私立東海大學

資訊工程研究所

碩士論文

指導教授：黃育仁 教授

抗核抗體紋路與統計特徵影像分類

HEp-2 Cell Images Classification Based on Textural and Statistic Featuresusing Self-Orgamizing Map

研究生：黃一宁

中 華 民 國 九 十 九 年 六 月

摘要

在組織性自我免疫的疾病中，HEp-2 細胞的間接性螢光切片檢查，常被用於檢測抗核抗體(antinuclear autoantibody, ANA)。本研究目的在於開發出一套對 HEp-2 細胞的間接性螢光抗核抗體影像，進行自動分類的系統。本論文的分類方法，是利用自我組織映射圖網路的演算法(self-organizing map, SOM)，結合 14 個細胞的紋理特徵與統計，對間接性螢光切片影像進行分類。本研究針對六種 ANA 類型影像(包含粗斑點型、細斑點型、分散斑點型、周邊型、均質型、核仁型)，共 1020 個抗核抗體細胞進行研究與分類。而最後實驗結果也證明本研究對於抗核抗體的螢光切片分類，提供一個高度正確率的判斷系統，因此對於自我免疫組織疾病的診斷，能提供有用的判斷與意見。

關鍵字：影像分類、自我組織映射網路、間接性螢光、抗核抗體、自我免疫疾病。

ABSTRACT

Indirect immunofluorescence (IIF) with HEp-2 cells has been used to detect antinuclear auto-antibodies (ANA) for diagnosing systemic autoimmune diseases. An automatic inspection system for ANA testing can be partitioned into HEp-2 cell detection, fluorescence pattern classification and computer aided diagnosis phases. The aim of this study is to develop an automatic classification scheme to identify the fluorescence patterns of HEp-2 cell in IIF images. The proposed method firstly utilizes a preprocessing procedure to reduce noise within IIF images. The feature extraction step calculates 14 practical textural and statistic features from fluorescence patterns. Then the proposed classification method performs an unsupervised neural network, the self-organizing map (SOM), to identify each fluorescence cell. This study evaluates 1020 autoantibody fluorescence patterns which were manually divided into six primary patterns, i.e. diffuse, peripheral, coarse speckled, fine speckled, discrete speckled and nucleolar patterns. The *k*-fold cross-validation method is used to evaluate the classification performance. From the simulations, the average accuracy using the proposed approach for identifying ANA patterns is 92.4%. The result shows that the proposed approach to identify autoantibody fluorescence patterns with a high accuracy and is therefore clinically useful to provide a second opinion for diagnosing systemic autoimmune diseases.

Keywords: immunofluorescence pattern, antinuclear autoantibodies, systemic autoimmune diseases, cell detection, image classification

INDEX

LIST OF TABLES

LIST OF FIGURES

CHAPTER1

INTRODUCTION

Autoimmune diseases have been confirmed to be in connection with the occurrence of disease specific autoantibodies, such as systemic autoimmune rheumatic diseases, primary biliary cirrhosis and dermatomyositis [1;2]. Exploiting indirect immunofluorescence (IIF) to identify the antinuclear autoantibodies (ANA) patterns of HEp-2 cell during the serological hallmark, the physician could diagnose the autoimmune diseases. The fluorescence patterns are usually identified by physician manual inspecting the slides with the help of a microscope [3]. This procedure still needs highly specialized and experienced physician to make diagnoses due to lacking in satisfied automation of inspection. For this reason, automatic identification and classification for fluorescence patterns in IIF image could offer physicians in making correct diagnosis without relevant experience. As ANA testing becomes more widespread used, clinical application of functional automatic inspection system is in great demand.

Perner et al. [4] performs the decision tree model with textural features for HEp-2 cell image classification. However, the error rate of classification achieves 25%. A recent approach combines expert system and classification technique to develop a staining pattern recognition system in ANA analysis [5]. The study utilizes

wavelet transformation to extract 360 features and then selects 180 features for identifying the fluorescence cells. Although the wavelet-based features achieve a satisfied classification result, the feature selection and classifier training steps for such a large feature space are very time-consuming.

