International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 9 (2015) pp. 24313-24322

http://www.ripublication.com

Performance of GRNN Classifier Used In Pap Smear Test on Cervical Cell

I. Immaculate
PG Scholar, Department of CSE (PG)
Sri Ramakrishna Engineering College, Coimbatore
Dr. A. Grace Selvarani
Professor & Head, Department of CSE (PG)
Sri Ramakrishna Engineering College, Coimbatore

Abstract

Most of females are affected by cervical cancer. The Pap smear could be a common cytological analysis that appears at cells from the cervix. It is a secure, efficient, and well established technique for the diagnoses of the cervical cancer. This paper proposes, to segment and classify single cell with help of fuzzy based c-means clustering technique into nucleus, cytoplasm and background. Classification of segmented cells will be trained by GRNN classifier method. The proposed method yields 99 % of accuracy with good performance.

Keywords: Cervical cell classification,Cervical cell screening, Cervical cell segmentation,,Pap smear.

Introduction

Image processing is the study of any algorithm that takes an image as input and returns an image as output. Image segmentation is the process of partitioning a digital image into multiple segments. Set of pixels are known as super pixel. Goal is to simplify and or change the representation of an image into something more meaningful and easier to analyze.[7]

Classification is the problem of identifying to which of a set of categories, a new observation, on the basis of training set of data containing observations (or instances) whose category membership is known. An example would be assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.).In the terminology of machine learning,classification is considered as an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as

clustering, and involves grouping data into categories based on some measure of inherent similarity ordistance.[2]

An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm, which maps input data to a category.[1]

Feature extraction starts from an initial set of measured data and builds features values intended to be informative, non redundant, facilitating the subsequent learning and generalization steps, in some cases leading to better human interpretations.

The extracted features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.[16]

Artificial intelligent techniques such as neural networks, genetic algorithms and fuzzy logic are among the most powerful tools available for detecting and describing subtle relationships in massive amounts of seemingly unrelated data. Neural networks can learn and are actually taught instead of being programmed. Teaching mode can be supervised or unsupervised.

Fuzzy c-means clustering algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more isitsmembership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated using the formula.[6]

One role of Pap smear test is directed to early detection of abnormal cancer cells. Once the abnormal cells are detected, the patient can undergo for a biopsy examination and further treatment. Consequently, the growth of the cancer can be stopped at an early stage.[7]

Related Work

In this section,studies related to segmentation and classifications of cervical cells are carried on. Cytological analysis is the analysis done beneath a magnifier of cells collected from an area of the body. This can be done to work out on how the cells appear theirtype and performance. This takes a look at typically accustomed hunt for cancers and metastatic tumor changes. It should even be accustomed hunt for infective agent infections in cells.[1]

M.-H. Tsai, Y.-K.Chan, Z.-Z Lin, S.-F.Yang-Mao, P.-C. Huang[15] developed a cytoplast and nucleus contour (CNC) detector to sever the nucleus and cytoplast from a cervical smear image. They alsoproposed the bi-group enhancer to make a clear-cut separation for the pixels laid between two objects, and the maximal color difference (MCD) method to draw the Pap test nucleus contour. The CNC detector adopts a median filter to sweep off noises, the bi-group enhancer to suppress the noises and brighten the object contours, the *K*-mean algorithm to discern the cytoplast from the background, and the MCD method to extract the nucleus contour.

A. Genctav, S. Aksoy, S. Onder,[2] proposed, automatic thresholding to separate the cell regions from the background, a multi-scale hierarchical segmentation algorithm to partition these regions based on homogeneity and circularity, and a binary classifier to finalize the separation of nuclei from cytoplasm within the cell regions.

M.E. Plissiti, C. Nikou, A. Charchanti[14], presented a fully automated method for cell nuclei detection in Pap smear images. In their method, they have examined the performance of an unsupervised (fuzzy C-means) and a supervised (support vector machines) classification technique. In both classification techniques, the effect of the refinement step improves the performance of the clustering algorithm.

Several research works have these automated systems indeed to improve the accuracy of the screening result [15].Automatic system should be useful for earlier detection and diagnosing. From theimage processing view, extracted each cell segments into different regions as nucleus, cytoplasm, and background, most of research was based on nucleus [5].

