

抗核抗體免疫螢光影像特徵萃取與分類之研究

謝宗佑¹ 張世旭¹ 謝祖怡²

¹大葉大學資訊工程學系

51591 彰化縣大村鄉學府路 168 號

²台中榮民總醫院

40705 台中市西屯區台中港路三段 160 號

摘要

在免疫疾病診斷上，抗核抗體（anti-nuclear antibodies, ANA）的分類可幫助病情的診斷，目前常用的檢驗方式是以間接免疫螢光法（indirect immunofluorescence）觀察螢光影像的抗核抗體的螢光形態。傳統上醫師以人眼觀察螢光影像進行分類，此作法傷眼及費時，因此本論文提出以電腦輔助對抗核抗體影像進行分類的方法。首先使用影像處理的技術先萃取影像特徵，再以特徵將影像分為均質型（homogeneous）、周邊型（peripheral）、粗點型（coarse speckled）、與離散點型（discrete speckled）共四類型。特徵抽取法是從影像中找出細胞區域，從各細胞區域分別萃取出 7 種特徵，特徵為拉普拉斯運算值（水平與垂直的 2 次微分、斜對角的 2 次微分）的平均值與標準差、曲率的平均值與標準差、及細胞內灰階值的標準差。當各細胞區域的特徵向量取出後，找出具代表性的特徵向量代表該影像的特徵。輸入測試影像計算該特徵與各類訓練影像特徵集合的距離，以最小距離分類法對特徵分類。實驗結果證明所提出方法的四類型的分類正確率至少為 94%，系統整體分類正確率達 98.75%。證明抗核抗體影像可以電腦輔助進行分類，所提出的分類方法簡單快速易於軟體實作。

關鍵詞：抗核抗體，螢光影像，特徵萃取，分類，電腦輔助系統

Extraction and Classification of the Features of Anti-nuclear Antibody Immunofluorescence Images

TSUNG-YU HSIEH¹, SHIH-HSU CHANG¹ and TSU-YI HSIEH²

¹Department of Computer Science and Information Engineering, Da-Yeh University
168 University Rd., Dacun, Changhua 51591, Taiwan

²Division of Allergy, Immunology and Rheumatology, Taichung Veterans General Hospital
No. 160, Sec. 3, Chung-Kang Rd., Taichung 40705, Taiwan

ABSTRACT

Identification of anti-nuclear antibodies (ANA) is an important step in the diagnosis of autoimmune diseases. The most commonly used procedure is indirect immuno-fluorescence (IIF), in which the physician expends considerable time in reading the image patterns visually. To ameliorate this problem, a computer-aided classification method is proposed in this report. This system (1) uses

image pre-processing techniques to obtain the regions of the nucleus, (2) extracts the feature information from the cell regions, (3) extracts the representative feature vectors from the features of the cell regions, (4) calculates the distance between the representative feature vectors in the test image and the training feature vectors, and (5) classifies the test image by a minimum distance classifier. In the experimental results, four patterns of anti-nuclear antibodies were used to evaluate the proposed method. These results indicated that the correct classification rate is above 94%, the average correct rate being 98.75%. This high classification rate shows that the proposed system is workable. Furthermore, this method is fast and easily implemented by software in a computer-aided system.

Key Words: anti-nuclear antibodies (ANA), immuno-fluorescence images, extraction of features, classification, computer-aided system

一、前言

抗核抗體 (anti-nuclear antibody, ANA) 是指人體中異常的抗體，屬於自體免疫疾病 (autoimmune disease)。正常的抗體會對抗外來的抗原 (antigen)，這些異常的抗體把人體本身的細胞當做外來的抗原，加以對抗，產生免疫相關疾病，常見的免疫疾病如系統性紅斑狼瘡 (systemic lupus erythematosus) 及系統性硬化症 (systemic sclerosis) 等。此種疾病的檢驗方式有酵素免疫分析法 (enzyme immunoassay) 與間接免疫螢光抗體法 (indirect immunofluorescence antibody)，其中間接免疫螢光法可顯示效價強度 (titer) 及螢光形態 (pattern)，有助於免疫疾病的分類及診斷。間接免疫螢光法的作法中會以紫外線激發螢光，以便醫療人員觀察抗核抗體附著於細胞核的位置，且為了減少判讀錯誤的發生，檢驗中醫療人員需觀察同一檢體 10 個觀景區塊，在長時間的觀察下對人眼易造成損傷，因此本論文為了維護醫療人員的健康與減少醫療人員的負擔，建立自體免疫疾病診斷上的電腦輔助醫療診斷系統 (computer aided diagnosis, CAD)。

電腦輔助診斷系統可提昇醫療水平與降低醫療人員的負擔，另可提供教學與訓練的功能。有關電腦科技輔助抗核抗體診斷的研究有：2001 年 Perner [7] 提出以資料探勘 (data mining) 技術仿效人類學習建立 n-ary 決策樹 (decision tree)，再以決策樹依影像特徵分類螢光影像，實驗結果分類正確率為 86.7%。Perner 等人 [8] 發表類似的論文，改變量化作法變更區間數，以及增加萃取的特徵，找出最佳參數。此外 Perner 等人 [9] 於 2002 年發表另一篇論文，將原本用於建立特徵分類決策樹的資料探勘技術，用於人類專家分類行為，建立分類行為決策樹，輔助人類專家分類，結果顯示經由輔助可使錯誤率由 23.6% 降至 16.6%。

