

# 基於邊緣變化的二階段子宮頸癌分類

戴紹國

朝陽科技大學&助理教授

sgdia@cyut.edu.tw

莊子慶

朝陽科技大學&碩士生

s10014613@cyut.edu.tw

## 摘要

本研究主要針對陰道鏡的子宮頸影像進行子宮頸癌偵測和分類。我們與其他相關研究不同的地方在於我們發現除了一般常用的判讀特徵之外，病灶邊緣白化的型態是一個可能的特徵。所以我們在分類之前先針對病灶的邊緣的型態進行分群的動作，讓每一個分群的病灶其邊緣有著一致的型態，然後針對每一個分群抽取白化和邊緣灰階對比的程度作為特徵來訓練分類器，進行子宮頸癌的分類。我們使用434張影像來進行實驗，測試分群前後的分類正確率，我們測試了其中一個分群證實分群之後可以讓正確率提升8.18%，此研究之成果可作為醫師診斷癌化等級時的輔助參考，更準確的診斷出癌化等級，使病人接受最適當的醫治，減少因誤判所造成之遺憾。

**關鍵字：**子宮頸癌、陰道鏡、影像切割、邊緣變化、自我組織圖網路

## Abstract

This study is focus on the detection and classification of the cervical cancer by the colposcopic cervical images. We found that the pattern of the edge changes could also be a feature for the interpreter of lesion upon cervix. In this study, we first cluster lesions according to the pattern of the edge changes. Then we train classifier for each cluster with the features of whiteness and gradient of

contour. We collect 434 images for the experiments and test one of the clusters for the precision and compared to the results generated from all images. In the experimental results, we demonstrated that after clustering the precision will be 8.18% higher. Our research can help doctor to improve their diagnosis and the quality of treatment as well.

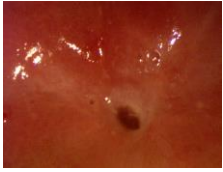
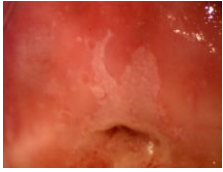
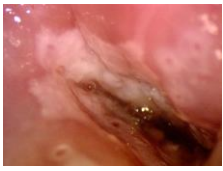

**Keyword:** Cervical cancer, colposcopy, segmentation, self-organizing map

## 1.前言

民國71年起癌症即為國人十大死因首位，而子宮頸癌為我國婦女高發生率、高死亡率之癌症[2]，子宮頸癌所帶來的不只是單方面的痛苦與死亡，更影響到經濟、社會、家庭等等；而子宮頸癌是可以透過檢查來得知是否有病變的徵兆，以早期發現早期治療，減少醫療成本支出[4]。陰道鏡檢查是醫師直接觀察子宮頸的實際情況判斷子宮頸的病變程度，確認是否癌前病變的發展情形[5,11,12,15]。

子宮頸癌初期是由細胞分化不良演變而來的，但還未屬於子宮頸癌，而分化不良可分為輕度、中度和重度，當分化不良過分嚴重時，子宮頸上皮內開始出現鱗狀的薄扁的細胞，慢慢的病灶發展成子宮頸癌，依嚴重程度分為低度病灶(Low SIL)、高度病灶(High SIL)與侵略性子宮頸癌(Invasive Carcinoma)，依上皮內細胞的異常增生的程度可分為 CIN I、

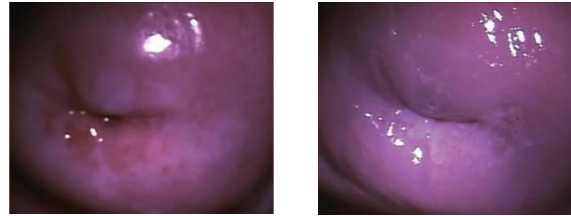
表 1：病灶程度與異常增生程度對應表

癌化程度		影像
LSIL	CIN I	
	CIN II	
HSIL	CIN III	
Invasive Carcinoma	CIS	

CIN II、CIN III 和 CIS，如表 1 所示。

在醫師尚未確認子宮頸是否病變前，醫師會先透過陰道鏡檢查，來彌補靈敏度的不足。陰道鏡檢查方法是使用 3% 到 5% 醋酸塗抹於子宮頸，使子宮頸上皮組織白化，如圖 1 [1] 所示。子宮頸腫瘤接觸到醋酸則會呈現出不透明的白色和可變的瞬態 [7]，藉由觀察醋酸白化上皮組織來預測子宮頸的病變程度 [14]。