However, when one would like to classify each cervical cell into categories with only nucleus information, it might not yield a good performance. Hence, segmenting whole cell is more desirable [15]. However, there is no classification result reported in these works. After the segmentation step,each cell is then classified using specific classifiers based on the extracted features from cell components as mentioned earlier or by using filters to discriminate classes without feature extraction process [2]. However, the classification performance in Ref. [2] is not quite high.

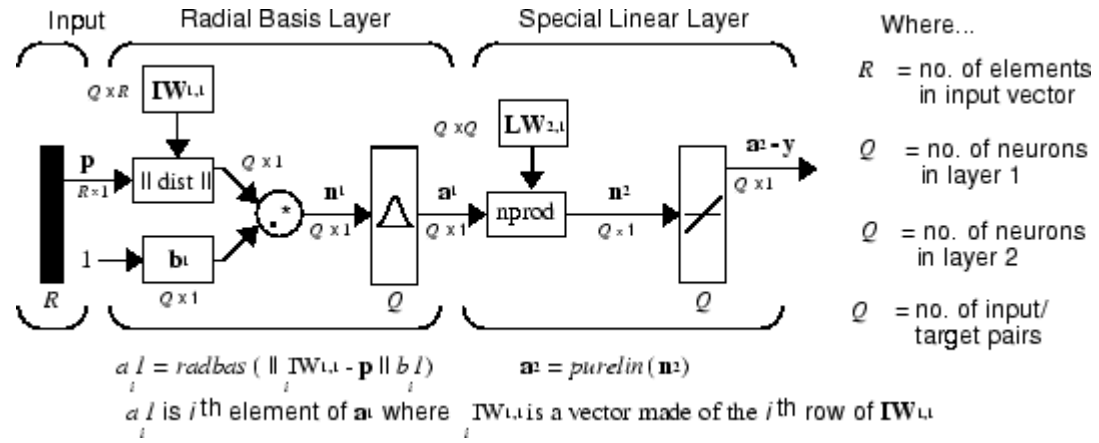
Patch based fuzzy c-means clustering method is used to segment nuclei and background from the conventional pap test[18].

ANN classifier with back propagation method is a non- linear classifier which can be used to classify complex relation between input and output. The number of hidden layer is based on the number of features used [2]. So, in the training period and testing time is high. It also provides less accurate identification.

Proposed System

In this study we proposed General Regression Neural Network (GRNN) method. The important feature of GRNN is that it uses only one hidden layer, so it reduces the testing time in training phase.A generalized regression neural network (GRNN) is often used for function approximation. It has a radial basis layer and a special linear layer. [4]

The architecture for the GRNN is shown in figure3.1. It is similar to the radial basis network, but has a slightly different second layer.



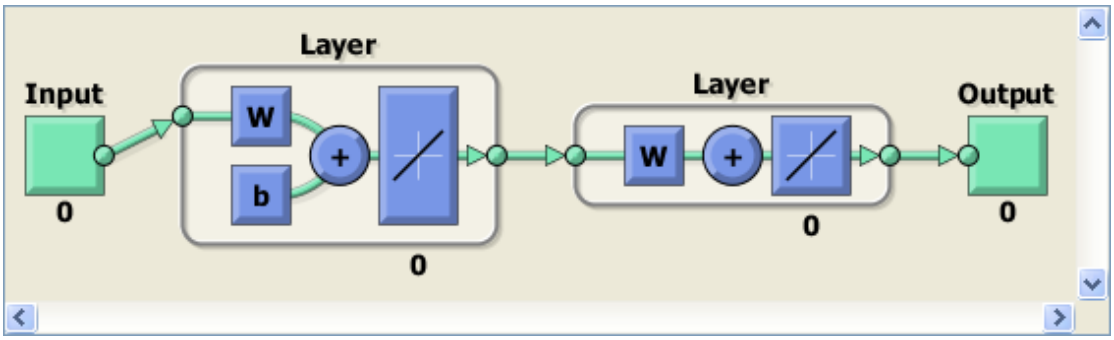


Figure 3.1: Architecture of GRNN

Segmentation Process

Patch based fuzzy c-means clustering technique is most suitable for clustering data with uncertainty. In segmentation, a single cell is converted into gray scale image, then apply median filter to reduce the noise and make it smooth. The processed image is segmented into nucleus and non-nucleus with the help of FCM clustering method. Nucleus and non- nucleus clustered into patches based on the threshold value of nucleus percentage in that all patches. Finally,the patch values of center is less than nucleus threshold value labeled as nucleus, else it labeled as non-nucleus. [7]

Quantitative Analysis of Features Extracted And Used

After segmentation by the help of FCM clustering, features will be extracted based on the nucleus. The coarseness of nucleus texture represents the distribution of chromatin. It increases according to the level of sensitivity.[11]

For automatic classification of cervical cancer, several classifier methods have been combined to separate normal and abnormal cells. This part is described in two sections: feature selection and classifiers.