2003 年 Sack [11] 發表基於 Perner 提出的分類方法 [7]，以電腦輔助分類免疫疾病中 HEp-2 螢光形態，實驗結果分類正確率為 83%。Soda 等人 [15] 在 2006 年提出針對抗核抗體螢光強度的專家系統分類方法，在同一個效價擷取螢光影像，分類螢光影像中的螢光強度。之後 2007 年 Rigon 等人 [10] 發表以抗核抗體數位影像輔助人類專家分類的助益，研究顯示醫檢師使用數位影像檢驗的信心度較高。同年 Soda 個人也發表另一篇研究 [14]，利用多專家系統分析螢光影像的螢光形態，結合 6 個最近鄰居分類器 (nearest neighbor classifier) 分類抗核抗體螢光影像，實驗結果顯示其方法的正確率為 70.5%。2007 年黃嘉政 [1] 發表應用資料探勘與類神經網路分類在 ANA 細胞的辨識率上，錯誤率為 18.8%。因此本文目的為提出一個高分類正確率的演算法，而欲達到高分類正確率，分類所採用的特徵非常重要，特徵得具有代表性，才足以將細胞類型分類成功。本文經實驗分析驗證，提出 7 種特徵分類抗核抗體螢光影像。

目前抗核抗體的螢光形態在醫學上可分為四大類型 [2, 4, 11]：均質型 (homogeneous)、周邊型 (peripheral)、斑點型 (speckled) 與核仁型 (nucleolar)，其中斑點型又可細分為細點型 (fine speckled)、粗點型 (coarse speckled) 與離散點型 (discrete speckled)。由於罹患細點型或核仁型的病人較少，本論文針對均質型、周邊型、粗點型、離散點型四類型進行分類。抗核抗體四類型影像由台中榮民總醫院免疫風濕科所提供，影像大小為 3136×2352 如圖 1 所示，為了清楚顯示圖 2 是圖 1 影像中間區域放大顯示。圖 2(a) 為均質型：細胞核內的螢光分佈均勻，核仁處因沒有染色質存在沒有螢光產生；圖 2(b) 為周邊型：細胞核內部如同均質型，在細胞核邊緣呈現較亮的螢光；圖 2(c) 為粗點型：細胞核內由許多亮點聚集成一螢光區域；圖 2(d) 為離散點型：細

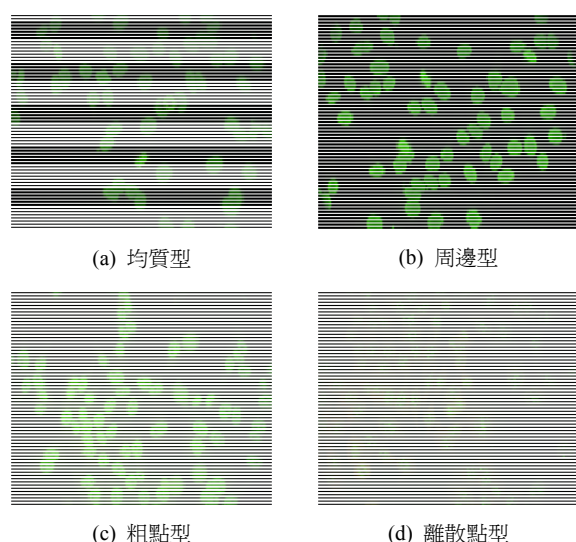


圖 1. 四類型的螢光影像

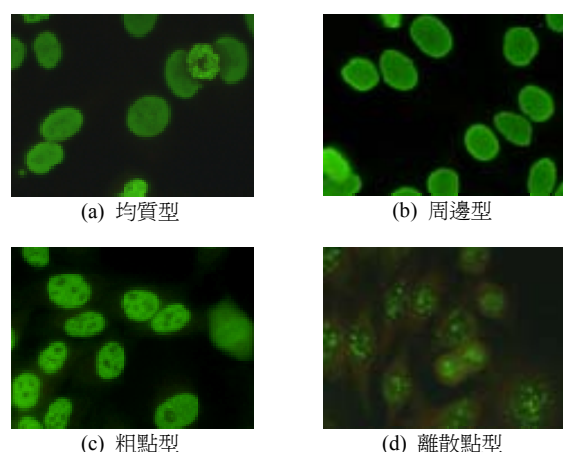


圖 2. 圖 1 影像中間區域放大顯示

胞核內猶如粗點型，其亮點較粗點型少但較亮。

由於間接免疫螢光法是利用紫外線激發螢光，醫療人員再透過顯微鏡觀察抗核抗體的螢光影像需花費大量時間在觀察檢體上，為了減少醫療人員的負擔與提升檢驗的速度，本論文提出針對抗核抗體螢光影像進行特徵萃取與分類之演算法。特徵萃取的作法是先將影像經由灰階化處理，再經影像平滑化與縮小處理。接著抽取螢光影像的細胞區域，並對每一細胞區域影像進行影像對比調整。之後對每一細胞區域的紋理進行分析，轉換成數據形式的特徵資訊，取出 7 個特徵組成特徵向量。最後挑選特徵，去除不相似細胞區域的特徵，取出代表性特徵。最後將代表性特徵與每一類訓練特徵計算相似度，依相似度對四類型投票，以獲得最大投票

