Content Based Image Retrieval with Relevance Feedback using Riemannian Manifolds", *IEEE Fifth International Conference on Signal and Image Processing (ICSIP)*, (2014), pp.26-29.
- [7] Aswini KM & Manas RS, "A Novel Image Mining Technique for Classification of Mammograms Using Hybrid Feature Selection", *International Journal of Advanced Research in Computer Science*, Vol.2, No.3, (2011).
- [8] Bugatti PR, "Content-Based Retrieval of Medical Images by Continuous Feature Selection", *21<sup>st</sup> IEEE International Symposium on Computer-Based Medical Systems*, (2008).
- [9] Baddeti S & Yarravarapu SR, "An Effective Similarity Measure via Genetic Algorithms for Content Based Image Retrieval with Extensive Features", *Multimedia Tools Applications*, Vol.17, No.17, (2013).
- [10] Durga Rajesh KV & Mukesh D, "Squared Euclidean Distance Matrix: A Heuristic Based Approach for Cell Formation Applications", *International Journal of Applied Engineering Research*, Vol.11, No.10, (2016), pp.7044-7048.
- [11] Liang L, Jun P & Bo Y, "Image Feature Selection Based on Genetic Algorithm", *Proceedings of the International Conference on Information Engineering and Applications (IEA)* (2012).
- [12] Rabeb M, Mohamed T & Hawete H, "A New Relevance Feedback Algorithm Based on Vector Space Basis Change", *International Conference on Intelligent Text Processing and Computational Linguistics*, (2014).
- [13] Hossein A & Abolfazl L, "A Feature Relevance Estimation Method for Content-Based Image Retrieval", *International Journal of Information Technology & Decision Making*, Vol.10, No.5, (2011).
- [14] Tatiana AC & Pavel PC, "Image Retrieval Using Multiple Evidence Ranking", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.4, (2004).
- [15] Begüm D and Lorenzo B, "Fellow, A Novel Active Learning Method in Relevance Feedback for Content-Based Remote Sensing Image Retrieval", *IEEE Transactions On Geoscience And Remote Sensing*, Vol.53, No.5, (2015).