pp.198:203

Effective Feature Selection for Pre-Cancerous Cervix Lesions Using Artificial Neural Networks

Farnaz Rouhbakhsh[1], Fardad Farokhi[2], Kaveh Kangarloo[3]

[1] Electrical Engineering Department, Islamic Azad University Central Tehran Branch, Tehran, Iran, Email: farnazruhbakhsh84@gmail.com
[2] Electrical Engineering Department, Islamic Azad University Central Tehran Branch, Tehran, Iran, Email: f_farokhi@iauctb.ac.ir
[3] Electrical Engineering Department, Islamic Azad University Central Tehran Branch, Tehran, Iran, Email: k_kangarloo@iauctb.ac.ir

Abstract

Since most common form of cervical cancer starts with pre-cancerous changes, a flawless detection of these changes becomes an important issue to prevent and treat the cervix cancer. There are 2 ways to stop this disease from developing. One way is to find and treat pre-cancers before they become true cancers, and the other is to prevent the pre-cancers in the first place. The presented approach uses precancerous images which are taken from a digital colposcope, and a set of texture and color features is extracted which includes low and high grade SIL (Squamous Interepithelial Lesion) .After extracting, features are fed to a classifier, which could be KNN,RBF,MLP and Neuro-Fuzzy network and after training effective features are selected using UTA algorithm for each classifier individually. Finally, results come in a comparison table, show that the landa fourteenth, theta-x and together with Neuro-fuzzy classifier have the best overall performance. This approach has an acceptable and simple early diagnosis of cervix cancer and may have found clinical application.

Keywords: Image classification; artificial neural network; Feature selection; Colposcopic images;

1. Introduction

1.1. Cervical precancerous cells

Low-grade squamous intraepithelial lesions (LGSIL) refer to early changes in the size, shape and number of cells that form the surface of the cervix. Some low-grade lesions disappear; however others may grow larger or become more abnormal, forming a high-grade lesion. Precancerous low-grade lesions also may be called mild dysplasia or cervical intraepithelial neoplasia. High-grade squamous intraepithelial lesions (HGSIL) means that the precancerous changes are more severe; like low-grade SIL, these precancerous changes involve only cells on the surface of the cervix. If abnormal cells spread deeper into the cervix or to the other tissues or organs, the disease is then called cervical cancer.

Cancer of the cervix is the second most common cancer in women worldwide, with about 500000 new cases and 250000 deaths each year (world health organization reports). The Papanicolaou (also called Pap smear, Pap test, cervical smear, or smear test) and colposcopy test is a conventional screening test used to detect potentially pre-cancerous and cancerous processes in the endocervical canal (transformation zone) of the female reproductive system. Changes can be treated, in order to prevent cervical cancer.

In March 2012, the U.S. Preventative Services Task Force updated its recommendations concerning cervical cancer screening guidelines. Cervical cancer screening is essential for the early detection and prevention of cervical cancer. The guidelines apply to women who are at an average risk of developing cervical cancer but in many countries due to the small number of skilled cytologists, screening is a time consuming procedure which includes human errors in diagnosis of pre-cancerous cells in uterine cervix. It means this kind of cancer like others, can be prevented earlier if it is applied a reliable and not time-consuming automated system.

1.2. Recent works

In recent years, early detection of cervix cancer has been widely discussed in relation to image processing science, and numerous algorithms have been proposed for cervix precancerous lesions classification. In [1] authors worked on connectionist methods such as multilayered perceptions, radial basis function (RBF) networks, and ensembles of such networks. They demonstrated that cervical tissue fluorescence spectra can be used to develop detection algorithms in order to differentiate SILs from normal tissue samples. Of the various algorithms explored, the RBF network ensemble proved to be the best alternative, surpassing single networks, and the multivariate statistical algorithm which proposed for only AW (Acetowhite) region of cervix. In [2] colposcopic image classification based on contour parameters using MLP artificial neural network with accuracy of 95.8% was presented. This method uses only texture feature which was fed to a 6 hidden layers network.

In [3] an automated segmentation and classification of major lesions observed in early stages of cervical cancer based on morphological texture feature was presented. This method considers only AW region of cervix for abnormality detection too.

In [4] they applied a cost-sensitive SVM (support vector machine) classification scheme to the cervical cancer images in order to separate diseased regions from healthy tissue. But it has poor TP and TN and it was time consuming. In all previously presented methods, the author deals with specular reflection or SR [5] which should be removed prior to classification as it is dramatically reduces the classification accuracy.