數之類型為判定的螢光類型。

以下第 2 節說明如何從抗核抗體螢光影像中萃取特徵的方法，第 3 節說明將影像分類的演算法，第 4 節為實驗結果與討論，最後系統的結論在第 5 節說明。

二、特徵萃取演算法

本節將介紹所提出之特徵萃取演算法，擷取的影像為一連串的色彩影像，影像中包含細胞核與黑色的背景區域，因此在特徵萃取前，先使用影像處理去除影像中不必要的資訊，以減少運算量。再取得影像中所有細胞區域，針對每一細胞區域萃取其特徵資訊，最後分析所有細胞區域的特徵資訊，去除細胞為非正常時期的特徵資訊。特徵萃取演算法簡述如下：

1. 將輸入的 ANA 彩色影像轉成灰階影像；
2. 將灰階影像進行縮小及平滑處理；
3. 將平滑後的影像分割成細胞區域及背景，並抽取出各細胞區域；
4. 調整影像中各細胞區域之對比度；
5. 計算各細胞區域內像素的梯度 (gradient) 強度，設梯度大的像素為邊緣 (edge) 點；
6. 在邊緣點上計算拉普拉斯值 L_1 與 L_2 ，取 L_1 大的邊緣點為特徵點，計算特徵點上拉普拉斯值的平均值與標準差作為特徵值；
7. 計算各特徵點的曲率與特徵；
8. 計算細胞內部灰階強度的標準差作為特徵，最後將上述抽取的特徵組成特徵向量；
9. 從各細胞的特徵向量中挑選幾個具代表性之特徵向量，作為此影像之特徵向量集合。

以下將分項詳細介紹所提出的特徵萃取演算法。

(一) 影像灰階化

擷取的抗核抗體螢光影像為彩色影像包含 RGB 三種頻道資訊，而激發的螢光為綠色，以單一顏色頻道為主。因此在進行後續處理之前，將影像由彩色轉換為單一顏色頻道之影像，可減少資料運算，以下為轉換公式 [3]：

$$Y(x, y) = 0.299 \times r(x, y) + 0.587 \times g(x, y) + 0.114 \times b(x, y), \quad (1)$$

其中 $r(x, y)$ 、 $g(x, y)$ 與 $b(x, y)$ 分別為像素 (x, y) 的紅綠藍三強度值， $Y(x, y)$ 是轉換後的灰階強度。

(二) 影像平滑化與縮小處理

擷取過程中為了減少照明光影響影像的擷取，整個過程必須在暗室中進行。取像過程中 CCD 或 CMOS 經長時間充、放電的影響，擷取的影像充斥著雜訊。為了減少雜訊的影響，將灰階影像 Y 進行 2 次 Haar 小波轉換處理 [3, 12]，得到 $LL2$ 頻帶的平滑影像。此作法可降低雜訊的影響，且影像大小會縮為原影像 $1/16$ 倍，減少運算的資料量，縮短運算時間。原始影像大小為 3136×2352 ，經過 2 次 2D Haar 轉換為 $LL2$ ，影像縮小為 784×588 。

假設影像為 2×2 如圖 3(a) 所示，共有 4 像素 a, b, c ，與 d ，則 1 次 2D Haar 轉換的結果如圖 3(b) 所示， $LL1$ 值 $= (a+b+c+d)/4$ ，為四個像素值的平均值。若運行 2 次 2D Haar 轉換， $LL2$ 值為 16 個像素值的平均值。因為 16 個像素的平均值可以將雜訊總和平均，因此可降低雜訊 [3]。因為只是像素平均值運算，所以運算簡單及快速。

(三) 細胞區域抽取

因各類型是細胞區域內不同特性的表現結果，因此將影像分成背景與細胞區域，最後並抽取出一細胞區域。本研究採用 Otsu 門檻法 [6] 自動找出門檻值 TH_{Otsu} ，若像素灰階值大於門檻值 TH_{Otsu} 則為細胞；否則視為黑色背景。圖 4 為圖 1 處理後的細胞影像，影像中標示白色為大於門檻值的像素，代表細胞位置；在影像中標示黑色為小於或等於門檻值的像素，是影像中的黑色背景。

當分離出背景與細胞區域後，從區域中取出單一細胞以作後續的特徵抽取處理。使用 8-連通 (8-connectivity) 追蹤相連的細胞像素 (connected pixel) 標記為同一細胞區域 [3]，得到區域 R_i 。因為雜訊影響可能產生小面積的的細胞區域，因此去掉在影像上下左右四邊上的細胞及細胞內像素不足面積門檻值 TH_{area} ，最後取得 n 個細胞區域 R_i ， $i=1, \dots, n$ 。 TH_{area} 的設定與原始影像大小有關，若原始影像較大，則門檻值定高些；若原始影像較低，則門檻值定低。在