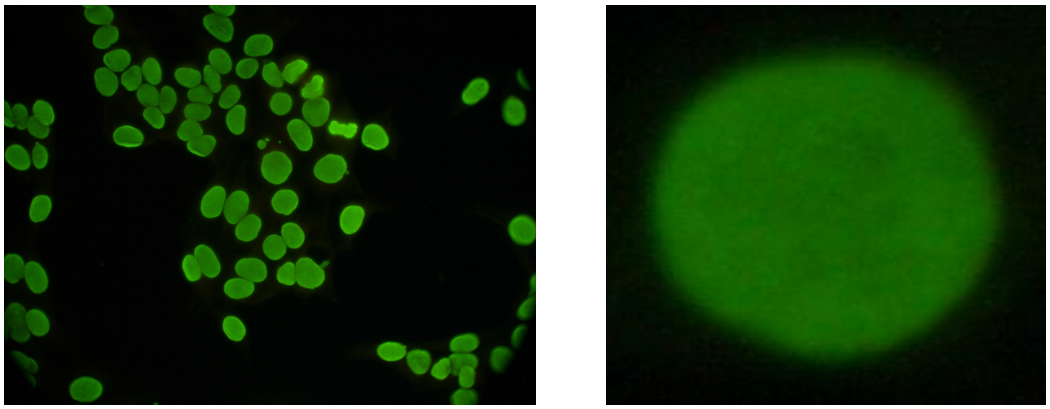
Accordingly, this study developed a classification scheme that adopted efficient textural and statistic features in a lower dimension to identify fluorescence cells. After image preprocessing, four textural features and ten statistics features were extracted from a fluorescence cell image. The proposed method utilized a well-known unsupervised neural network model, i.e. the self-organizing map (SOM) [6] [7], to classify the homogenous cells and then accomplished automatic identification for the six primary ANA patterns. This study evaluated 1020 fluorescence cells images. The computer simulations revealed that the average accuracy using the proposed classification method was 92.4%. The result shows that the proposed SOM model with the textural and statistics features is clinically feasible for fluorescence pattern classification.

CHAPTER2

MATERIALS AND METHODS

2.1 DATA ACQUISITION

This study used slides of HEp-2 substrate, at a serum dilution of 1:80. A physician takes images of slides with an acquisition unit consisting of the fluorescence microscope coupled with a commonly used fluorescence microscope (Axioskop 2, CarlZeiss, Jena, Germany) at 40-fold magnification. The immunofluorescence images were taken by an operator with a color digital camera (E-330, Olympus, Tokyo, Japan). The digitized images were of 8-bit photometric resolution for each RGB (Red, Green and Blue) color channel with a resolution of 3136×2352 pixel. Finally, the images were transferred to a personal computer and stored as *.orf-files (Raw data format) without compression. The samples were collected from January 2007 to May 2009 and test image database containing 1020 cells with manual segmentation form the samples. Figure 1 shows the result of manual segmentation for sample.



(a) (b)

Fig. 1. Indirect immunofluorescenc (IIF) image acquisition: (a) an original image (b) the manually segmented cell image

2.2 AUTOANTIBODY FLUORESCENCE PATTERNS

This study evaluated six primary ANA patterns that included diffuse pattern, peripheral pattern, coarse speckled pattern, fine speckled pattern, discrete speckled pattern and nucleolar pattern. For the diffuse pattern, the inside area of cell is always apparent homogeneous. In the peripheral pattern, the fluorescence intensity surrounding nucleus is higher than that of the outer region. The nucleolar pattern shows that large fluorescence speckled staining within nucleus. The speckled patterns can be further separated into three distinct types, i.e. coarse speckled, fine speckled and discrete speckled patterns. In the coarse speckled and fine speckled pattern, the nucleuses apply in homogeneous form and some dark speckled areas may appear within nucleus. For the discrete speckled pattern, there are many tiny fluorescence speckled stains in nucleus, and the other areas exhibit weaker fluorescence intensity. Figure 2 illustrates the distinct autoantibody fluorescence patterns.

