

QUANTUM CLUSTERING-BASED FEATURE SUBSET SELECTION FOR MAMMOGRAPHIC IMAGE CLASSIFICATION

N. Hamdi^{*}, K. Auhmani^{**}, M.M. Hassani^{*}

^{*}Department of Physics, Faculty of sciences Semlalia, Cadi Ayyad University,
Marrakech, Morocco

^{**}Department of Industrial Engineering, National school of applied sciences, Cadi Ayyad
University, Safi, Morocco

ABSTRACT

In this paper, we present an algorithm for feature selection. This algorithm labeled QC-FS: Quantum Clustering for Feature Selection performs the selection in two steps. Partitioning the original features space in order to group similar features is performed using the Quantum Clustering algorithm. Then the selection of a representative for each cluster is carried out. It uses similarity measures such as correlation coefficient (CC) and the mutual information (MI). The feature which maximizes this information is chosen by the algorithm.

This study is carried out for mammographic image classification. It is performed in three stages: extraction of features characterizing the tissue areas then a feature selection was achieved by the proposed algorithm and finally the classification phase was carried out. We have used the KNN classifier to perform the classification task. We have presented classification accuracy versus feature type. Results show that Zernike moments allowed an accuracy of 99.5% with preprocessed images.

KEYWORDS

Feature selection, Classification, Feature extraction, Mammographic image, Quantum Clustering, Correlation coefficient, Mutual information.

1. INTRODUCTION

In recent years data to process, store or to transmit became large. The data size can generally be presented in two dimensions: the number of variables (attributes or descriptors) and the number of samples. These two dimensions can take very high values, which can lead to distortions in the exploration and processing of such data. For this reason, it is essential to implement data processing tools to extract knowledge included in this data without loss of information. The feature selection is one of the most used approaches to reach this objective.

Feature selection allows extracting optimal subset of relevant descriptors according to criteria previously set. This selection makes it possible to reduce the size of the initial space while preserving all of the representative data.

This reduction is twofold, the first is to reduce redundancy and the second facilitates post-treatment (extraction, classification and pattern recognition ...) and therefore the data interpretation.

Two approaches are used to classify unsupervised and non-parametric dimensionality reduction techniques: those based on linear or non-linear projection of the large-scale data on well-chosen space dimension, those based on selection of spectral bands operating criteria of information complementarity and redundancy (entropy, contrast, correlation, mutual information ...) which aim to select the most representative spectral bands in the measurement space [1].

For the approach based on linear projection, we can cite the following methods: principal component analysis PCA [2], locality Preserving Projections (LPP) [3,4], the Multidimensional Scaling method (MDS) [5].

In the non-linear approach we take into account that features are not always comparable. The most used methods in this case are: The isometric mapping method (Isomap) introduced by Tenenbaum [6], the Locally Linear Coordination LLC [7], and Locally Linear Embedding (LLE) [8].

Spectral bands selection approach is based on the band selection of an optimal subset of relevant bands from a set of bands. The goal of the band selection is to reduce the dimensionality of hyperspectral data; the information cost and accelerate knowledge discovery, and other tasks in areas such as pattern recognition. We present here some of this methods; Progressive Bands selection (PBS) [9], hierarchical methods 'clustering' [10], Selection by Prediction and Projection [11], method based on mutual information [12,13] and method based on the Histogram [14].

In this paper we will present a new algorithm for feature selection. The basic idea of this algorithm, which can be classified as filter, is to use similarity measures via a clustering.

2. MATERIELS AND METHODS

2.1. Quantum clustering for Feature selection (QC-FS)

The proposed unsupervised selection algorithm which is labeled QC-FS. (Quantum Clustering for Feature Selection) performs the selection in two steps:

- Partitioning the original features space in order to group similar features, is performed using the Quantum Clustering algorithm presented in [15].
- The selection of a representative for each subset. It operates similarity measures such as correlation coefficient (CC) and the mutual information (MI). The features which maximizes this information is chosen by the algorithm.