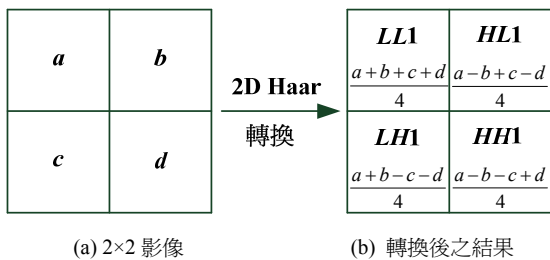


圖 3. 2D Haar 轉換的示例

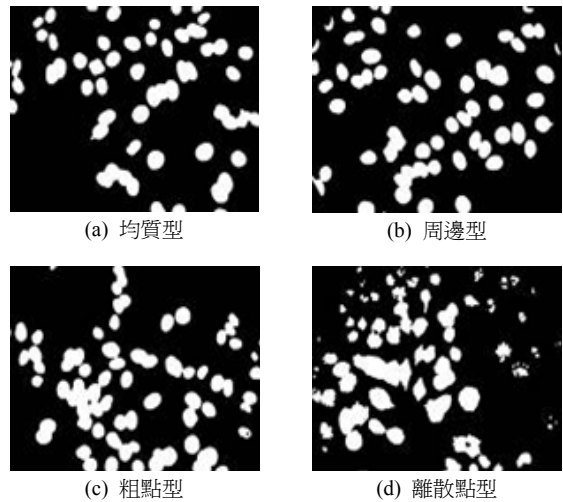


圖 4. 圖 1 經細胞區域抽取之結果，其中白色為細胞區域，黑色為背景

實驗中調整門檻值使得訓練影像的分類正確率最高。當原始影像大小為 3136×2352 ，本文在實驗調試時門檻值 TH_{area} 設為 500。

圖 4 的四種類型影像經標示並刪去面積低於門檻值的區域，結果如圖 5 所示。圖 5(a) 為均質型經過刪去面積不足區域的結果影像，區域數量從 42 個減少為 31 個區域；圖 5(b) 為周邊型經過刪去面積不足區域的結果影像，區域數量從 47 個減少為 35 個；圖 5(c) 為粗點型經過刪去面積不足區域的結果影像，區域數量從 42 個減少為 30 個；圖 5(d) 為離散點型經過刪去面積不足區域的結果影像，由於離散點型的螢光區域較為破碎，可以很明顯的看出區域數量從 84 個減少至 33 個。

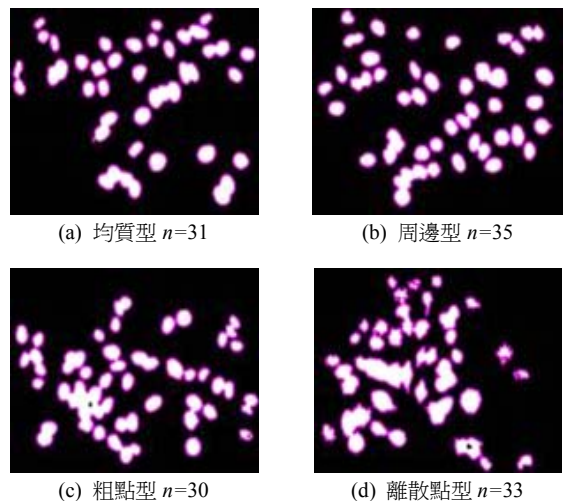


圖 5. 圖 4 經細胞標記法處理後之結果， n 為細胞區域個數

(四) 影像對比調整

由於來源影像是在不同的快門時間下拍攝，使不同影像的對比會有所差異，即使在同一張影像上，位於不同區域的影像對比也有所不同。為了減少每一細胞區域之間對比差異，研究中使用一種簡單且快速的技術，使每一細胞區域經由處理後，影像的對比呈現一致的狀態。首先計算每一細胞區域 R 的亮度最大值後，取得對比調整比例 $scale$ ，

$$scale = 255 / \max(R), \quad (2)$$

其中 $\max(R)$ 代表該細胞區域 R 的灰階最大值，將細胞內的像素灰階值乘上 $scale$ ，使調整後的灰階值介於 0~255 之間。

(五) 細胞內邊緣像素偵測

從圖 2 觀察週邊型在細胞外圍呈現較亮特性，斑點型在細胞內部的紋理 (texture) 變化較大。因此研究中以影像之梯度大小 (magnitude of the gradient) [3] 代表暗亮度或紋理變化之強度，細胞區域各像素梯度計算公式如下：

$$\Delta_x \approx I(x+1, y) - I(x-1, y), \quad (3)$$

$$\Delta_y \approx I(x, y+1) - I(x, y-1), \quad (4)$$

$$G(x, y) = \sqrt{\Delta_x^2 + \Delta_y^2}, \quad (5)$$

其中像素 (x, y) 位於細胞區域， $I(x, y)$ 是像素 (x, y) 的灰階值， Δ_x 代表像素 (x, y) 的水平灰階變化量， Δ_y 代表像素 (x, y) 的垂直灰階變化量， $G(x, y)$ 是像素 (x, y) 的梯度強度。當梯度越大代表明暗度變化大或紋理變化強，若梯度 $G(x, y)$ 大於門檻值 TH_G ，則像素 (x, y) 判定為邊緣 (edge) 點，門檻值 TH_G 設為該細胞區域內所有梯度 G 值的平均值 (mean)，後續的特徵抽取演算法將在細胞區域上的邊緣點上進行特徵萃取。