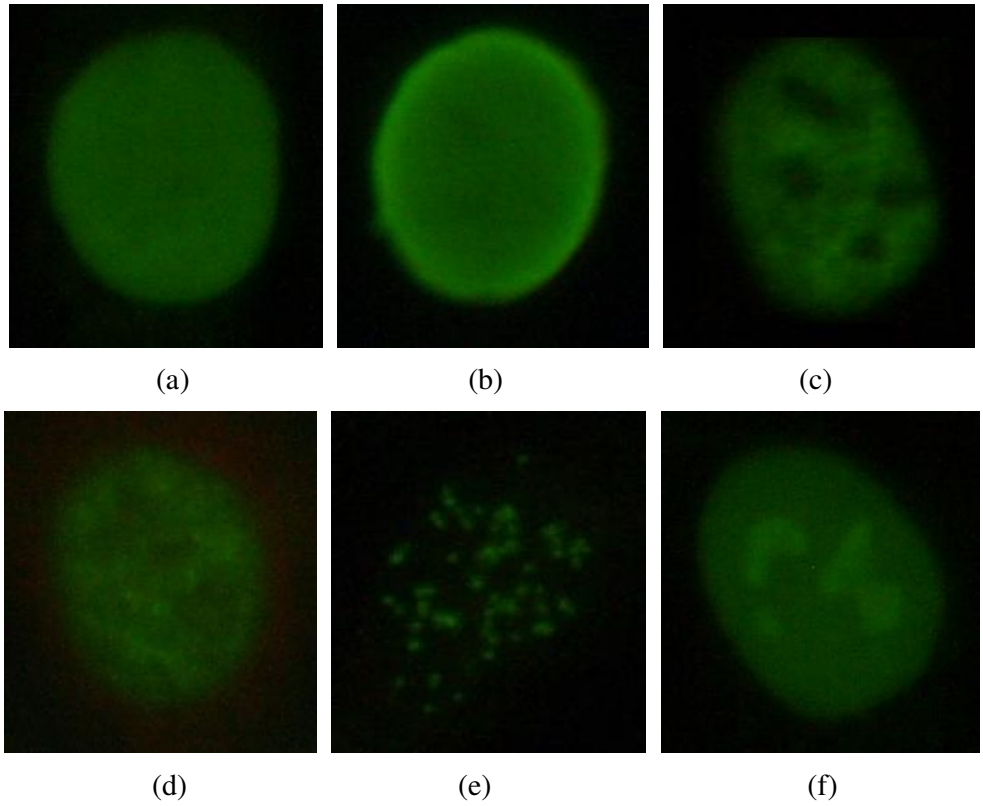


Fig. 2. The six primary types of ANA pattern: (a) diffuse, (b) peripheral, (c) coarse speckled, (d) fine speckled, (e) discrete speckled and (f) nucleolar patterns

2.3 THE PROPOSED CLASSIFICATION METHOD

The overall processes of the proposed method included the preprocessing, feature extract and pattern classification phases. This approach could be classified into two main stages: (1) the feature extraction stage and (2) the test set sample identifying stage. In the preprocessing and feature extraction phases, the proposed method transformed the RGB channels of an IIF image to HSI (Hue, Saturation and Intensity) color spaces and then extracted 14 features from the intensity channel. The pattern classification phase created the map of SOM with the 14 features and mapping with test samples to evaluate the classification results.

2.4 PREPROCESSING

An original IIF image always comprises image noises, speckles and textures. This study performed the Wiener method [8-10] as the preprocessing procedure to reduce the noises and preserve beneficial textures. The Wiener filtering is a pixel-wise adaptive low-pass technique which based on statistics estimated from the local neighborhood of each pixel. Due to the cells inside of diffuse and peripheral patterns appear smooth; the tiny speckle would be reduced of inside area after Wiener filtering. The cell images used the filter of Wiener can make the discriminative feature values between the speckle and smooth patterns. Thus the proposed method used the image with preprocessing of Wiener to extract the textural feature.

2.5 FEATURE EXTRACTION

The proposed classification approach utilized self-organizing map (SOM) based on the similarity of cells. This study performed the Otsu's method to automatically segment the entire cell region (denotes *E-region*) and then obtained the boundary region (denotes *B-region*) and inner region (denotes *I-region*) of the cell. Figure 3 illustrates the defined regions for an ANA cell. Ten statistic features and four textural features from the three distinct regions were performed to identify fluorescence

patterns. The features are listed as follows:

- *Area* (Area of a region in the cell) - The $Area_{E\text{-}region}$, $Area_{B\text{-}region}$ and $Area_{I\text{-}region}$ were calculated as statistic features in this study.
- *Intensity* (Average intensity of a region in the cell) - The $Intensity_{B\text{-}region}$ and $Intensity_{I\text{-}region}$ were estamted.
- *Higher intensity ratio* (Higher intensity ratio of *E-region*, *B-region* and *I-region*) - The ratio which the area of higher intensity compared with the whole cell, the rate based bright region of the *E-region*, *B-region* and *I-region* generated as features. The higher intensity define was used the Otsu's method characteristic to discriminate cell and bright speckle between the *E-region*, *B-region* and *I-region*. These features can detect higher intensity range of a cell, and using these features to classify the cells pattern was intuition.
- *Darker intensity ratio* (Darker intensity ratio of *E-region*, *B-region* and *I-region*) - The way of extract these features were same as *higher intensity ratio.* First step was inverse the image region of *Area*, and next used the result to find the darker region of *E-region*, *B-region* and *I-region*. These features were sensitive with speckle, when the speckle occur the features value were increase.

- *Uniformity* - The uniformity is a useful texture measure based on the histogram of an image, and the smoother fluorescence cell would obtain the larger value. "Uniformity" given by equation (1), where Z is a random variable denoting gray levels and *p(Zi)*, $i = 0, 1, 2, \ldots, L\text{-}1$, is the corresponding histogram, L is the number of distinct gray levels. Because the p's have values in the range [0, 1] and their sum equals 1, measure U is maximum for an image in which all gray levels are equal (maximum uniform), and decreases from there.

$$U = \sum_{i=0}^{L-1} P^2(\mathrm{Z}_i) \tag{1}$$

- *Standard deviation* (Standard deviation of *E-region*) - The standard deviation of *E-region* indicated that the distribution of histogram, the rougher inside area would also obtain the larger value.

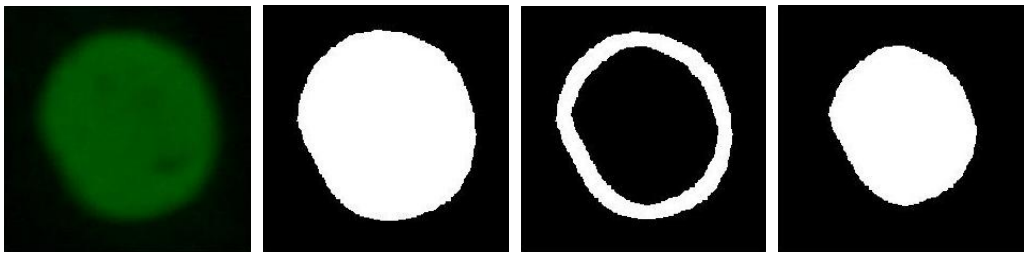
- *Auto-correlation coefficients* - The normalized auto-correlation coefficients can be used to reflect the inter-pixel correlation within an image. The 2-D normalized auto-correlation coefficient between pixel (i, j) and pixel ($i + \Delta m$, $j + \Delta n$) in an image with size $M \times N$ is defined as

$$r(\Delta m, \Delta n) = \frac{A(\Delta m, \Delta n)}{A(0,0)}, \tag{2}$$

where

$$A(\Delta m, \Delta n) = \frac{1}{(M-\Delta m)(N-\Delta n)} \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} f(x, y) f(x+\Delta m, y+\Delta n) \quad (3)$$

The auto-coefficients used the 2-D normalized auto-coefficients to measure the features, and the parameter was set as 5 and 10 in this study.



(a) (b) (c) (d)

Fig. 3. The area of features extraction: (a) a diffuse pattern, (b) the detected cell area, (c) the *B-region* features extract area and (d) the *I-region* features extract area

2.6 CLASSIFICATION

Self-organizing maps (SOM), an unsupervised learning of artificial neural network technique, has been widely used to classify data in many clusters. This technique have been applied extensively for data classification [11-14], region identify [15;16], image analysis [16;17], and the result is convincing and feasible. The proposed method exploited the advantages of SOM mode for identifying fluorescence patterns in IIF images. In this study, the 14 statistic and textural features from the fluorescence cells were utilized as input of SOM classifier. The output of the SOM was a map address, the address values corresponded with a type of fluorescence

pattern.