在陰道鏡內可看出子宮頸的皮上細胞病變和血管異變的跡象，從這些跡象中得到子宮頸病灶的形態學特徵，進而將癌化程度分類到 Normal、CIN I、CIN II、CIN III 與 CIS。陰道鏡檢查對子宮頸癌白化程度可依三個部分來分類，上皮組織的雪白程度、上皮白化區的邊緣變化和上皮白化區內的血管變化，醫師透過這三種病變程度加以分類癌化程度，由於病變情況沒有客觀的量化標準，所以必須完全依靠醫師的主觀判斷。若是醫師經驗不足，則嚴重



(a) 使用醋酸前影像 (b) 使用醋酸後影像

圖 1：塗抹醋酸前後影像 [1]

缺乏特異性上與靈敏度上的準確性。而資深醫師也平均大概是 80% 的準確性，據估計約有三分之一的重度病徵錯過初期陰道鏡檢查 [6]。而陰道鏡檢查是確定子宮頸是否病變的重要檢查，若病人陰道鏡檢查時被誤診，導致延誤就醫的黃金時刻而造成無法挽回的悲劇。如果能運用電腦影像技術來輔助醫師確認病灶位置，並且提供病灶癌化程度的量化數據與分級結果，將會讓醫師的診斷更加的精準 [1]。

近來，由於影像技術的快速發展，使得醫療影像的自動化輔助診斷出現了很多相關的研究。而在陰道鏡子宮頸癌影像自動化判讀方面，Yang al. 使用 K-means 分群法和確定性退火技術 (SA) 檢測醋酸白色上皮 [16]，Gordon et al. 發現醋酸白色區域確定是提取最高的平均強度集群之間的平滑區域 [9]，Pogue et al. 的研究專注在子宮頸影像上不同區域的度量評估，並表示基於電腦處理過的子宮頸影像可以提供辨識一些組織特徵，提及使用歐拉數可以作為臨床辨識腫瘤的化生，有助於臨床評估 [13]。

我們觀察白化病灶的邊緣變化和其癌化的情形有著一定的關聯性，本研究將針對子宮頸癌的白化區邊緣變化加以量化並且分群，一致化白化區的邊緣變化之後，透再過分類並組合白化區與邊緣變化之特徵，來針對各種癌化級別組合進行最優化的分類。

本研究將分為三大章節，第一章節為前言，介紹子宮頸癌的重要性、影響、病灶程度、異增程度、相關研究；第二章節為研究方法，介紹流程大綱，再介紹每細部流程的目的與做法；第三章節為實驗與結論，驗證每個細部流

程之結果與研究結論。

## 2.研究方法及流程

本研究分成幾個步驟，分述如下：

### 2.1. 子宮頸癌區域切割

我們的切割方法結合自我組織圖網路技術(Self-Organizing Map, SOM)[10]、模糊K-means分群法(Fuzzy k-means clustering)[8]，以分群品質測量進行陰道鏡子宮頸影像的分割。醫師在病患的子宮頸上塗抹醋酸之後，透過陰道鏡對子宮頸檢查，子宮頸病灶區因醋酸而呈現乳白化，對此乳白區域進行影像切割。

1. 首先以自我組織圖網路將特徵向量映射到二維的空間上，它是屬於無監督式學習網路模式的一種。它的基本原理是模仿人類大腦中，具有相似功能的腦細胞會聚集在一起的特性，所發展出來的類神經網路。與其他分群演算法的最大不同在於，SOM 有一個拓樸空間，此拓樸圖用來表達每個輸出值的分布狀況。因此，可以簡化影像的結構，讓分群的計算更加有效率。