This paper attempts to present a simple and reliable approach by applying a set of colour and texture features as a hybrid vector to increase the classification accuracy, independent of the tissue abnormalities, such as mosaicism ,punctuations and furthermore without any need to SR removal [5],[6] as a time consuming process.

2. Method

In this algorithm, first colposcopic images including considered region of interest (ROI) are cropped into size 300*400 pixel and equalized for brightness and contrast and then features including correspondence analysis (CA) [7], color gradient[8], quasi invariants [8], exceed red [10], energy and skewness are extracted and after constructing input matrix, it is fed to the classifiers. Finally, after training of neural networks, by UTA algorithm [11], effective and ineffective features are determined respectively. Fig.1 shows the flowchart of the algorithm.

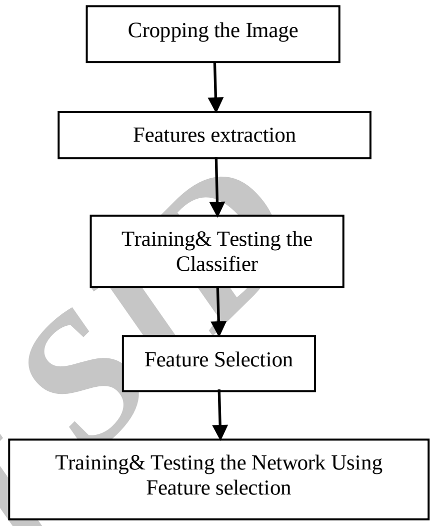

Fig.1. presented method.

2.1. Texture and color features extraction

Selecting appropriate features is the one of the most important part of the image classification. In this study, 27 features were used to determine abnormality in cervix images. Any features which exists evidences in previous papers are selected for evaluation and some other features which are powerful for the classification were evaluated as well. Features are as follows:

2.1.1. Correspondence Analysis

Correspondence Analysis (CA) [7] is a multivariate statistical tool, like principal components analysis, which enables to proceed on a large dataset made of p individuals described by n variables. This dataset can be represented by a matrix X with p rows and n columns. The extraction is done by defining a q-dimensional sub-space in which the *q* axes are representative of the significant information contained in the original dataset although in this approach the whole n axis representatives were used. This algorithm samples the original image into blocks of 4 × 4 pixels, using a sliding technique. It then performs a correspondence analysis (CA) on these blocks and finally reconstructs the image one block after another using only the factors which significantly contribute to the signal variance of each block. The flow chart is shown in Fig. 2.As in this analysis, it has been proven that the original matrix could be reconstructed using those *n* eigenvalues, the obtained 16 eigenvalues were selected as features [7].

2.1.2. Colour gradient

These features are the gradient magnitude of R, G and B colours in x and y direction [8].

2.1.3. Spherical derivatives

Spherical derivative used to compute the Shadow-Shading Quasi-Invariants (SSQI) [8], by calculating these parameters:

Theta-x: spherical derivatives in the shadow-shading invariant x direction
Theta-y: spherical derivatives in the shadow-shading invariant y direction
Phi-x: spherical derivatives in the shadow-shading invariant x direction
Phi-y: spherical derivatives in the shadow-shading invariant y direction
r_x: spherical derivatives in the shadow-shading variant x direction
r_y: spherical derivatives in the shadow-shading variant y direction

Note that the Jacobian is taken into account, so theta_x = r×sin (phi)×THETA_x and phi_x = r×PHI_x, where THETA_x and PHI_x are the true spatial derivatives of THETA and PHI.

2.1.4. Skewness

Skewness is a measure of the asymmetry of the data around the mean sample. If skewness is negative, the data are spread out more to the left of the mean than to the right and vice versa.
The skewness of a distribution is defined as (1):

$$\text{skewness} = \frac{E(x-\mu)^3}{\sigma^3} \quad (1)$$

Where E is expectation operator,μ is the mean and σ is the standarddeviation

2.1.5. Energy

Energy measures the uniformity of the entries in the joint distribution. It is lowest when all entries are equal. It is formulated as equation (2).

$$E = \{p(i,j)\}^2 \quad (2)$$

Where P, is input image, i and j is row and column of the image respectively.