In the training stage, a number of maps were created in order to determine the best-fit map. The best-fit map depended on the quantization and topographic error. The training procedure would select a map with the less quantization and topographic error for making the best-fit map. Firstly, SOM was used the features to created map and decide the best-fit map, the next was labeling the map and output the map address. Figure 4 illustrates the final map of the proposed method. When SOM map was created, the test set could use the map to found the similar map address and identify the fluorescence pattern.

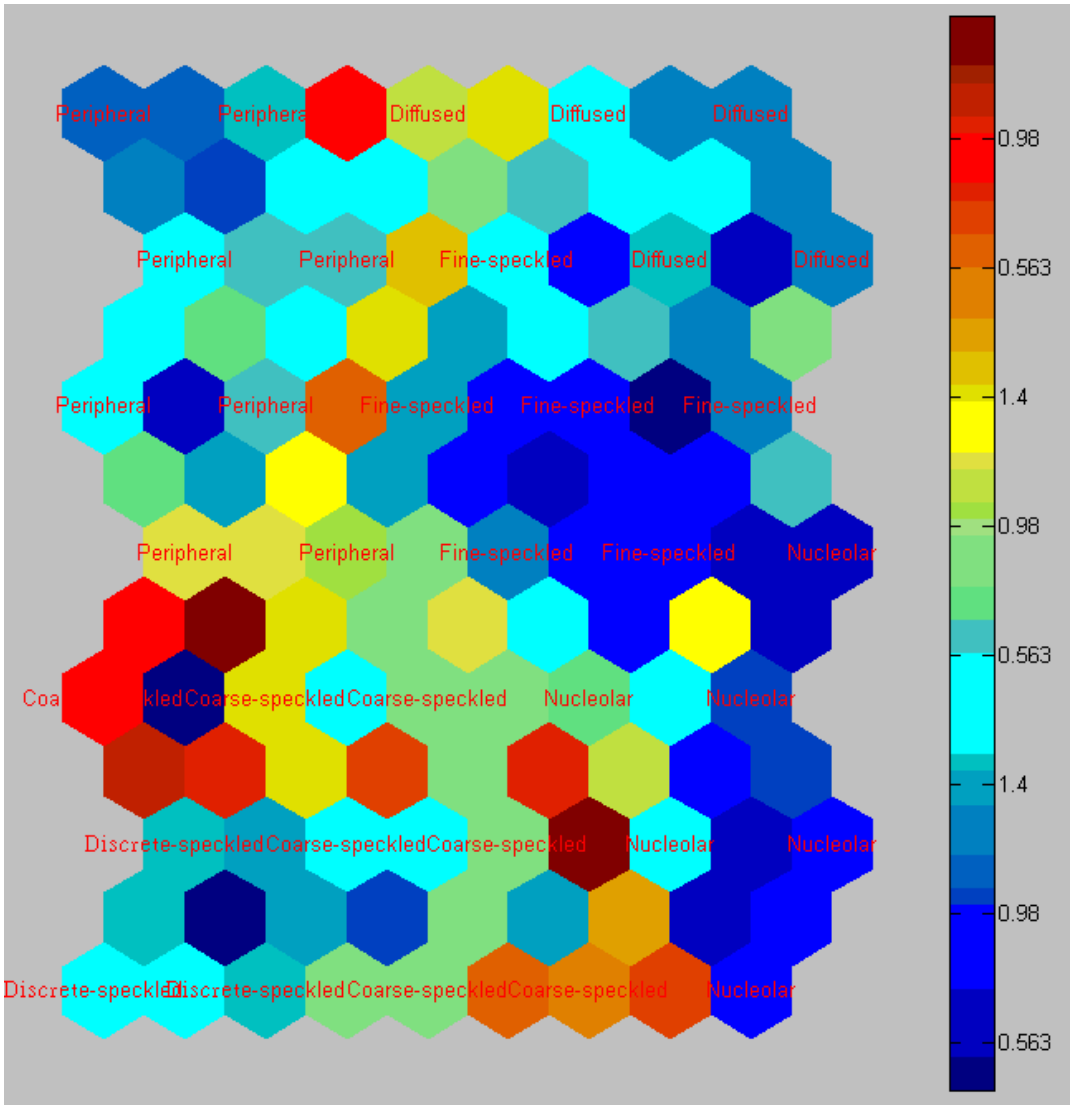


Fig. 4. The obtained SOM map in this study

2.7 EVALUATION

This study used *k*-fold cross-validation method [18-20] to evaluate the classification performance of the proposed classification method. When the data amount is not enough, the *k*-fold method can solve the situation and construction the normal train and test evaluation. Firstly, every ANA pattern was divided into *k* groups. Once trained one of *k* groups, the classifier tested the remainder groups. This process repeated same steps until all *k* groups have been used in turn as the group that is used for testing.