圖 5 經對比調整後，經過本步驟處理後細胞區域的邊緣點影像如圖 6 所示，為顯示清楚只取中間區域放大顯示，白色點為細胞邊緣像素點，黑色為非邊緣點。

(六) 拉普拉斯運算與特徵

從圖 6 可觀察影像內的細胞邊緣點不少，且大量相同的資訊易使細微的變化被忽略，因此本文考慮將邊緣點進一步分類，只保留具有特暗或特亮的邊緣像素點，將這些點當作特徵萃取點。作法是採用拉普拉斯運算 (Laplacian operator)

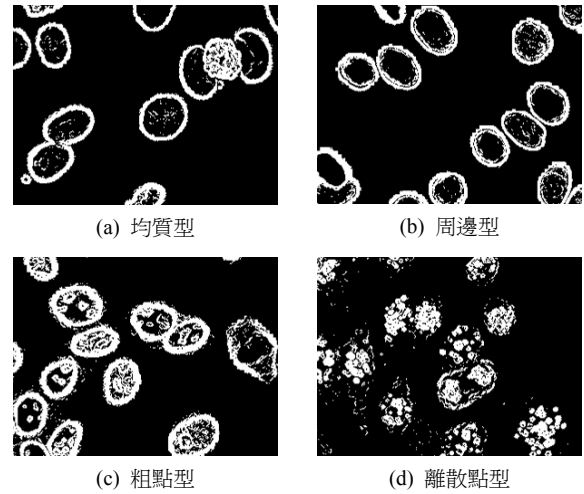


圖 6. 各類型的細胞邊緣影像，白點為 $G(x, y) > TH_G$ 的邊緣點；黑色點為非邊緣點

[3]，在邊緣點 (x, y) 上進行拉普拉斯遮罩運算，拉普拉斯遮罩 M_1 與 M_2 如圖 7 所示，依式 (6) 與式 (7) 運算結果得到拉普拉斯數 $L_1(x, y)$ 與 $L_2(x, y)$ 如下式

$$L_1(x, y) = \{I(x-1, y) + I(x+1, y) + I(x, y-1) + I(x, y+1)\} / 4 - I(x, y), \quad (6)$$

$$L_2(x, y) = \{I(x-1, y-1) + I(x+1, y-1) + I(x-1, y+1) + I(x+1, y+1)\} / 4 - I(x, y). \quad (7)$$

當像素 (x, y) 計算得到的拉普拉斯數較大，代表該點與鄰近四周點的灰階差異大，此點為較亮或較暗的點；若得到的數值較小，代表該點與鄰近四周點的灰階差度小，該點位於平坦之處。因此根據拉普拉斯值，可將細胞邊緣像素劃分兩部分，一為較暗或較亮的點，另一為亮度較平坦區域上的點。假設集合細胞 R 內所有邊緣點為集合 E ，計算集合 E 內所有 L_1 值的平均值為 $mean_E$ ，劃分的門檻值 TH_{L1} 設定為 $TH_{L1} = mean_E$ 。當邊緣像素點 (x, y) 且其拉普拉斯值 $L_1(x, y) > TH_{L1}$ ， (x, y) 視為特徵點。圖 8 為圖 6 判為特徵點的結果圖，白色點為特徵點，黑色點為非特徵點。集合細胞

$$(a) M_1 = \begin{bmatrix} 0 & 1/4 & 0 \\ 1/4 & -1 & 1/4 \\ 0 & 1/4 & 0 \end{bmatrix} \quad (b) M_2 = \begin{bmatrix} 0 & 1/4 & 0 \\ 1/4 & -1 & 1/4 \\ 0 & 1/4 & 0 \end{bmatrix}$$

圖 7. 拉普拉斯遮罩 M_1 與 M_2

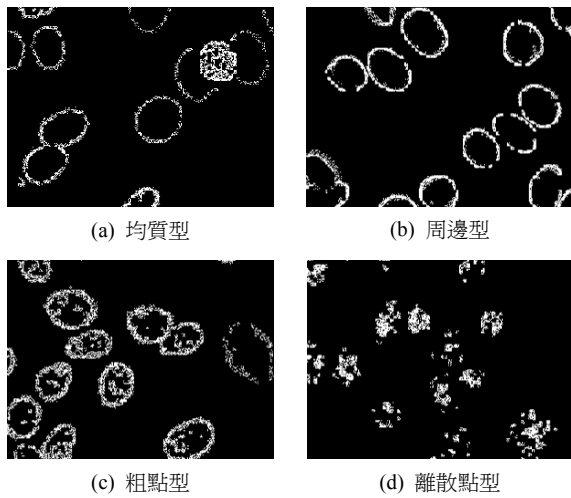


圖 8. 圖 6 判為特徵點的結果圖，白色點為特徵點；黑色點為非特徵點

R 內所有特徵點為集合 A ，計算集合 A 內所有 L_1 值的平均值 (mean average) μ_1 與標準差 (standard deviation) σ_1 作為特徵值 f_1 與 f_2 。計算集合 A 內所有 L_2 值的平均值 μ_2 與標準差 σ_2 作為特徵值 f_3 與 f_4 。因此每一細胞目前可取得 4 個特徵 $(f_1, f_2, f_3, f_4) = (\mu_1, \sigma_1, \mu_2, \sigma_2)$ 。