We used filter approach and specifically the correlation coefficient and mutual information methods as they provide information on the relevance of a variable relative to the others. They use the general characteristics of the training set for classifying variables by assigning a weight to each of them comprised between -1 and 1. Thus, the d selected variables are to be considered in order to ultimately construct a $d*n$ matrix, such that n is the number of used images.

In order to measure the dependence degree between components, we use two standard measures of statistical dependence: the correlation coefficient and mutual information. The correlation coefficient is easy to calculate and is a measure of second order of statistical dependencies between the components. The mutual information measures the overall statistical dependencies between the components. It is therefore more difficult to calculate [16].

2.2 Bravais-Pearson Correlation coefficient (CC)

The correlation coefficient of Bravais-Pearson is a statistical index that expresses the intensity and direction (positive or negative) of the linear relationship between two quantitative variables. It is a measure of the linear link, i.e the ability to predict a variable x by another variable y using a linear model.

It allows to measure the relationship intensity between two quantitative variables. It is therefore an important parameter in the linear regressions (single or multiple) analysis. However, this coefficient is zero ($r = 0$) when there is no linear relationship between the variables (which does not exclude the existence of a non linear relationship). Moreover, the coefficient has a positive sign if the relationship is positive (direct, increasing) and negative sign if the relationship is negative (inverse, decreasing).

The Bravais-Pearson coefficient r between two variables x and y is calculated by using the covariance and the standard deviation as follows:

$$r_{xy} = \frac{cov_{xy}}{s_x \times s_y}$$

With:

$$cov_{xy} = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{n}$$

2.3 Mutual Information (MI):

The mutual information of a pair of variables (X, Y) represents the dependence degree in the probabilistic sense. It measures the amount of information brought by a random variable on another. It is a reduction of uncertainty about a random variable due to the knowledge of another. Let (X, Y) a pair of random variables with probability density given by $P(x, y)$. Note the marginal distributions $P(x)$ and $P(y)$. Then the mutual information is in the discrete case:

$$I(X, Y) = \sum_{x,y} P(x,y) \log \left(\frac{P(x,y)}{P(x)P(y)} \right)$$

The mutual information and the correlation coefficient procedures are not based on a particular model. The construction of a K-nearest neighbours (KNN) and decision tree model is feasible on the d selected variables. The obtained results for the QC-CC and QC-IM models will be presented.

2.4 Feature Extraction

Features were calculated on a set of two classes labelled images (normal and abnormal) [17]. These images are firstly pre-processed and transformed into the frequency domain by three types of wavelet transforms. In order to exploit the advantage of each feature, we merge them into a single feature vector. The extracted features are shown in Table 1.

Table I. Description of extracted features

Feature type	Number
Zernike's Moment	49
Radon's Transform	12
Multi-scale Histogram	24
Tamura's features	6
The 4 Moments	48

2.5 Dataset

The proposed algorithm was tested on the original mammographic images of MIAS database [18]. These images of original size 1024*1024 are accompanied by the opinion of the radiologists

that specify the class of the image, the microcalcifications type, the center of clusters and other informations. We decomposed the images into regions of interest (ROI) of size $256 * 256$ and applied our benchmark on the ROI. We have used 107 mammographic images; 56 samples are tumoral and 51 samples are normal.

2.6 Scheme of realized system

Our scheme (Figure 1) is divided in four parts. First, the original mammographic image will undergo pre-processing to reduce noise, enhance the features and enhance the presentation. The next part deals with the extraction of features as described above. In the third part, the proposed feature selection method will be applied to the feature vector in order to select a sub set of pertinent features. The low-dimensional sub set will then be used to classify images by th K nearest neighbours classifier. To evaluate and compare the performance of the proposed algorithm, we present the classification accuracy versus feature type.

