

基於形態學雙構模型的 子宮頸抹片細胞分割與分類

戴紹國
朝陽科技大學&助理教授
sgdia@cyut.edu.tw

黃俊豪
朝陽科技大學&碩士生
s10114619@gm.cyut.edu.tw

林淑娟
台中榮民總醫院&病理師
sara@vghtc.gov.tw

摘要

子宮頸癌是台灣婦女癌症發生機率的前十位，然而子宮頸癌能夠透過子宮頸抹片檢查早發現早治療，以降低死亡率與發病率，子宮頸抹片檢查是一項容易且便宜的篩檢工具，子宮頸抹片是以人工方式觀看玻片內的細胞，以核質比作為判斷是否為異常細胞的依據，人工方式不但耗時也會發生誤判的情形，因此開發一套自動化的子宮頸細胞圖像分割與分類，可以加速作業時間，而且一個完善的分割方法，可以正確計算出核質比，以利做出正確判斷。本實驗的方法，先以形態學雙構模型讓細胞核與細胞質區域更明顯，再使用k-means分群演算法以及動態門檻值分割目標細胞的細胞質與細胞核，並使用support vector machine (SVM) 分類器可以達到87.22%的正確率，足夠輔助醫生或病理師進行子宮頸癌篩選作業。

關鍵字：子宮頸抹片、影像分割、k-means分群演算法、動態門檻值、SVN分類器

Abstract

In Taiwan, cervical cancer is one of the top ten most common cancers among women. It can be detected in the early stage with some screening tool and reduce the mortality and morbidity by the early treatment. Pap smear is an easy and

inexpensive screening tool for the cervical cancer screening. Whereas, it is time consuming and has no quantitative criteria. Therefore, the development of an automatic detector for the abnormal cervical cell is thus necessary. This paper proposed an effective method, combined with k-means clustering algorithm and a dynamic threshold method based on Dual Reconstruction approach. Then the support vector machine (SVM) classifier is adopted for the discrimination between normal and abnormal cells. In the experimental results, we achieved precision of 87.22% and some examples for the well segmented results.

Keyword: Pap smear, Segmentation, k-means clustering algorithm, Dynamic threshold, support vector machine classifier

1. 前言

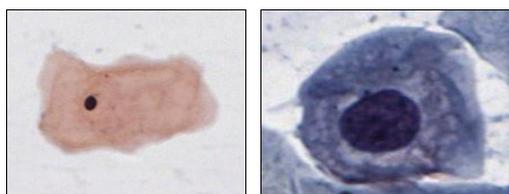
子宮頸癌是台灣婦女癌症發生機率的前十位，然而子宮頸癌能夠透過子宮頸抹片檢查及早發現就能及早治療，以降低死亡率與發病率。在子宮頸癌篩檢的工具中，子宮頸抹片檢查是一項容易且便宜的篩檢工具[9]，子宮頸抹片檢查可以有效的檢驗出癌前病變細胞、細胞異常增生以及癌症，其敏感度與特異性分別為58%和95%[10]，敏感度代表發現疾病細胞的能力，特異性代表辨識正常細胞的能力，子宮頸抹片檢查因為其特異性與低成本的優點，成為

台灣目前子宮頸細胞癌前病變的優先篩檢工具。

子宮頸抹片檢查的步驟為醫生使用陰道擴張器將陰道擴大，再使用木製刮棒或子宮頸刷，將子宮頸周圍的細胞刮下並塗抹於玻片上，迅速浸泡於固定液以保持固定後，送往病理科進行細胞學檢驗。細胞學的診斷為細胞判斷正常與異常的重要依據，人類的細胞構造共有三個部分，分別為細胞核、細胞質與細胞膜，一般不正常細胞是指異常增生細胞與癌細胞，其中異常增生是侵襲性鱗狀細胞癌之前期變化，其主要特徵為細胞核數量增加、形狀增大、顏色加深，所以在判斷是否為異常增生的細胞時，主要是使用細胞核與細胞質的比值，稱為核質比，當核質比上升，則細胞異常程度即上升。目前子宮頸抹片的檢測，是以人工的方式觀測，不但人事成本較高，更會因為人為的疏失，造成錯誤判斷的情況，因此開發自動篩檢系統，在提高準確度與降低耗時，將能帶來利益與效益，其重點在於細胞核與細胞質的圖像分割，因為能完整的分割出細胞核與細胞質的輪廓區域，就能計算出準確的核質比，以利分類正常與異常之細胞。