2. 接下來使用模糊 K-means 分群法，針對這些神經元進行分群動作，藉由分群品質分析藉由計算群內距離(intra-distance)判斷分群的結果是否恰當[4]，其最適分群數演算法為：

*Algorithm: M-Search*

*Input: {  $d_i; i=4$  to  $m$  } The maximum complete diameter intra-cluster distance of all cluster*

*Output:  $i$ , the best cluster number*

*Step 1  $i$ =initial value*

*Step 2 repeat increase  $i$  for step 3*

*Step 3 until  $d_i < d_4 * \sigma$*

*Step 4 repeat increase  $i$  for step 5*

*Step 5 until  $d_{i+1} > d_i$  and  $d_{i+2} > d_{i+1}$*

*Step 6 output  $i$  as the best cluster number*

3. 在取得神經元分群之後，每個像素根據其相對應神經元分群來指定其分群標籤，由此

取得子宮頸白化的區域。

### 2.2. 邊緣樣式分群

在取得白化區域之後，為了取得白化邊緣的變化，我們使用型態學的擴張(dilate)和侵蝕(erode)來取得邊緣的帶狀區域及其平均灰階度，然後計算相鄰兩個帶狀的斜率，並且4等份量化斜率的角度，進行邊緣變化之編碼，演算法如下：

*Algorithm: Edge coding*

*Input:  $X$ 、 $Y$ ;  $X$ =灰階影像， $Y$ =白化的區域*

*Output:  $C$*

*Step 1. For  $k=\{-8,-7,\dots,8\}, k \neq 0$  repeat step 2 to 6*

*Step 2.  $se_k = Dilate_k(Y)$*

*Step 3  $R_k = X .* (se_k - se_{k+1})$*

*Step 4  $AR_k = avg(R_k)$*

*Step 5  $G_k = atan(AR_{k+1}/AR_k)$*

*Step 6  $G_i = max|G_k|$  where  $i$  is the index of the max. of  $G_k$ .*

*Step 7  $Q = \{G_{i-m}, \dots, G_i, \dots, G_{i+m}\}$  where  $Q$  is the take by figure of  $G$  vector*

*Step 8  $C = floor(Q/45)$  where  $C$  is figure of feature code*

取得邊緣變化之編碼之後 就直接以該編碼當做分群代號，達成分群目的。

### 2.3. 特徵抽取

在分群之後針對每一群進行分類器的訓練，首先是特徵的抽取，我們使用白化程度和邊緣的斜率來辨識病灶的癌化程度，說明如下：

● 白化程度:飽和度特徵

將影像轉成HSV色彩模型， $R_s$ 是切割出來的癌化區域的s plain(飽和度)，然後取前j大值的平均，作為特徵

$S_m = avg(Max_m(R_s))$  where  $m$  is the amount of feature

- 斜率特徵：

$$GX_1 = [G_i]$$

$$GX_2 = [G_{i-1}, G_i, G_{i+1}]$$

$$GX_3 = [G_{i-2}, G_{i-1}, G_i, G_{i+1}, G_{i+2}]$$

$$GX_4 = [G_i, G_{i-1}/G_{i+1}]$$

$$GX_5 = [G_i, G_{i-1}/G_{i+1}, G_{i-2}/G_{i+2}]$$

我們合併 S 特徵與 GX 特徵，即每種 S 特徵與每種 GX 特徵相互組合補助，也考慮到 S 特徵與 GX 特徵可能產生相互干擾的情形，故 S 特徵與 GX 特徵各自組合來測試其區分能力，如下：

$$f_{mn} = [S_m, GX_n]$$

最後分類器的設計，因為每個癌前病變分級對特徵的表現差異很大，所以我們採用決策樹的分類方式，每個分支都使用其合適的特徵來分類。

### 3. 實驗結果

#### 3.1 影像來源

實驗中使用所有影像的子宮頸癌之級別皆由醫師透過陰道鏡，針對病灶的部分拍攝影像，並在影像中標示出病灶區塊，取出病灶區內的子宮頸組織進行切片檢查來診斷其癌前病變的程度，然後給予該病灶癌前病變級別，以確認其正確性。樣本總數量如表 2：