Feature Selection:

Feature selection technique can be distinguished from feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are

often used in domains where there are many features and comparatively few samples. [12]

Feature selection is also useful as part of the data analysis process, as it shows which features are important for prediction, and how these features [2] arerelated. The below table 3.2.1 list the features which are used in this paper.

Table 3.2.1: Features used for selection

Feature No	Feature name	Remarks
F1	Area of nucleus	
F2	compactness of nucleus	
F3	Major axis of nucleus	
F4	Minor axis of nucleus	
F5	Aspect ratio of nucleus	
F6	Homogeneity of nucleus	Nucleusbased features
F7	Nucleus to cytoplasm ratio	
F8	Compactness of entire cell.	
F9	Area of entire cell.	Non nucleus features

F1: Area of nucleus

A_{nu}=Total number of pixels in the nucleus region (3.2.1)

F2: compactness of nucleus

C_{nu}= (perimeter of nucleus)^2/Area of nucleus (3.2.2)

F3: major axis of nucleus

L_{nu}=the length of the major axis of an ellipse which completely encloses the nuclear region.(3.2.3)

F4: **minor axis of nucleus**

D_{nu}= the length of the minor axis of an ellipse which completely encloses the nuclear region. (3.2.4)

F5:Aspect ratio of nucleus

R_{nu}= Width of the nucleus(W_{nu})/Height of the nucleus(H_{nu}) (3.2.5)

F6 : homogeneity of nucleus

$H_{nu}=\sum^{k}_{i=1} \sum^{k}_{j=1} p(i,j)/ 1+|i-j|'$ (3.2.6)

F7: Nucleus to cytoplasm (N/C) ratio

NC=area of nucleus/area of cytoplasm (3.2.7)

F8: compactness of entire cell

C_{en}=(perimeter of entire cell)^2/area of entire cell (3.2.8)

F9 :area of entire cell
A_{en}=Total number of pixels in the entire cell region. (3.2.9)

Classifiers

For classification process
Artificial neural networks (ANN): This classifier is a non-linear statistical classifier which can be used to classify the data with complex relation between input and output. The ANN consists of a number of neurons organized in layers. This research uses a 3-layer back propagation neural network, consisting of 5–9 neurons in the input layer depending on the number of features used, 1 hidden layer with 15 neurons, and 2, 4, or 7 neurons in the output layers depending on the number of classes being considered[2]

GRNN Classifiers:
General Regression Neural Network (GRNN) is one of the neural networks applied in this paper. A GRNN can be thought of as a normalized Radial Basis Function (RBF) network in which there is a hidden unit centered at every training case. These RBF units are known as kernels and are usually probability density functions (e.g., Gaussian). The hidden-to-output weights are the target values, so the output is a weighted average of the target values of training cases close to the given input case. [4]

SVM classifiers:
This classifier finds the optimal hyper plane that separates clusters of data in the feature space. The data is mapped by a kernel function into a different space where the data can be separated by a selected hyperplane. [2]

Data description:
In this paper, weuse ERUDIT Pap smear dataset.This dataset contains 552 single-cell images, including 138 normal cellsand 414 abnormal cells, obtained from Ref. [6].

The resolutionof the images in this dataset is 0.201 _m per pixel. They wereclassified into 4 classes based on the degree of abnormality as follows:

Normal cell images with 63 superficial cells
75 intermediate cells (138 cells in total),
Low-grade squamous intraepithelial lesions
(LSIL) (138 cells),
High-grade squamous intraepithelial lesions (HSIL) (138cells),
Squamous cell carcinoma (SCC) (138 cells)

Experimental Evaluation

The classification is evaluated based on six measures like sensitivity, specificity, spearman Rank order correlation co efficient (Rs)[], the Cohen's Kappa coefficient(k),and the weighted kappa coefficient(kw)[2]

Accuracy: It shows the percentage of data are classified to the correct class(i.e.)normal as normal class or abnormal as abnormal class[2]

$$\text{Accuracy} = \frac{\sum_{i=1}^{c}(\text{No.of correctly classified cellsa in classi})}{\sum_{i=1}^{c}(\text{Total no.of cells in classi})} \quad (4.1)$$

