Extracting valuable associations among textural features of medical images

Fatma Helmy Ismail
Misr International University
Egypt
fatma.helmy@miuegypt.edu.eg

Aboul Ella Hassanien
Cairo University
Egypt
aboitcairo@gmail.com
Scientific Research Group in Egypt (SRGE)
http://www.egyptscience.net

***Abstract*—As a matter of fact, the textural features extracted from medical images have physical meaning. Some of them express the contrast, the uniformity, the homogeneity, the distortion and the suavity of the image. The idea of information digging for finding associations among these features in retina images is introduced. The proposed technique works in three stages. In the first stage, the Gray Level Co-event Matrix (GLCM) textural features are extracted and averaged for the 0 $^{\circ}$, 45 $^{\circ}$, 90 $^{\circ}$and 135 $^{\circ}$directions. In the second stage, feature selection is adopted using The relief algorithm to choose the most relevant textural features and discretize them to fall between two ranges of values (high and low). Finally, the FP-growth algorithm is applied on the selected features to discover useful associations among them. The results show that the normal retina images are highly associated with certain textural features in a certain range of values. Hence, These associations can be utilized for effective diagnosis of normal retina images.**

***Keywords*—Grey Level Co-occurrence Matrix (GLCM), Association Rules, FP-growth, Image Mining**

I. Introduction

Nowadays, medical images are investigated to find valuable data that guide the grouping of unclassified images. Therefore, there is a major need of image mining frameworks. The way toward diagnosing retina ailments has turned out to be basic. In this paper, we extract association rules, which speak to visit repetitive patterns that happen frequently in normal/diseased retina images and these rules can be utilized in the future for compelling determination of retina illnesses. Our proposed technique utilizes textural features of images. The proposed approach contains three stages. The main stage is textural features extraction and choice and the second stage is discretizing the extracted features. The last stage is to discover frequent patterns in retina images utilizing FP-growth algorithm of association rule mining. Agrawal et al. [1] very first time introduced the apriori algorithm to discover association rules from a transaction dataset. Beyer et al. [2] presented that problems arouse when the count of features that portray the image increases. Thus reducing the number of features becomes a mandatory task as well as improving the accuracy of classification. Textural features in retina images speak to the distinctions in thickness of retina tissue. Properties, for example, roughness, smoothness and regularity are portrayed via textural features. A database of images can be questioned by utilizing textural features to recover comparative patterns. The spatial order of pixel intensities describes textural information. [3] [4]. [5] introduced a method to discover effective and strong association rules from medical images using apriori algorithm. Carson et.al. [6] introduced a new image presentation which offers a transformation from raw pixel data to a small set of localized coherent regions in color and textural space. Ji Zhang et al. [7] presented various image mining research issues in image mining. A framework for texture information of an image and achievement of higher retrieval efficiency than the shape features of an image is presented by Monika Sahu et al. [8] . Marcela Y. Ribeiro et al. [9] introduced a method to improve a mammogram classification using association rule mining. This method produces two types of association rules, non-sensitive and sensitive association rules. These non-sensitive association rules are not useful for the diagnosis process. Also they are finding region of interest (ROI) of mammogram manually and then to these ROI, feature extraction techniques were applied. Maria-Luiza Antonie et al. [10] [11] [12] proposed a mammogram classification method using association rules. the rough set theory combined with association rules mining is used for mammogram clarification by Jiang Yun et al. [13]. Sumeet Dua et al. [14] proposed a classification method relying on weighted association rules. This method takes advantage of the inter-class and intra-class weight of each association rule for classification. Jawad Nagi et al. [15] introduced a technique utilizing morphological processing and seeded region growing algorithm for automatic segmentation of breast tissues. Albeit a large number of the specialists have created distinctive procedures for mining of mammogram images to discover solid and proficient association rules, still it is a testing assignment. Subsequently we proposed a framework to discover association rules among textural features extracted from retina images. The remaining part of the paper is divided into the following; section 2 introduces the framework proposed for image mining method. Section 3 introduces the experimental outcome. Section 4 introduces the conclusion and the future work.

II. Proposed Image Mining Method

The introduced framework contains three phases. The first phase is extracting (GLCM) textural features from retina images. The second phase is selecting the most relevant textural features. The final phase is preparing transactional database from the set of the selected features and generating association rules from them using FP-growth. Each input retina image is associated with a class i.e. normal or cotton wool diseased image.