表 2：實驗樣本分佈表

Normal	CIN I	CIN II	CIN III	CIS
64	138	139	99	42

邊緣變化共為 1024 種，但因為收集樣本時間還不夠，所以樣本數量不夠，在邊緣樣式分群之後，各個分群樣本過於稀少，所以只能選擇群內數量最多之類群來實驗，如表 3 所示，我們選擇編號 22322 的群組來做實驗：

表 3：邊緣變化分群數量

邊緣編號	樣本數量
22322	193
23322	105
33322	55
22222	52
32222	19
32322	10

表 4：22322 分群之樣本分佈

Normal	CIN I	CIN II	CIN III	CIS
27	46	54	43	23

#### 3.2 實驗設備

以桌上型電腦(Intel Core CORE i7, 4GB RAM)操作實驗，使用 Matlab(版本為 R2011a)所提供之影像處理工具來進行子宮頸影像的切割與分類。

#### 3.3 實驗結果與討論

於本章節我們將實際驗證我們的方法，使用最適合分類各級別的特徵，測試使用邊緣分群之前後正確率的比較，分類器我們使用 SVM，透過組合特徵，然後將建立最優化決策樹。

首先我們實驗在未使用邊緣變化方法，建立出最優化的決策樹(圖 3)樣本數量(表 2)，在第一層我們使用  $f_{11}$  分類出 Normal，第二層使用  $f_{21}$  分類出 CIN I，第三層使用  $f_{23}$  分類出 CIS，第四層使用  $f_{43}$  分類出 CIN II 與 CIN III，實驗結果(表 5)可看出在第一層與第三層分類時表現比較好，反之第二層與第四層的辨識率比較低。

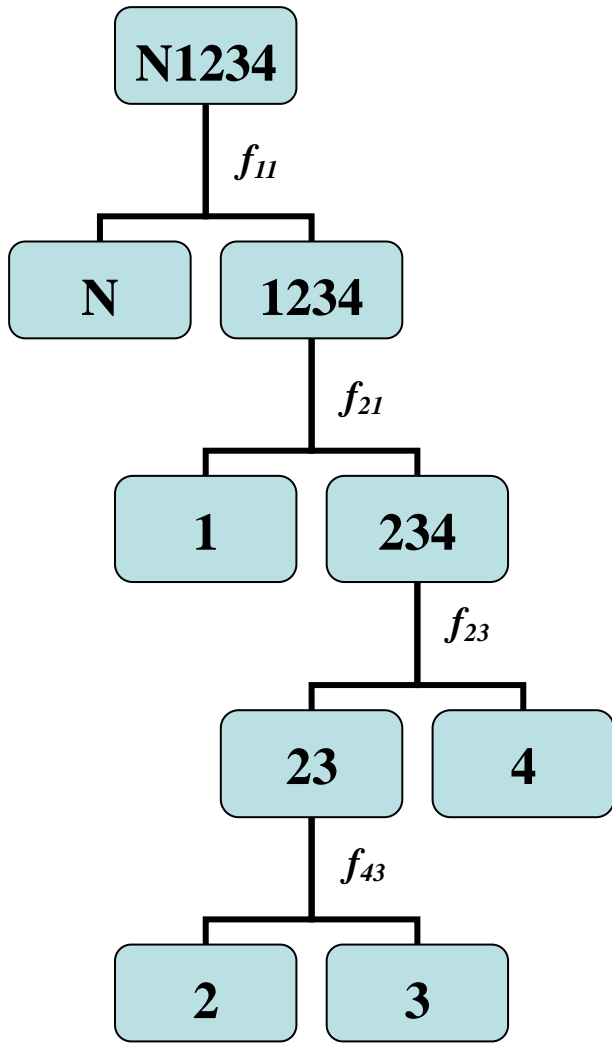


圖 3：未使用邊緣變化之最優化決策樹

表 5：未使用邊緣變化之辨識準確率

第一層	第二層	第三層	第四層	平均
90.46	75.35	98.57	90.34	88.68

然後我們實驗使用邊緣變化方法，建立起最優化的決策樹(圖 4)樣本數量(表 4)，在第一層我們使用  $f_{23}$  分類出 Normal，第二層使用  $f_{53}$  分類出 CIN I，第三層使用  $f_{53}$  分類出 CIS，第四層使用  $f_{12}$  分類出 CIN II 與 CIN III，實驗結果(表 6)可看出在第一層與第三層分類時表現比較好，反之第二層與第四層的辨識率比較低。