(七) 曲率的平均值與標準差

從圖 2 觀察均質型和周邊型，通常其細胞核膜外緣的曲率值 (curvature) 較小；而粗點型與離散點型較為不同，其點通常位於細胞核內，為一突起的亮點或亮點包圍的暗點，因此得到的曲率值會較大。因此本文考慮使用曲率作為一個特徵。

本文採用 Level Set [5, 13] 文中的方法計算曲率值，計算特徵點 (x, y) 上曲率的公式如下：

$$\Delta_{xx} \approx \Delta_x(x+1, y) - \Delta_x(x-1, y), \quad (8)$$

$$\Delta_{yy} \approx \Delta_y(x, y+1) - \Delta_y(x, y-1), \quad (9)$$

$$\Delta_{xy} \approx \Delta_x(x, y+1) - \Delta_x(x, y-1), \quad (10)$$

$$C(x, y) = \left| \frac{(\Delta_x^2 \Delta_{yy} - 2\Delta_x \Delta_y \Delta_{xy} + \Delta_{xx} \Delta_y^2)}{(\Delta_x^2 + \Delta_y^2)^{3/2}} \right|, \quad (11)$$

其中 Δ_{xx} 為水平的二次微分、 Δ_{yy} 為垂直的二次微分、及 Δ_{xy} 為水平微分上的垂直微分，再以這些微分值計算曲率 C 值。

計算細胞特徵點集合 A 內所有 C 值的平均值 μ_c 與標準

差 σ_c 作為特徵值 f_5 與 f_6 。因此每一細胞目前可取得 6 個特徵 $(f_1, f_2, f_3, f_4, f_5, f_6) = (\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_c, \sigma_c)$ 。

(八) 細胞內強度的標準差

前述的特徵萃取作法只考慮細胞區域內的邊緣點，對於細胞紋理的描述不夠全面性。為了健全特徵的全面性，進而減少分類的誤差，考慮細胞內部區域影像的灰階變化特性，增加細胞內強度變化作為特徵。因細胞的外圍區域易受背景切割的影響，例如 TH_{Obsu} 大則細胞區域變小， TH_{Obsu} 小則細胞區域變大。因此將細胞劃分為內與外兩部份，只針對內部細胞區域計算特徵。

細胞劃分的示意圖如圖 9 所示，從細胞外圍起距離寬度 w 內的像素為細胞外緣 *Outer*，剩下區域定為細胞內部 *Inner*，以此將細胞劃分為內部與外部。 w 是外部區域的指定寬度，實驗值設 $w=42$ 。

當在細胞區域取得內部 *Inner* 區域後，計算內部區域所有灰階值的標準差 σ_{gray} 作為特徵值 f_7 。綜合以上特徵萃取方法，對於每一細胞區域 R 共可萃取出 7 種特徵，組成特徵向量 F ，其中

$$F = (f_1, f_2, f_3, f_4, f_5, f_6, f_7)^T = (\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_c, \sigma_c, \sigma_{gray})^T. \quad (12)$$

由於一張抗核抗體螢光影像上不只一個細胞區域，若影像上有 N 個細胞區域，可取得 N 個特徵向量，將這些向量組成特徵集合 $S = \{F_1, F_2, \dots, F_N\}$ 。

(九) 代表特徵向量挑選

由於抗核抗體在螢光取像時部份細胞可能正處於有絲分裂，有絲分裂細胞的螢光形態會不同於正常細胞，導致有絲分裂細胞萃取的特徵也會不同於正常細胞的特徵。為了避免使用有絲分裂細胞的特徵導致影響分類的正確性，本文解

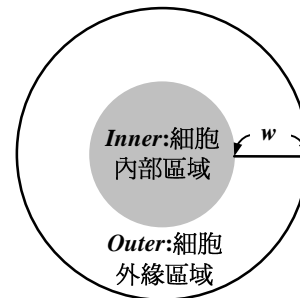


圖 9. 細胞區域劃分為內、外部示意圖， w 是外部區域的指定寬度

決的作法是採取從特徵集合 S 中挑選出一些代表向量，以去除含有絲分裂細胞區域的特徵。首先定義兩特徵向量 F_i 與 F_j 之間的距離 $Dist$ 為

$$Dist(F_i, F_j) = |F_i - F_j|^2 / |F_i + F_j|^2, \quad (13)$$

其中 $|F|$ 代表向量 F 的範數 (norm)，即 $|F| = \sqrt{f_1^2 + \dots + f_7^2}$ ，分母 $|F_i + F_j|^2$ 將距離常規化 (normalize)，使 $0 \leq Dist \leq 1$ 。若 $Dist(F_i, F_j)$ 越大代表向量 F_j 離 F_i 越遠；若 $Dist(F_i, F_j)$ 越小代表 F_j 離 F_i 越近。

計算特徵向量集合 S 內任兩特徵向量的距離，計算所有距離的平均值 μ_{Dist} 和標準差 σ_{Dist} ，代表特徵向量彼此之間的關係。以平均值 μ_{Dist} 和標準差 σ_{Dist} 建立相似度門檻值 $TH_{Dist} = \mu_{Dist} + \sigma_{Dist}$ ，用以劃分每一特徵向量與其它特徵向量的相近度。若特徵向量 F_j 與 F_i 的距離小於門檻值 TH_{Dist} ，表示向量 F_j 與 F_i 相近，則定相近度 $sim(F_i, F_j) = 1$ ；否則定相近度 $sim(F_i, F_j) = 0$ ，即

$$sim(F_i, F_j) = \begin{cases} 1, & \text{if } Dist(F_i, F_j) < TH_{Dist} \\ 0, & \text{otherwise} \end{cases}. \quad (14)$$