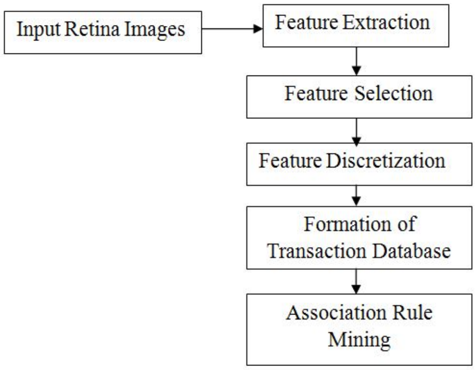


Fig. 1: Flow chart of the proposed system

A. GLCM Feature Extraction

Textural Features are extracted from retina images that are classified into two classes (normal/cotton wool diseased images) and organized into feature vectors. In this step, several state of the art works are studied to choose the most powerful features that can improve the classification results. The following features are chosen as they achieve the most reliable results: auto-correlation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation, inverse difference, inverse difference normalized and inverse difference moment normalized. All the listed features are extracted and averaged for the 0 °, 45 °, 90 °and 135 °angles.

B. Feature Normalization

In order to apply the feature selection algorithm, the data have to be scaled to a fixed range (usually 0 to 1). A Min-Max scaling is typically done via equation 1:

$$X_i = \frac{X_i}{X_{min} - X_{max}} \tag{1}$$

Where X_i is the original feature vector, X_{min} is the minimum value of that feature vector, and X_{max} is the maximum value of it.

C. Feature selection

The Relief is an algorithm developed by [16] [17] in 1992 that takes a filter-method approach to feature selection that is notably sensitive to feature interactions. The pseudo code of the algorithm is listed.

D. Discretization of Features

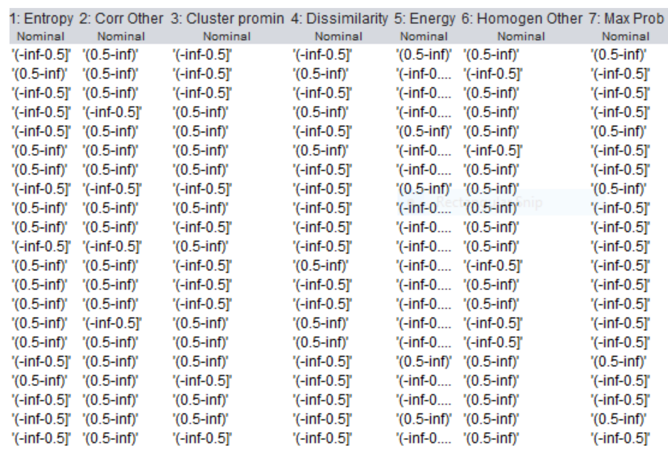
The selected features values are discretized. Fig 2 and 3 show a sample of the discretized features into two bins values the first bin value ranges from -inf to 0.5 and the second bin value is larger than 0.5. Therefore, each feature value lies inside one of the two ranges the first range below 0.5 (low value) and the second range above 0.5 (high value).

Algorithm 1 Relief Algorithm for Feature Selection

Input/Output: Input dataset with n instances and p features belonging to two known classes (normal and cotton wool diseased image), threshold value c . **Output** selected features with weight value w higher than c.
initialization Set the iteration number i=1 and the total number of iterations to M. initialize a p-long weight vector w to zero.
while $i \leq M$ **do**
    Take the feature vector x belonging to one random instance, and the feature vectors of the instance closest to x (by ecludian distance) from each class. The closest same class instance is called near-hit (nh) and the closest different class instance is called near-miss (nm). such that
    update the weight vector

$$w_i = w_i - (x_i - nh_i)^2 + (x_i - nm_i)^2 \tag{2}$$

    i=i+1.
end while
divide each element of the weight vector w by M
select features with weight value w higher than c.