2.1.6. Local standard deviation

It returns an array, where each pixel of it contains the standard deviation of the 3-by-3 neighbourhood pixels around the corresponding pixel of the input image. We considered this features on the red channel.

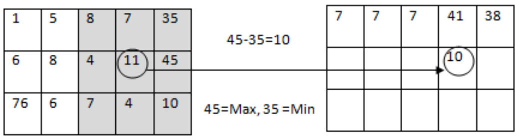

Fig.2. local standard deviation

2.1.7. Red excess-mean

This feature measures the excess of the red colorand calculated as equation (3).

$$Ex=2*R-(G+B) \quad (3)$$

2.2. Feature Vector

In this study , feature vector were constructed using 27 features including 16 features of CA, 6 features from qausi-invariant, one feature for colour gradient, skewness, energy, local standard deviation and red excess-mean respectively. The feature vector was fed to 4 different types of the classifiers in order to classify the input instances to the two classes of low grade and high grade SIL.

2.3. Classifiers

In this study, we used four classifiers as presented in A to D section:

2.3.1. K-Nearest Neighbour algorithm

The k-nearest neighbour algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. The k-nearest neighbour algorithm is the simplest type of machine learning algorithms. An object is classified by a majority vote of its neighbours, with the object being assigned to the most common class amongst its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbour.

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists of only storing the feature vectors and class labels of the training samples.

In the classification phase, k is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

Usually Euclidean distance is used as the distance metric.

2.3.2. RBF

In the context of a neural network, the hidden units provide a set of “functions” that constitute an arbitrary “basis” for the input patterns (vectors) when they are expanded into the hidden space; these functions are called radial basis function. Radial Basis functions were first introduced in the solution of the real multivariate interpolation problem.

The construction of a radial-basis functions (RBF) network, in its most basic form, involves three layers with entirely different roles as it is shown in Fig.3 [12]. The input layer is made up of source node (sensory units) that connects the network to its environment. The second layer, the only hidden layer in the network, applies a nonlinear transformation from the input space; in most applications the hidden space is of high dimensionality. The output layer is linear, supplying the response of the network to the activation pattern (signal) applied to the input layer. Radial basis functions are typically used to build up function approximations of the form:

$$y(x) = \sum_{i=1}^{N} w_i \emptyset(\|x - x_i\|) \quad (4)$$

Where the approximating function y(x) is represented as a sum of N radial basis functions, each associated with a different center x_i, and weighted by an appropriate coefficient w_i. The weights w_i can be estimated using the matrix methods of squares, because the approximating function is linear in the weight sand$\emptyset$ is Gaussian function called green function.

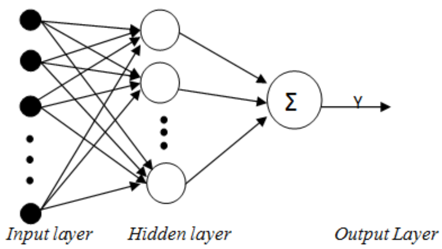

Fig.3. Radial Basis Function Network

2.3.3. Multi-Layer Perceptron

A multilayer Perceptron (MLP) is a feed forward artificial neural network model that maps input sets of data onto a set of the appropriate output. An MLP consists of multiple layers of neurons which are fully connected to the next layer neurons as it is shown in Fig.4.Each node in the first layer connects with a proper weight to the other nodes respectively; each neuron (or processing element) has a nonlinear activation function. MLP utilizes a supervised learning technique called “back propagation” [13] for training of the network. MLP is a modification of the standard linear perceptron and can distinguish data that is not linearly separable.

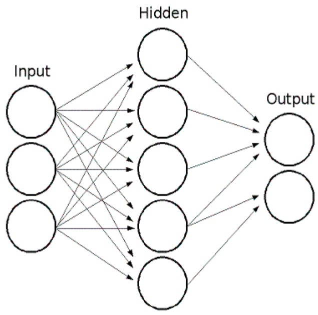

Fig.4. MLP neural network with one hidden layer

2.3.4. Neuro-Fuzzy system

A Fuzzy system has been designed into two different approaches. In the first approach, fuzzy IF-Then [9] rules are first generated and then fuzzy system is constructed according to certain choices of fuzzy inference engine, fuzzifier, and defuzzifier. In the second approach that was used in this study, the structure of fuzzy system is specified first and some parameters in the structure are free to change. The parameters are so determined according to the input-output pairs of the training data. The selected fuzzy system is formulated as (5) [10].