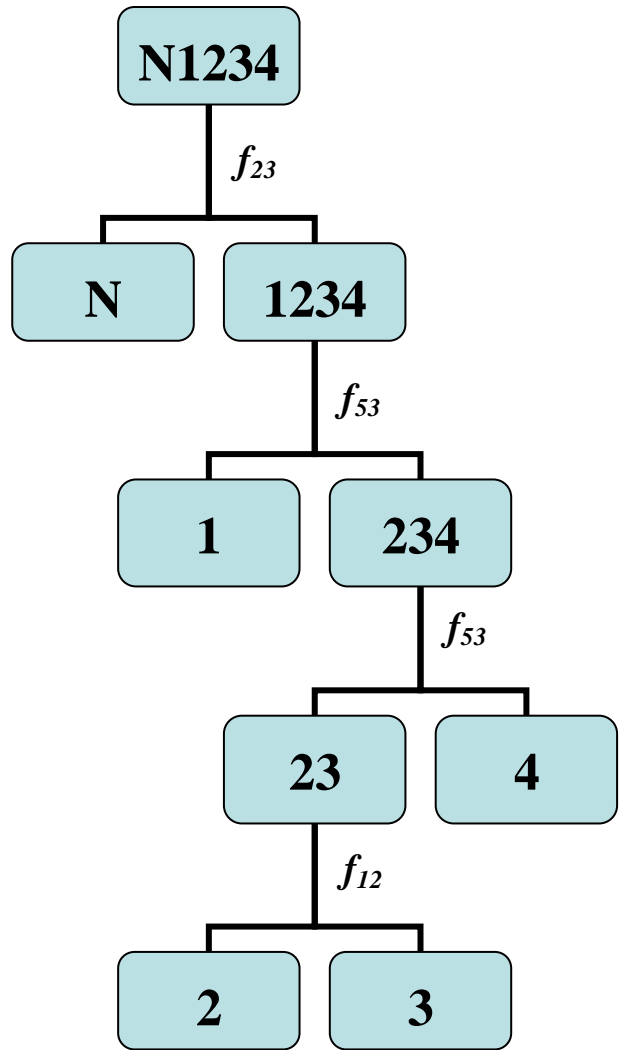


圖 4：使用邊緣變化之最優化決策樹

表 6：未使用邊緣變化之辨識準確率

第一層	第二層	第三層	第四層	平均
97.93	91.57	100	97.94	96.86

由上表 5 與表 6 可看出，雖然在第二層的辨識率雖然只有 75.35% 與 91.57%，辨識率偏低的原因是因為我們取特徵的方法是從人類視覺為出發點，CIN I、CIN II 與 CIN III 的特徵在人類視覺上是非常相似的，故 CIN I、CIN II 與 CIN III 辨識準確率偏低是合理的，但是使用邊緣變化分群後的平均辨識率為 96.86%

與沒使用邊緣變化分群的 88.68% 提升了 8.18 個百分比，而且使用邊緣變化方法後每層的辨識率皆高於未使用邊緣變化方法，因此證實我們的方法是有效的。

#### 4. 結論

本研究主要在使用自動化影像切割與分類，讓子宮頸癌影像可以透過系統自動找出病灶區域，並完成分級動作來幫助醫師達到輔助診斷的功能。本研究結合 Fuzzy k-mean clustering、自我組織圖網路和分群品質測量進行影像的切割，並且加入了邊緣變化的分群，將同類邊緣的子宮頸癌影像分群，進而達到最佳的子宮頸癌病變的分級準確率；使用各級最佳的特徵來分類級別，研究結果顯示分級的正確率達 96.86%。此研究的分級結果主要是可以提供給醫師作為參考之用，亦可提供給經驗不足的醫師當作訓練時的工具。