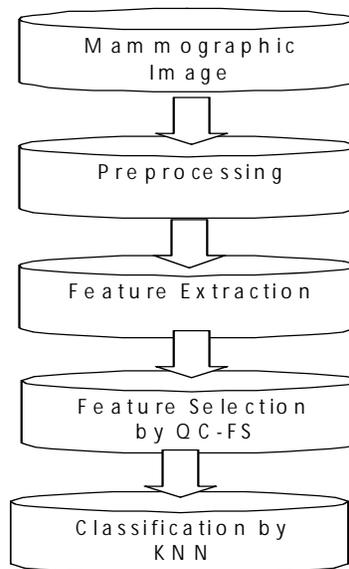


Figure 1 Bloc Scheme of the realized automatic system

3. RESULTS AND DISCUSSIONS

We have proposed a pre-processing procedure of mammographic image where the evaluation is based on visual quality of the obtained image [19, 20]. We have then applied our proposed system on the images firstly without pre-processing and then after pre-processing.

In order to measure the performance of the QC-FS algorithm we have used the KNN classifier with $K=1$. Table 2 dresses different measures of classification accuracy (%) after feature selection operation by QC-FS algorithm for different type of features.

Table 1 Classification accuracy versus feature type :

	Correlation Coefficient		Mutual information	
	With Pre-processing	Without Pre-processing	With Pre-processing	Without Pre-processing
Radon Transform	91.1%	82.6%	92%	79.4%
Zernike Moments	99.5%	71.9%	71.9%	71.4%
The first 4 Moments	72.8%	71%	72.8%	72.8%
Multi scale histogram	94.8%	92%	94.8%	92.2%

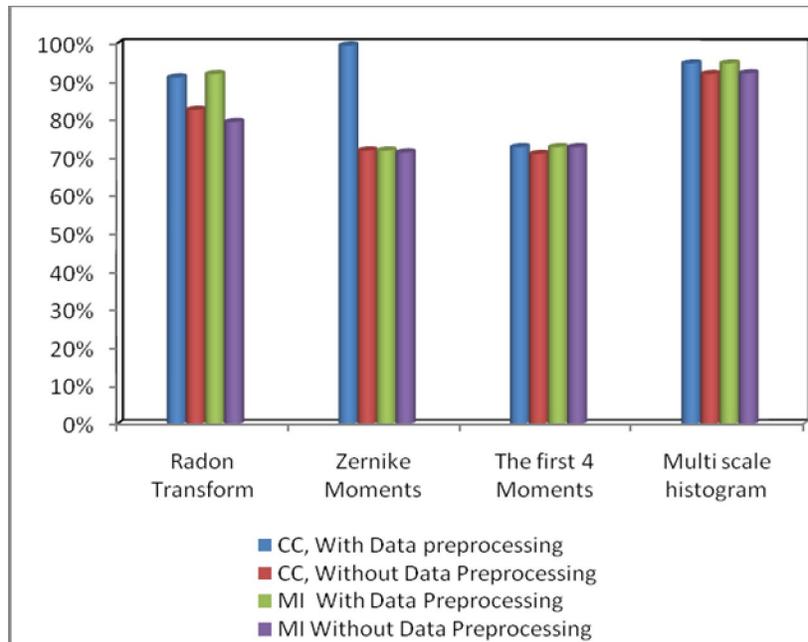


Figure 2: Classification accuracy versus feature type, with and without Data pre-processing.

Table 2 and figure 2 show classification accuracy of mammographic images according to the type of descriptors and depending also on the similarity measure used in the feature selection process. The results are also presented in two cases: Descriptors are extracted before or after pre-processing the image. From this table we see that in all cases:

The contribution of pre-processing is very important and its effect is particularly noticeable in figure 2. The classification performance without pre-processing is significantly lower than those obtained with the pre-processing; we observe more than 20% improvement in some cases (for Zernike moment for example).

Zernike Moments allow the best accuracies for the two similarity measures.

Data pre-processing generally accelerates the learning step in the classifiers and improves the classification rate because the data is less noise and better conditioned. However, for the descriptors from the first four static moments, improving the average rates do not exceed 0.1%.

4. CONCLUSION

In this paper we have presented an algorithm to select relevant features for mammographic image classification. The algorithm performs the selection task in two steps. The first step deals with the clustering of data. This clustering is carried out by the quantum clustering algorithm. The second step deals with the selection of a representative feature of each cluster. This selection is based on mutual information or on correlation coefficient.