1: Entropy	2: Corr Other	3: Cluster promin	4: Dissimilarity	5: Energy	6: Homogen Other	7: Max Prob
Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'
'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'
'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'
'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'
'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'
'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(-inf-0....	'(0.5-inf)'	'(-inf-0.5]'

Fig. 2: Snapshot of some features values

E. Extracting Association rules using FP-growth algorithm

The original definition of association rule was first discussed in [1] as: assume

$$I = i1, ..., in$$

be a set of n literals called items. Let

$$D = t1, ..., tm$$

be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. An association rule is presented as indicated

$$A \to B$$

. $A, B \subseteq I$ and $A \bigcap B = \phi$. A is called the head of the rule and B is called body of the rule. An item set is a set of items contained in the antecedent and the consequent. Support and confidence values are used determine the importance of

8: Diff Var	9: Diff Entr	10: InfoMes 2	11: InvDiff Norm	12: InvDiff Mm Norm	13: **Norm/Disea**
Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	Cotton Wool
'(-inf-0....	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	Cotton Wool
'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	Cotton Wool
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	Cotton Wool
'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	Cotton Wool
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	Cotton Wool
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	Cotton Wool
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	Cotton Wool
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	Cotton Wool
'(-inf-0....	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	Cotton Wool
'(0.5-inf)'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	Normal Image
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	Normal Image
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	Normal Image
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	Normal Image
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	Normal Image
'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	Normal Image
'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	Normal Image
'(-inf-0....	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	'(0.5-inf)'	Normal Image
'(0.5-inf)'	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(-inf-0.5]'	Normal Image
'(-inf-0....	'(-inf-0.5]'	'(-inf-0.5]'	'(0.5-inf)'	'(0.5-inf)'	Normal Image

Fig. 3: Snapshot of other features values

the rules generated by the mining process 3 4. Support value defines how repetitivly an association rule is valid in a given transaction data set. Confidence value defines how often items in B appear in transactions that contain A.

$$support, S(A \rightarrow B) = \frac{No.oftuplescontainingbothAandB}{Totalno.oftuples} \tag{3}$$

$$confidence, C(A \rightarrow B) = \frac{No.oftuplescontainingbothAandB}{Totalno.oftuplescontainingA} \tag{4}$$

It is required to discover the association rules with support and confidence values that are larger than the support and confidence values supplied by the user.Records of retina images are input to FP-growth algorithm, to discover the association rules. There are many other correlation measures to evaluate association rules significance, we select interestingness correlation measures such as lift 5 and conviction 6.

$$Lift(A \rightarrow B) = \frac{confidence(A \rightarrow B)}{support(B)} \tag{5}$$

$$Conviction(A \rightarrow B) = \frac{1 - support(B)}{1 - confidence(A \rightarrow B)} \tag{6}$$

III. Experimental Results

The images are taken from the DR1 dataset created by the Department of Ophthalmology, Federal University of Sao Paulo. The number of images in each class (normal/cotton wool diseased image) is 70. The relief algorithm selected only 12 features: maximum probability, energy, information measure of correlation, difference entropy, correlation, entropy, dissimilarity, inverse difference normalized, inverse difference moment normalized, cluster prominence, homogenity and difference variance. The extracted association rules show high association between certain values of textural features and normal images. Thus, these associations can guide the diagnosis of normal images. The following rules show some of the extracted associations along with their confidence, lift and conviction values.

1) Inverse difference normalized is high and the image is normal $\rightarrow$ Homogeneity is high.
   (confidence 100%,lift 1.18, conviction 9.75)
2) Inverse difference moment normalized is high and the image is normal $\rightarrow$ Inverse difference normalized is high .
   (confidence 100%,lift 1.18, conviction 7.35)
3) Entropy is high and the image is normal $\rightarrow$ Information measure of correlation is high .
   (confidence 100%,lift 1.33, conviction 10.5)
4) Inverse difference normalized is high and Information measure of correlation is high and the image is normal $\rightarrow$ Homogeneity is high .
   (confidence 100%,lift 1.18, conviction 7.95)
5) Inverse difference normalized is high and Correlation is high and the image is normal $\rightarrow$ Homogeneity is high .
   (confidence 100%,lift 1.18, conviction 6.9)
6) Homogeneity is high and Entropy is high and the image is normal $\rightarrow$ Information measure of correlation is high .
   (confidence 100%,lift 1.33, conviction 9.25)
7) Inverse difference normalized is high and Homogeneity is high and Correlation is high and Inverse difference moment normalized is high and Entropy is high and the image is normal $\rightarrow$ Information measure of correlation is high .
   (confidence 100%,lift 1.33, conviction 7)

Rule number 7 shows strong associations among the values of some textural features and normal images. The rules from number 1 to number 6 are subsets of rule number 7. Hence, Rule number 7 covers the above 6 rules because of the apriori property. The apriori property mentions that any subset of frequent itemset is also frequent. The dicovered associations help in the diagnosis of unclassifed image as long as its textural features are extracted.