#### 參考文獻

- [1] 李新澤，**陰道鏡對子宮頸癌的自動化輔助診斷**，朝陽科技大學資訊管理系學位論文，2012。
- [2] 陳美如，”由國外經驗檢視我國子宮頸癌篩檢政策”，**台灣大學衛生政策與管理研究所學位論文**，pp. 1-103，2005。
- [3] 陳雅萍、葉淑惠、龔福財、何美瑤，”子宮頸癌病人症狀困擾、憂鬱及生活品質之相關性研究”，**實證護理**，Vol. 3, No. 2, pp. 89-98, 2007。
- [4] 簡大任、李瑋珠，”台北市有偶婦女接受子宮頸抹片檢查影響因素之探討”，**中華公共衛生雜誌**，Vol. 14, No. 2, pp. 111-128, 1995。
- [5] Cancer Facts & Figures 2006, available at: <http://www.cancer.org>.
- [6] Cox, J. T., Schiffman, M. and Solomon, D., “Prospective follow-up suggests similar risk of subsequent cervical intraepithelial neoplasia grade 2 or 3 among women with cervical intraepithelial neoplasia grade 1 or negative colposcopy and directed biopsy,” **Am. J. Obstet. Gynecol.**, 188, 1406-1412, 2003.
- [7] Ferris, D., Cox, J. T., O’Connor, D. M., Wright, V. C. and Foerster, J., “Modern

Colposcopy Textbook and Atlas, American Society for Colposcopy and Cervical Pathology”, **Dubuque, IA**, 2004.

- [8] Gasch, A. P.; Eisen, “M. B. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering,” **Genome. Biol.** 2002, 3 (11), RESEARCH0059.
- [9] Gordon, S., Zimmerman, G., Long, R., Antani, S., Jeronimo, J. and Greenspan, H., “Content analysis of uterine cervix images: initial step towards content based indexing and retrieval of cervigrams,” **Proc. SPIE**, 6144, 1549-1556, 2006.
- [10] Kohonen, T., “The self-organizing map,” **Dept. of Comput. Sci., Helsinki Univ. of Technol., Espoo**, Vol. 78, No. 9, pp. 1464-1480, 1990.
- [11] Li Wenjing, Venkataraman Sankar, Gustafsson Ulf, Oyama Jody C., Ferris Daron G., Lieberman Rich W., “Using acetowhite opacity index for detecting cervical intraepithelial neoplasia,” **j. Biomed. Opt.**, 14(1), 014020, 2009.
- [12] Park, S. Y., Follen, M., Milbourne, A., Rhodes, H., Malpica, A., Mackinnon, N., MacAulay, C., Markey, M. K. and Richards-Kortum, R., “Automated image analysis of digital colposcopy for the detection of cervical neoplasia,” **J. Biomed. Opt.** 13, 014029-1-014029-10, 2008.
- [13] Pogue, B. W., Mycek, M. A. and Harper, D., “Image analysis for discrimination of cervical neoplasia,” **J. Biomed. Opt.**, 5, 72-82, 2000.
- [14] Reid, R. and Scalzi, P., “Genital warts and cervical cancer. VII. An improved colposcopic index for differentiating benign papillomaviral infections from high-grade cervical intraepithelial neoplasia,” **Am. J. Obstet. Gynecol.**, 153, 611-618, 1985.
- [15] Sankaranarayanan, R., Rajkumar, R., Theresa, R., Esmay, P. O., Mahe, C., Bagyalakshmi, K. R., Thara, S., Frappart, L., Lucas, E., Muwonge, R., Shanthakumari, S., Jeevan, D., Subbarao, T. M., Parkin, D. M. and Cherian, J., “Initial results from a randomized trial of cervical visual screening in rural south India,” **Int. J. Cancer**, 109, 461-467, 2004.
- [16] Yang, S., Guo, J., King, P., Sriraja, Y., Mitra, S., Nutter, B. Ferris, D., Schiffman, M., Jeronimo, J. and Long, R., “A

multi-spectral digital cervigram analyzer in the wavelet domain for early detection of cervical cancer," *Proc. SPIE*, 5370, 1833-1844, 2004.