針對每一特徵向量 F_i 統計共有多少個特徵向量與其相近，統計量 $sum(F_i)$ 計算如下

$$sum(F_{ji}) = \sum_{F_j \in S, F_j \neq F_{ji}} sim(F_i, F_j) \quad (15)$$

若 $sum(F_i)$ 為 0 代表無特徵向量與向量 F_i 相近；若 $sum(F_i)$ 越大代表越多特徵向量與向量 F_i 相近， F_i 越具有代表性。因此本文挑選 $sum(F_i)$ 大於門檻值 TH_{sum} 的向量 F_i 作為代表性之特徵向量。計算集合 S 中每一向量的 sum 值，計算所有 sum 值的平均值 μ_{sum} 和標準差 σ_{sum} ，設定門檻值 $TH_{sum} = \mu_{sum} + \sigma_{sum}$ 。

藉由門檻值 TH_{sum} 決定特徵向量的保留，當 $sum(F_i) > TH_{sum}$ ，表示向量 F_i 具有代表性，將特徵向量 F_i 加入代表性特徵向量集合 Ω 中，最後以集合 Ω 作為影像的代表性特徵。

三、分類演算法

取得抗核抗體螢光影像的代表性特徵向量集合 Ω 後，再根據特徵向量集合進行分類。首先從四類型 ANA 影像

中，每一類型取 M 張影像當做訓練影像 (training images)，以前述特徵抽取演算法對每一訓練影像抽取特徵向量集合 Ω ，聯集同一類影像的特徵集合 Ω 為該類訓練特徵向量集合 G_k , $k=1, 2, 3, 4$ ，其中 G_1 代表第 1 類均質型的訓練特徵向量集合， G_2 代表第 2 類周邊型的訓練特徵向量集合， G_3 代表第 3 類粗點型的訓練特徵向量集合， G_4 代表第 4 類離散點型的訓練特徵向量集合。

當輸入向量 F 的分類方法說明如下，首先計算向量 F 與各訓練集合 G_k 的距離 $MinDistofG$ ，計算公式如式 (16) 所示。假設第 a 集合具有最小距離，則將向量 F 分類為第 a 類。對向量 F 的分類公式如式 (17) 所示，

$$MinDistofG(F, G_k) = \min_{F_i \in G_k} \{Dist(F, F_i)\}, \quad (16)$$

$$Classify(F) = \arg \min_{k=1,2,3,4} \{MinDistofG(F, G_k)\}, \quad (17)$$

其中 $Classify(F)$ 代表向量 F 的分類結果。

假設輸入測試影像進行分類時，先計算該影像的代表性特徵向量集合 Ω 。由於集合內的特徵向量 F 不只一個，因此考慮每一特徵向量 F 的分類結果代表 1 投票數，每一類代表 1 票箱，因此共有 4 個票箱。對集合內每一向量 F 以式 (17) 進行分類結果後，以分類結果對該類票箱投票，當每一向量均分類投票後，最後統計獲得最大投票數的類型 k ，代表此測試影像被分成第 k 類。例如統計獲得最大投票數為第 3 類，則測試影像被分為第 3 類。對特徵向量集合 Ω 分類投票的公式以 $Vote(\Omega)$ 表示，

$$Vote(\Omega) = \arg \max_{k=1,2,3,4} \left\{ \sum_{F \in \Omega} check(F, k) \right\}, \quad (18)$$

$$check(F, k) = \begin{cases} 1, & \text{if } k = Classify(F) \\ 0, & \text{otherwise} \end{cases}, \quad (19)$$

其中向量 F 是集合 Ω 的元素 (element)， $check(F, k)$ 檢查向量 F 經過 $Classify(F)$ 分類後是否歸為第 k 類，若是的話則設 $check(F, k) = 1$ ；否則設 $check(F, k) = 0$ 。透過 $Vote(\Omega)$ 統計獲得最大投票數的類型結果。

四、實驗結果

實驗使用的抗核抗體螢光影像由台中榮民總醫院免疫風濕科所提供，將病人的血清以 1:80 的比例稀釋，以間接免疫螢光法，透過螢光顯微鏡 (Axiopoint 2, CarlZeiss, Germany)，使用數位影像擷取設備 (E-330 Olympus, Tokyo Japan) 擷取抗核抗體螢光影像 (倍率為 640)，影像為 24 位元的 RGB 彩色影像，三個頻道各以 8-bits 表示色階，影像大小為 3136×2352，約七百萬像素，存成 ORF 檔案為未壓縮的檔案格式。

實驗使用均質型、周邊型、粗點型及離散點型四種類型影像進行測試，各類影像數分別為：均質型 36 張、周邊型 22 張、粗點型 94 張及離散點型 28 張，共 180 張影像。各螢光形態類型隨機抽取 5 張為訓練影像集，共有 20 張訓練影像，餘 160 張為測試影像。表 1 為實驗的分類結果，粗點型和離散點型均完全分類正確，分類正確率皆為 100%；均質型有 1 張影像分錯為粗點型，分類正確率為 97%；周邊型有 1 張影像分錯為均質型，分類正確率為 94%。因 160 張測試影像共有 158 張影像正確分類，系統整體分類正確率為 98.75%。