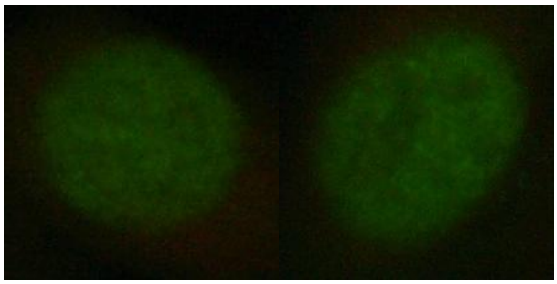
CHAPTER3

RESULTS

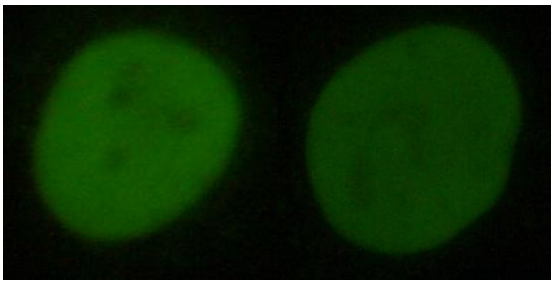
The ANA pattern database, including 130 diffuse patterns, 200 peripheral patterns, 200 coarse speckled patterns, 130 fine speckled patterns, 200 discrete speckled patterns, and 160 nucleolar patterns. Every IIF image was acquisition from the patients, and the difference disease mapping to specific ANA pattern. In this study, some ANA simples not enough 200 because of the mapping disease was unusual(like diffuse, fine speckled and nucleolar patterns). The database totally contained 1020 cells, and assign equally to the *k*-folds (k =10). The map size of the SOM model would affect the classification performance. This study adopted a 9×9 SOM classifier by observing the simulations. Table 1 lists the accuracy of the proposed SOM classification for identifying the fluorescence patterns. Table 2 lists the classification results of the ANA database. The best classification appeared for the peripheral pattern and the worse classification found in the fine speckled pattern. Figure 5 illustrates the samples which acquired incorrect result of classification. The fine speckled patterns, as shown in Fig. 5(a), were classified into the diffuse pattern; the diffuse patterns in Fig. 5(b) were categorized into the fine speckled pattern. In table 2, the red mark area was the incorrect classification cells number of diffuse and fine speckled, and the gray area was the correct classification cells number. This thesis

used the statistic and textural features to classify the ANA patterns, and texture of the diffuse samples (as in Fig. 5) was not regular within the normal samples. The normal diffuse patterns should not have speckles inside the cells, but some samples (like Fig. 5(b)) still contain these speckled regions. In the textural feature extraction procedure, it would lead to the different feature values which were decided by the speckles inside the cells or not. Those diffuse simples like Fig. 5(b) may be identified with fine speckled patterns because of some speckles inside the cells. On the contrary, fine speckled patterns would also encounter the similar problem. These two situations could lead to the fine speckled and diffuse patterns be identified failure. The proposed SOM classifier for the same type of the patterns would generate the unreasonable feature values and then obtained an incorrect classification result.

The simulations were made on a single CPU Intel(R) Core(TM)2 2.13 GHz personal computer (ASUSTek Computer Inc., Taipei, Taiwan) with Microsoft Windows Vista® operating system. The system was performed by using MATLAB (R2008.a) software (The MathWorks, Inc., Natick, MA).