$$f(x) = \frac{\sum_{l=1}^{M} y^{-l}[\prod_{i=1}^{n} \exp\left(-(\frac{xi - xi^{-l}}{\sigma i^{l}})^2\right)]}{\sum_{l=1}^{M}[\prod_{i=1}^{n} \exp\left(-(\frac{xi - xi^{-l}}{\sigma i^{l}})^2\right)]} \quad (5)$$

Where M is the number of the rules, and y^{-l} ,xi^{-1} and σi^{l} are free parameter. After specification of these parameters fuzzy system is designed completely. Generally input x is passed through a product Gaussian operator to obtain z^l:

$$Z^l = \prod_{i=1}^{n} \exp\left(-(\frac{xi - xi^{-l}}{\sigma i^{l}})^2\right) \quad (6)$$

Then z^l are passed through a summation operator and a weighted summation operator to obtain a and b as follow:

$$a = \sum_{l=1}^{M} y^{-l} z^{l} \quad (7)$$

$$b = \sum_{l=1}^{M} z^{l} \quad (8)$$

Finally, the output of the fuzzy system is computed as *f(x)=a/b* as shown in Fig.5.

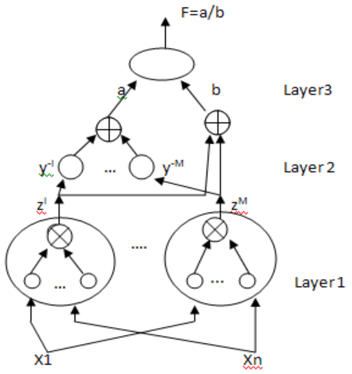

Fig.5. Neuro-Fuzzy network

2.4. Feature selection

Feature selection is the technique of selecting a subset of relevant features from a large set of applied features for building robust learning models. It means that by removing most irrelevant and redundant features from the data, the performance of learning models is improved. There are several methods to select an optimized subset of features. One of them is the UTA method proposed by Utans [11] which is based on trained artificial neural networks. In UTA algorithm, after training the network, average of a feature in all patterns in the input matrix is computed. Then the considered feature has been replaced in all instances in rows with the mean value.

The theory is based on the fact that if there is a feature which is constant in value in all instances then it could not make any difference between instances in classification process and it is not a consistent feature. After replacing, the network with new matrix has been tested and finally a comparison error is defined as (9) for each feature individually.

$$E=(FP+FN)old - (FP+FN)new \qquad (9)$$

Where $(FP+FN)_{old}$ is the sum of false positive and false negative of trained network using original matrix and $(FP+FN)_{new}$ is the sum of false positive and false negative of trained network using the matrix with replaced feature with the mean value.
There are 3 conditional states for E:
1. If E < 0, then related feature is relevant, so higher value means higher importance of that feature.
2. If E=0, then related feature is ineffective, so removing this feature has not any changes in networks performance.
3. If E>0, then related feature not only is irrelevant, but also is destructive and should be removed.

3. Results

We have investigated 96 bmp images 300*400 pixel; contain 48 LSIL and 48 HSIL conditions. 27 features are extracted from each image and after constructing the input matrix with 29 columns and 96 rows, the input vectors was fed to 4 different classifiers respectively. The obtained results are presented in Table..1.

In KNN classifier the number of neighborhood was K=6 and the classification accuracy was 70.83%. It had no changes after applying UTA algorithm due to E<0 for all features and so it uses all features in classification. After KNN, we used RBF networks with 10 neurons including 10 green functions. The result of training is 72.92 % accuracy and like KNN it had no changes in performance by applying UTA algorithm. Then we apply MLP network with 2 hidden layers with 30 neurons in each of them and learning rate of 0.01. It had 93.75% efficiency and by applying UTA it reached 97.92 % with 2 minutes process time. UTA determined that this network uses 15 features. Finally we test our matrix on Neuro-Fuzzy network and the accuracy result of 100% was achieved, and by applying UTA algorithm, we Fig.d out that this network strongly uses all features with large E values. Fourteen landa coefficients has most effect in network efficiency. So we tried to apply landa fourteenth separately and we get result of 95.83% accuracy, then we add theta-x to and the accuracy increased to 97.92%. It seemed that one more feature is needed to reach a better accuracy, thus we tried all features one by one in addition with the landa fourteenth and theta-x. Finally local standard deviation leads us to the accuracy of 100 %.