Sensitivity: it shows the percentage of the abnormal data that are correctly classified as abnormal. [2]

$$\text{Sensitivity} = \frac{\text{Number of True positive(TP)}}{\text{Number of True positive}+(\text{Number of False Positive})'(\text{FP})} \quad (4.2)$$

Specificity: It shows the percentage of the normal data that are correctly classified as true negative (Normal) [2]

$$\text{Specficity} = \frac{\text{Number of True Negative(TN)}}{\text{Number of True Negative} + (\text{Number of False Positive})'(\text{FP})} \quad (4.3)$$

Spearman rank order correlation coefficient (Rs): It is the correlation between ground truth ranking (Ui) and algorithm ranking (Vi) of cell,i=0,1,2….N

If the cells are in the same class have the same rank [2]

$$Rs = \frac{\sum_{i=1}^{n}(\text{Ui}-\text{U}')(\text{Vi}-\text{V}')}{\sum_{i=1}^{n}(\text{Ui}-\text{U}')\text{^}2\sum_{i=1}^{n}(\text{Vi}-\text{V}')\text{^}2} \quad (4.4)$$

The Cohen's kappa coefficient (*k*): it is a chance correlated agreement between desired class and assigned class [1] .This coefficient can be measured from the number of data from the number of data in assigned class and assigned class j from the confused matrix [2]

$$Kw = \frac{\text{p0}-\text{pe}}{1-pe} \quad (4.5)$$

And also, Positive Predictive Value (PPV) and Negative Predictive Value (NPV), both are used to evaluate the classifiers performance

$$\text{PPV} = \frac{\text{Number of True positive(TP)}}{\text{Number of True positive(TP)}+\text{Number of False Positive(FP)}} \quad (4.6)$$

$$\text{NPV} = \frac{\text{Number of True Negative(TN)}}{\text{Number of True Negative} + \text{Number of False Negative(FP)}} \quad (4.7)$$

Table 2: Performance of classifiers

Classifier	**Parameters**	**%**	**Classifier**	**Parameters**	**%**
ANN	Accuracy	92.33	GRNN	Accuracy	99.22
	Sensitivity	92.67		Sensitivity	98.64
	Specificity	92		Specificity	99.58
	PPV	89.6057348		PPV	99.78
	NPV	97.9423868		NPV	99.88
	Rs	0.85		Rs	0.99
	K	0.85		K	0.99
	Kw	0.85		Kw	0.99

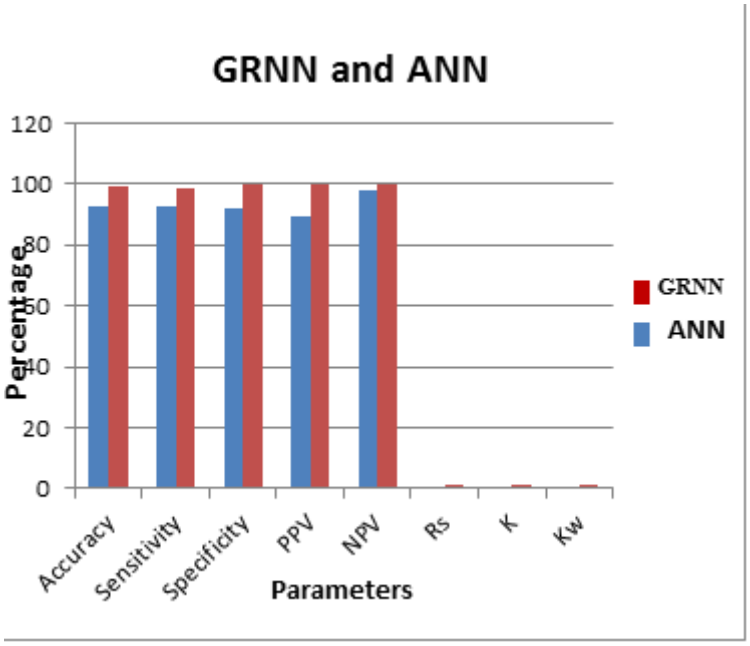


Figure 2: Performance of ANN classifier and GRNN classifier

The accuracy of ANN produced 92% whereas GRNN produces best accuracy of 99%. The ANN classifier produced specificity (93%), sensitivity (92%). But GRNN produces very good results like 98%, 99% in specificity, sensitivity respectively.