Classification accuracy of mammographic image was computed without data pre-processing and then with data pre-processing. Accuracies were also calculated versus the feature type and the used similarity measure.

Presented results showed that the best accuracies were obtained if data were preprocessed and the similarity measure is the correlation coefficient with Zernike moments features. An accuracy of 99% was reached.

REFERENCES

- [1] Jihan Fawaz Khoder, "Nouvel Algorithme pour la Réduction de la Dimensionnalité en Imagerie Hyperspectrale" , L'Université de Versailles S t Quentin-en-Yvelines Et L'Université Libanaise, Soutenue publiquement le 24 Octobre 2013.
- [2] L.Van Der Maaten, E.Postma, J. Van Den Herik. Dimensionality Reduction: A Comparative Review. TiCC, Tilburg University.
- [3] X.He, P.Niyogi, Locality preserving projections, in: Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2004, pp. 37–44.
- [4] S.Chen, H.Zhao, M.Kong, & B.Luo. 2D-LPP: a two-dimensional extension of locality preserving projections. *Neurocomputing*, 70(4), 912-921. 2007.
- [5] D.Demers and G.Cottrell. Non-linear dimensionality reduction. *Advances in Neural Information Processing Systems*, Vol. 5, pp 580–587, San Mateo, CA, USA, 1993.
- [6] J.B.Tenenbaum, V.de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): pp 2319–2323, 2000.
- [7] Y.W.Teh and S.T. Roweis. Automatic alignment of hidden representations. *Advances in Neural Information Processing Systems*, 2002.
- [8] S.T.Roweis and L.K. Saul. Nonlinear dimensionality reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.
- [9] K.Fisher and Chein-I Chang. Progressive band selection for satellite hyperspectral data compression and transmission. *Journal of Applied Remote Sensing*, Vol. 4, Issue 1, Special Section on Satellite Data Compression.
- [10] S.Lee and Melba M. Crawford. Hierarchical Clustering Approach for Unsupervised Image Classification of Hyperspectral Data. *IEEE International Symposium of Geoscience and Remote Sensing, IGARSS '04*, vol.2, pp 941- 944, 2004.
- [11] Q.Du and H.Yang. Similarity-based unsupervised band selection for hyperspectral image analysis. *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
- [12] C.Cariou, K.Chehdi, and S. Le Moan. BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing. *Geoscience and Remote Sensing Letters, IEEE*. Vol. 8, Issue 3, pp 565 – 569, May 2011.
- [13] Martínez-Usó, F.Pla, J. M. Sotoca, and P. García-Sevilla. Clusteringbased hyperspectral band selection using information measures. *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.

- [14] Kermad, K. Chehdi. Multi-Bands Image Segmentation: A Scalar Approach. International Conference on Image Processing, Vol.3, pp 468 -471, 2000.
- [15] H. David ; GOTTLIEB Assaf, Algorithm for data clustering in pattern recognition problems based on quantum mechanics, Physical review letters , 2002, vol. 88, no1, pp. 018702.1-018702.4
- [16] Michel NAROZNY, “ L'Analyse en Composantes Indépendantes (ACI) et la compression de données”, These au sein de l'Équipe Systèmes de Traitement des Signaux, Supélec, campus de Metz, Decembre 2005
- [17] N. Hamdi, K. Auhmani¹², M.M.Hassani, A Comparative study of Dimension Reduction Methods combined with wavelet transform applied to the classification of mammographic images,
- [18] <http://peipa.essex.ac.uk/info/mias.html>.
- [19] K.Auhmani, N.Hamdi, M.M.Hassani; Detection and visualization of microcalcifications in digital mammograms using mathematical morphology and wavelet algorithm In proceedings of the International Conference on Smart Materials and Adaptive Structures: Mathematical Modeling and Computation; pp 233-235 Tangier Morocco; 14-16 April 2008.
- [20] N.Hamdi, K.Auhmani, M.M.Hassani, Computer Aided Diagnosis: Automated detection and enhancement of microcalcifications in digitized mammograms using wavelet decomposition and local gray thresholding, PCN Journal, 47, 75-78, May 2009.