(a)



(b)

Fig. 5. The samples with inaccurate classification: (a) fine speckled patterns and (b) diffuse patterns

Table 1 The simulation results of SOM classification

Fluorescence patterns	***NT***	***NC***	***Accuracy***
Diffuse	130	112	86.1%
Peripheral	200	200	100%
Coarse speckled	200	191	95.5%
Fine speckled	130	104	80.0%
Discrete speckled	200	196	98.0%
Nucleolar	160	140	87.5%
Total	1020	943	92.4%

NT: number of test cells

NC: number of correct classifications

Accuracy= NC / NT × 100%

Table 2. The classification results of fluorescence pattern database

Fluorescence patterns	Diffuse	Peripheral	Coarse speckled	Fine speckled	Discrete speckled	Nucleolar
Diffuse	112	0	2	11	0	6
Peripheral	3	200	0	0	0	0
Coarse speckled	0	0	191	8	4	4
Fine speckled	12	0	5	104	0	9
Discrete speckled	0	0	0	0	196	1
Nucleolar	3	0	2	7	0	140
Total	130	200	200	130	200	160

CHAPTER4

DISCUSSION

This study developed a classification scheme to identify IIF image fluorescence patterns. The database including 1020 fluorescence cells which manually segmented from IIF images. The proposed classification method performed the SOM classifier with 14 features to identify the ANA patterns. For the six distinct ANA patterns, the classification accuracy was 86.1%, 100%, 95.5%, 80%, 98% and 87.5% in the diffuse, peripheral, coarse speckled, fine speckled, discrete speckled and nucleolar patterns. The average accuracy achieved 92.4%. In the experiment, we use the difference SOM map size to classification the ANA patterns, and the fig. 6 show the evaluate result. In fig. 6, the size of 9×9 can make better classification result than other maps size. The SOM classifier rule is Winner-Takes-All, when use same database mapping to distinct SOM map size may get the difference result. Finally, this propose choose the map size of 9×9 to identify the ANA patterns.

From the experimental results, the proposed method for IIF cells image classification using SOM artificial neural network with ten statistic features and four textural features is feasible. As this automated classification system is useful for identifying autoantibody fluorescence patterns and decrease the time of physician diagnosis. In the future, we hope to improve the proposed system by using dynamic

map size selected to fit every database automatically. Future work should combine the proposed fluorescence pattern classification method with automatic IIF image cells segmentation system, and apply them to computer-aided diagnosis (CAD) system for systemic autoimmune diseases.

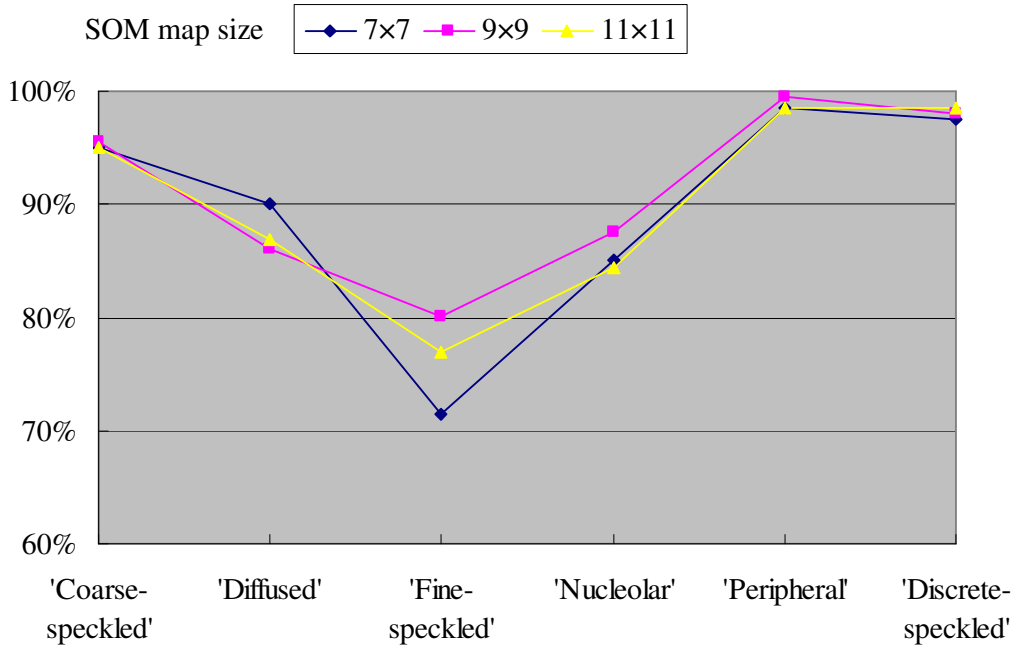


Fig. 6. the evaluation result of difference SOM map size

REFERENCE

[1] M. J. Fritzler, "Challenges to the use of autoantibodies as predictors of disease onset, diagnosis and outcomes," *Autoimmunity Reviews*, vol. 7, no. 8, pp. 616-620, Sept.2008.