Conclusion and Future Work

This paper proposes performance of the classifier in automatic cervical cell image segmentation and classifications.With help of Fuzzy C-means clustering method used to separate nucleus and cytoplasm. Comparing the performance of ANN classifier with GRNN classifierproduces good result. In this work, implementing, the segmentation and classification is based on singlecervical cell.

A single cervical cell segmented as ,Nucleus and Non- Nucleus, or ,two to three regions,with the help of fuzzy c means clustering method. With help of extractedfeatures, GRNN produces 99%of accuracy, 98% of specificity,99 % of sensitivity. But this paper concentrates only in the single cell, noton overlapped multiple cells. In future work, this can be extended to segment and classify the overlapped and multiple cells

References

[1] A.Bergmeir, M.G. Silvente, J.M. Benitez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software frame work", Computer Methods and Programs in Biomedicine 107 (2012)497–512.[10]

[2] A.Genctav, S. Aksoy, S. Onder, "Unsupervised segmentation and classification of cervical cell images", PatternRecognition 45 (2012) 4151–4168.[2]

[3] C.G. Loukas, A. Linney, "A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia, and vasculature", Computer Methods and Programs in Biomedicine 74 (2004) 183–199[7]

[4] Donald F. Specht ,"**A** General Regression Neural Network", IEEE Transactions on Neural Networks 2(6)(1991)568-576.[14]

[5] H. Doornewaard, Y.T. van der Schouw, Y. van der Graaf,Reproducibility in double scanning of cervical smears with the PAPNET system, ActaCytologica 44 (4) (2000) 604–610.[6]

[6] http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html

[7] http://www.oreilly.com

[8] J. Jantzen, The ERUDIT Pap smear Tutorial, availableat: http://fuzzy.iau.dtu.dk/smear/[40-c]

[9] J. Kong, O. Sertel, H. Shimada, K.L. Boyer, J.H. Saltz, M.N.Gurcan, "Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation", Pattern Recognition 42 (2009)1080–1092.[9]

[10] J.S. Lee, L. Kuan, S. Oh, F.W. Patten, D.C. Wilbur, A feasibility study of the AutoPap system location-guided screening, ActaCytologica 42 (1) (1998) 221–226.[5]

[11] Jasdeep Singh, Charanjit Singh ,"Design of Low Pass FIR Filter Using General Regression Neural Network (GRNN)", International Journal of Advanced Research in Computer Science and Software Engineering 10(3)(2013)[13]

[12] L. He, L.R. Long, S. Antani, G.R. Thoma, "Histology image analysis for carcinoma detection and grading", Computer Methods and Programs in Biomedicines 107 (2012)538–556.[8]

[13] L.G. Koss, C. Gompel, C. Bergeron, Introduction to Gynecologic Cytopathology: With Histologic and Clinical Correlations, Lippincott Williams & Wilkins[3]

[14] M.E. Plissiti, C. Nikou, A. Charchanti, Automated detection of cell nuclei in Pap smear images using morpho logicalre construction and clustering, IEEE Transactions on Information Technology in Biomedicine 15 (2) (2011)233–241.[22-c]

[15] M.-H. Tsai, Y.-K. Chan, Z.-Z Lin, S.-F. Yang-Mao, P.-C. Huang, Nucleus and cytoplast contour detector of cervical smearimage, Pattern Recognition Letters 29 (2008)1441–1453.[23-c]

[16] Mark S. Nixon Alberto S. Aguado ,Feature Extraction and Image Processing, published British Library Cataloguing in Publication Data, by ISBN 0 7506 5078 8

[17] N. Theera-Umpon, S. Dhompongsa, "Morphological granulo metric features of nucleus in automatic bone marrow white blood cell classification", IEEE Transactions on Information Technology in Biomedicine 11 (3) (2007) 353–359[12]

[18] R. Ashfaq, B. Solares, M.H. Saboorian, "Detection of endo cervical component by PAPNET TM system on negative cervical smears", Diagnostic Cytopathology 15 (2) (1996)121–123[4]

[19] T.Chankong, N. Theera-Umpon, S. Auephanwiriyakul," Cervical cell classification using segmented-nucleus features", in: International Symposium on BiomedicalEngineering, 2009.[11]

[20] Thanatip Chankong, Nipon Theera-Umpon, Sansanee, Auephanwiriyakul," Automatic cervical cell segmentation and classification in papsmears", computer methods and programs in Biomedicine 113(2014)539-556.