Table.1
Comparison of the system performance after applying UTA

Classifier	Accuracy	Specificity	Sensitivity
K-NN	70.83%	70.83%	70.83%
RBF	72.92%	83.33%	62.50%
MLP	97.92%	95.83%	100%
Neuro- Fuzzy	100%	100%	100%

4. Conclusion

In this paper, a new set of powerful hybrid features was presented that improved the efficiency of precancerous cervix lesions classification without any need to remove specular reflection process.

First 27 features were extracted from 96 digital cervix images, and then they were applied to the different types of classifiers. By using UTA algorithm, neuro-fuzzy network with just 3 features among 27 features at the least elapsed time, approximately 30 sec lead us to a fast and accurate classification of pre-cancerous cervix lesions. Table.2 shows that by applying UTA algorithm we can Fig. out that landa fourteenth feature in neuro-fuzzy system with 100% accuracy has most effective role in classification result according to UTA algorithm result.

Also we can conclude that however the texture feature had high effectiveness in classification, we cannot ignore the influence of color features.

Our future work will be focused on classification of 4 types of cervix texture, including 2 type of precancerous, normal and cancerous cervix.

Table.2
Effective feature selection by UTA

Effective feature	**Number of used feature**	**classifier**
Energy	27	**K-NN**
Mean-red	2	**RBF**
$Landa_{13}$,theta-x	15	**MLP**
$Landa_{14}$	3	**Fuzzy-system**

References

[1] K. Tumer, N Ramanujam, J. Ghosh, R. Kortum, “Ensembles of Radial Basis Function Networks for Spectroscopic Detection of Cervical Precancer”, IEEE Transactions on Biomedical Engineering, Vol.45, No.8, 1998.

[2] I. Claude, R. Winzenrieth, P. Pouletaut, J. Charles Boulanger, ICPR '02 Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1, Page 10771.

[3] B. Tulpule, Sh. Yang, Y. Srinivasan, S. Mitra, B. Nutter, “Segmentation and Classification of Cervix Lesions by Pattern and Texture Analysis”, The 14th IEEE International Conference Fuzzy Systems FUZZ '05, 2005.

[4] Y. Artan, X. Huang, ”Combining Multiple 2ν-SVM Classifiers for Tissue Segmentation”, 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp.488-491, 2008.

[5] Y. Srinivasan, E. Corona, B. Nutter, S. Mitra, S.Bhattacharya, “A Unified Model-Based Image Analysis Framework for Automated Detection of Precancerous Lesions in Digitized Uterine Cervix Images”, IEEE Journal of Selected Topics in Signal Processing, Vol.3, No.1, 2009.

[6] A. Das, A. Kar, D. Bhattacharyya, ”Elimination of specular reflection and identification of ROI: The first step in automated detection of Cervical Cancer using Digital Colposcopy”, IEEE International Conference on Imaging Systems and Techniques (IST), pp.237–241, 2011.

[7] P. Hannequin and J. Mas, “Statistical and heuristic image noise extraction (SHINE): a new method for processing Poisson Noise in Scintigraphic Images,” Phys. In Med. & Biol., Vol.47, pp.4329–4344, 2002.

[8] J.van de Weijer, T. Gevers, J. M Geusebroek, “Edge and Corner Detection by Photometric Quasi-invariants”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.4, pp.625-630, 2005.

[9] D. Nauck, R. Kruse, “A Fuzzy Neural Network Learning Fuzzy Control Rules and Membership Functions by Fuzzy Error Back propagation”, IEEE International Conference on Neural Networks, vol.2, pp.1022-1027, 1993.

[10] Li Xin Wang, “A Course in Fuzzy Systems and Control”. Prentice-Hall Edition, chapter 13, pp.168-172., 1997.

[11] M.F.Redondo, C.H.Espinosa, “A Comparison Among Feature Selection Method Based on Trained Network”, Neural Networks for Signal Processing IX Proceedings of the IEEE Signal Processing Society Workshop, pp.205-214, 1999.

[12] S. Haykin, “Neural Networks, A Comprehensive Foundation”, Second edition, Prentice-Hall Edition, chapter 5, 1999.

[13] S. Haykin, “Neural Networks, A Comprehensive Foundation”, Second edition, Prentice-Hall Edition, chapter 3, 1999