在文獻 [1, 7, 9, 11, 14] 中的分類正確率不超過 88%，本文所提出的整體分類正確率為 98.75%，證明所提出的 7 個特徵與分類法有效。由於所提的方法是簡單的平均值、標準差、微分、曲率、最大值、最小值、與投票等運算，與文獻中使用決策樹 [7, 9]、多分類器 [14]、與類神經網路分類 [1] 比較，本文所提出的方法較簡單且運算快速。

五、結論

本論文提出一抗核抗體免疫螢光影像特徵萃取與分類之系統，從實驗結果證明所提出特徵抽取及分類的演算法，對於四種 ANA 螢光影像類型的分類，最差的分類正確率都有 94%，系統整體分類正確率高達 98.75%。由於高的分類正確率使系統可進一步作為電腦輔助螢光影像分類，輔助醫療人員判讀 ANA 影像之類別。

實驗結果能得到高的正確率，證明所提出的特徵具有代表性，因特徵具有代表性才能將類別分類成功。另外考慮對各細胞區域進行特徵抽取組成特徵向量集合，再從這些特徵向量中取得代表性特徵向量的作法，如此去除有絲分裂細胞區域之特徵也使正確率提高。

系統所提出的特徵抽取與分類方法均簡單快速易於軟

體實作並可實時操作應用，系統參數中除了面積門檻值 TH_{area} 與外圍寬度 w 兩參數依靠影像大小手動設置實驗值，當取像環境改變時，醫療人員可以經過訓練影像而調整門檻值，使得分類正確率最高。若影像大小固定後，則不需再更改此兩參數；其餘參數值均自動由影像中取得，醫療人員不必嘗試設置系統參數，使系統能穩定操作，有助醫療人員操作系統輔助分類。

由於罹患細點型或核仁型的病人較少，本文只針對均質型、周邊型、粗點型、離散點型四類型進行分類，未來將收集細點型、核仁型、或其它類型的影像，納入系統加以分類實驗，使所提出的分類系統更有公信力與全面性。

本文在決定分類特徵時，針對所提出 7 個特徵均加以組合測試，當此 7 個特徵全部組合採用，所得的分類正確率最高。未來將嘗試找出新的更具代表性的特徵使分類的正確率更高。

誌謝

本研究由大葉大學與台中榮民總醫院共同提供經費補助，計畫編號 TCVGH-DYU-978303，特此致謝。

參考文獻

1. 黃嘉政 (民 97)，應用資料探勘技術於抗核抗體免疫螢光顯影分析，國立嘉義大學資訊管理學系暨碩士班碩士論文。
2. Bradwell, A. R., R. G. Hughes and E. L. Harden (2003) *Atlas of HEp-2 Patterns*, 2nd Ed., Binding Site, UK.
3. Gonzalez, R. C. and R. E. Woods (2001) *Digital Image Processing*, 2nd Ed., Prentice Hall, New Jersey, NJ.
4. Heffernan, M. P., J. H. Do and J. Mehta (2001) Antinuclear antibodies in dermatology. *Seminars in Cutaneous Medicine and Surgery*, 20(1), 2-13.
5. Malladi, R. and J. A. Sethian (1996) Image processing: Flows under min/max curvature and mean curvature. *Graphical models and Image Processing*, 58(2), 127-141.
6. Otsu, N. (1979) A threshold selection method from gray-level histograms. *IEEE Transactions on System, Man, and Cybernetics*, 9(1), 62-66.
7. Perner, P. (2001) Classification of HEp-2 cells using fluorescent image analysis and data mining. *Proceedings of the Second International Symposium on Medical Data Analysis*, Springer-Verlag, New York, NY.

-
8. Perner, P. and H. Perner (2002) Texture classification based on the Boolean model and its application to HEp-2 cells. Proceedings of the 16th International Conference on Pattern Recognition, Quebec, Canada.
 9. Perner, P., H. Perner and B. Muller (2002) Mining knowledge for HEp-2 cell image classification. *Artificial Intelligence in Medicine*, 26, 161-173.
 10. Rigon, A., P. Soda, D. Zennaro, G. Iannello and A. Afeltra (2007) Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose. *Cytometry Part B: Clinical Cytometry*, 72B(6), 472-477.
 11. Sack, U., S. Knoechner, H. Warschkau, U. Pigla, F. Emmrich and M. Kamprad (2003) Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics. *Autoimmunity Review*, 2, 289-304.
 12. Sayood, K. (2006) *Introduction to Data Compression*, 3rd Ed., Morgan Kaufmann, San Francisco, CA.
 13. Smereka, P. (2003) Semi-implicit level set methods for curvature and surface diffusion motion. *Journal of Scientific Computing*, 19(1-3), 439-456.
 14. Soda, P. (2007) Early experiences in the staining pattern classification of HEp-2 slides. Proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems, Marior, Slovenia.
 15. Soda, P. and G. Iannello (2006) A multi-expert system to classify fluorescent intensity in antinuclear autoantibody testing. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, Washington, DC.

收件：98.10.06 修正：99.02.01 接受：99.03.11