IV. Conclusion

We introduced a method to find valuable relationship among textural features from retina images. FP-growth algorithm is utilized to find such associations. Results demonstrate that image mining is achievable and gives helpful relationship among specific estimations of textural features and the class of the image. Hence the proposed technique improves and conveys more certainty to the finding procedure of retina images. Promote this calculation can be effortlessly connected on other therapeutic image informational collection, for example, MRI pictures. Future work incorporates evacuating redundant association rules.

References

[1] Agrawal R et al., Mining association rules between sets of items in large databases, in proceedings of the ACM SIGMOD ICMD, Washington DC,1993,pp 207-216.

[2] Beyer,K. et al. When is nearest neighbor meaningful? In Proc. Int. Conf. Database Theo. (ICDT), 1999 pp 217-235.

[3] Felipe.J. C. et al., Retieval by content of medical images using texture for tissue identification In Proc. 16th IEEE Symp. Computer -Based Med Systems CBMS 2003, New York, 2003 pp 175-180.

[4] Haralick. R. M. et al., Textural features for image classification, IEEE Trans Syst. Man. Cybern., Vol. SMC-3.ppc610-621, 1973.

[5] Deshmukh, Jyoti and Bhosle, Udhav. (2016). "Image Mining Using Association Rule for Medical Image Dataset". Procedia Computer Science. 85. 117-124. 10.1016/j.procs.2016.05.196.

[6] Carson, Chad, Serge Belongie, Hayit Greenspan, and Jitendra Malik. "Region-Based Image Querying", in Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on, pp. 42-49. IEEE, 1997.

[7] Ji Zhang, Wynne Hsu, Mong Li Lee, Image Mining: Trends and Developments, in Proceedings of Journal of Intelligent Information Systems, 19:1, pp. 7-23, 2002.

[8] Sahu Monika, Shrivastava Madhup, Image Mining: A New Approach for data mining based on texture, in Proceedings of IEEE International Conference on Computer and Communication Technology,2012

[9] Marcela X. Ribeiro, Agna J.M. Traina, Caetano Traina,Jr., and Paulo M. Azevedo-Marques, An Association Rule-Based method to support medical images diagnosis with efficiency., IEEE Transactions on Multimedia, Vol.10,No. 2, pp. 277-285, February 2008.

[10] Maria-Luiza Antonie, Osmar R. Zaiane, Alexandru Coman, Application of Data Mining Techniques for Medical Image Classification, in Proceedings of Second International Workshop on Multimedia Data Mining (MDM/KDD2001) in conjunction with ACM SIGKDD conference, SanFrancisco, USA, Aug 26, 2001; pp. 94-101 .

[11] Maria-Luiza Antonie, Osmar R. Zaiane, Alexandru Coman, Associative Classifier for Medical Images, LNICS, Vol. 2797, MMCD, Berlin/Heidelberg: Springer; 2003, pp. 68-83.

[12] Maria-Luiza Antonie, Osmar R. Zaiane, Alexandru Coman, Mammography classification by an association rule-based classifier, in Proceedings of Third International Workshop on Multimedia Data Mining, 2002, pp. 62-69.

[13] Jiang Yun et al., "Joining associative classifier for medical images," Hybrid Intelligent Systems, 2005. HIS '05.

[14] Sumeet Dua, Harpreet Sigh and H. W. Thompson, Associative Classification of Mammograms using weighted rules, Expert System Application: An International Journal, Volume 36, Issue 5, 2009 July, pp. 9250-9259

[15] Jawad Nagi et al., Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms, in Proceedings of IEEE EMBS Conference on Biomedical Engineering and Science(IECBES 2010), KulaLumpur, Malaysia, pp87-92, December 2010.

[16] Kira, Kenji and Rendell, Larry (1992). The Feature Selection Problem: Traditional Methods and a New Algorithm. AAAI-92 Proceedings.

[17] Kira, Kenji and Rendell, Larry (1992) A Practical Approach to Feature Selection, Proceedings of the Ninth International Workshop on Machine Learning, p249-256

[18] Jiawei Han, M. Kamber, J. Pei, Data Mining Concepts and Techniques, Morgan Kaufmann publishers, 2012.