[2] H. J. Worman and J. C. Courvalin, "Antinuclear antibodies specific for primary biliary cirrhosis," *Autoimmunity Reviews*, vol. 2, no. 4, pp. 211-217, June2003.

[3] K. Conrad, W. Schoessler, and F. Hiepe, *Autoantibodies in systemic autoimmune diseases* Pabst Science Publishers, 2002.

[4] P. Perner, H. Perner, and B. Muller, "Mining knowledge for HEp-2 cell image classification," *Artificial Intelligence in Medicine*, vol. 26, no. 1-2, pp. 161-173, Sept.2002.

[5] P. Soda and G. Iannello, "Aggregation of Classifiers for Staining Pattern Recognition in Antinuclear Autoantibodies Analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 3, pp. 322-329, May2009.

[6] W. Kurdthongmee, "Colour classification of rubberwood boards for fingerjoint manufacturing using a SOM neural network and image processing," *Computers and Electronics in Agriculture*, vol. 64, no. 2, pp. 85-92, Dec.2008.

[7] E. Kita, S. Kan, and Z. Fei, "Investigation of self-organizing map for genetic algorithm," *Advances in Engineering Software*, vol. 41, no. 2, pp. 148-153, Feb.2010.

[8] E. Ercelebi and S. Koc, "Lifting-based wavelet domain adaptive Wiener filter for image enhancement," *IEEE Proceedings-Vision Image and Signal Processing*, vol. 153, no. 1, pp. 31-36, Feb.2006.

[9] V. Katkovnik, K. Egiazarian, and J. Astola, "A spatially adaptive nonparametric regression image deblurring," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1469-1478, Oct.2005.

[10] R. Hardie, "A fast image super-resolution algorithm using an adaptive Wiener filter," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2953-2964, Dec.2007.

[11] M. Guerra, C. Pinuela, A. Andres, B. Galan, and J. R. Viguri, "Assessment of Self-Organizing Map artificial neural networks for the classification of sediment quality," *Environment International*, vol. 34, no. 6, pp. 782-790, Aug.2008.

[12] M. Mohamed, "Clustering the ecological footprint of nations using Kohonen's self-organizing maps," *Expert Systems with Applications*, vol. 37, no. 4, pp. 2747-2755, 2010.

[13] H. C. Zheng, G. Lefebvre, and C. Laurent, "Fast-learning adaptive-subspace self-organizing map: An application to saliency-based invariant image feature construction," *IEEE Transactions on Neural Networks*, vol. 19, no. 5, pp. 746-757, May2008.

[14] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 902-912, July2003.

[15] G. F. Lin and L. H. Chen, "Identification of homogeneous regions for regional frequency analysis using the self-organizing map," *Journal of Hydrology*, vol. 324, no. 1-4, pp. 1-9, June2006.

[16] C. Vijayakumar, G. Damayanti, R. Pant, and C. M. Sreedhar, "Segmentation and grading of brain tumors on apparent diffusion coefficient images using self-organizing maps," *Computerized Medical Imaging and Graphics*, vol. 31, no. 7, pp. 473-484, Oct.2007.

[17] I. Guler, A. Demirhan, and R. Karakis, "Interpretation of MR images using self-organizing maps and knowledge-based expert systems," *Digital Signal Processing*, vol. 19, no. 4, pp. 668-677, July2009.

[18] S. M. Weiss and I. Kapouleas, "An empirical comparison of pattern recognition neural nets and machine learning classification methods," *Proc 11th Int Joint Conf Artificial Intelligence*, pp. 234-237, 1989.

[19] T. S. Wiens, B. C. Dale, M. S. Boyce, and G. P. Kershaw, "Three way k-fold cross-validation of resource selection functions," *Ecological Modelling*, vol. 212, no. 3-4, pp. 244-255, Apr.2008.

[20] S. Kawaguchi and R. Nishii, "Hyperspectral image classification by bootstrap AdaBoost with random decision stumps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3845-3851, Nov.2007.