本研究中只使用核質比分類出正常與異常的子宮頸細胞，在一般病理師的觀測步驟也是先進行核質比的判斷。下面介紹正常子宮頸細胞與異常子宮頸細胞的差異：

圖1(a)為正常子宮頸細胞圖像；圖1(b)為異常子宮頸細胞圖像，兩個圖像的放大倍率雖然不同，但是由圖中的細胞核與細胞質的輪廓區域可以目視出，圖1(b)的細胞核是異常增大



(a)正常細胞 (b)異常細胞

圖1 子宮頸細胞

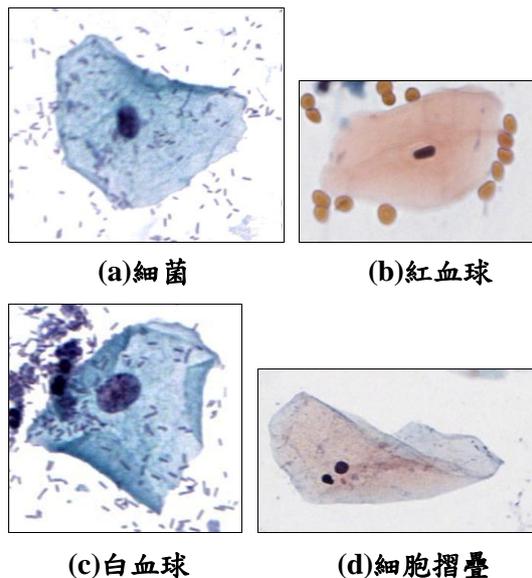


圖2 子宮頸抹片圖像雜訊

的，其細胞核與細胞質的比值也會比圖1(a)來得高，由這樣的核質比判斷，即可做出初步的正常與異常的判斷。

子宮頸抹片細胞圖像並不是每一張都能夠呈現完整的細胞，在抹片中有非常多的雜訊，會影響分割的結果，由圖2(a)到圖2(d)是雜訊的舉例圖像，除了細菌、紅血球、白血球等雜訊，在不同的子宮頸抹片上也可能存在著精子、黏液、破裂細胞等情形，所以在進行圖像分割時，這些可能存在的雜訊都必須排除，例如：當白血球位於細胞內或貼近細胞邊緣時，可能會誤判白血球為細胞核；當紅血球貼近細胞邊緣或是細胞質有摺疊的情況，可能會誤判細胞質正確的區域範圍。

近年來有許多的分割方法應用在子宮頸抹片細胞圖像，分割的目標為細胞的細胞質與細胞核，取得細胞質與細胞核的區域，才能進一步的計算其細胞的核質比，以判斷其細胞為正常細胞或異常細胞。V. Dadeshidze等人[4]使用高倍顯微鏡，將子宮頸抹片細胞圖像擷取出來，判斷異常細胞核與正常細胞核的大小差異，並自動分割出細胞的細胞核區域，達到子宮頸癌細胞的分割與分類。H. M. Cole[2]透過子宮頸病變細胞的細胞核，建構出一套分類方

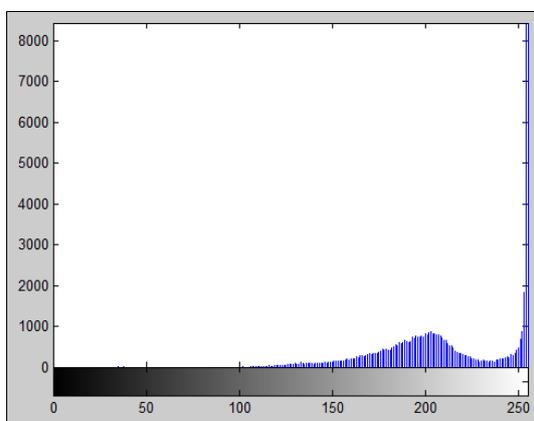
法，使用正常細胞與異常細胞的細胞核大小不同來分類。N. A. Mat-Isa1 等人[7]使用門檻值以及區域成長特徵方法，分別取得細胞質與細胞核的區域，使用分割目標的區域，計算其面積大小之核質比，進行簡單的細胞判斷，但是其門檻值必須使用人工的方式進行設定。Ashidi and Isa[1]使用k-means分群演算法以及區域成長演算法等技術，分割子宮頸抹片細胞的細胞質與細胞核輪廓區域。上述學者所針對的目標都為細胞質與細胞核，尤其是細胞核的面積大小能夠判斷出細胞為正常或異常，而細胞核與細胞質的面積比值就是核質比，所以使用核質比作為特徵值，是可以有效的分類出正常細胞與異常細胞。

2. 研究方法

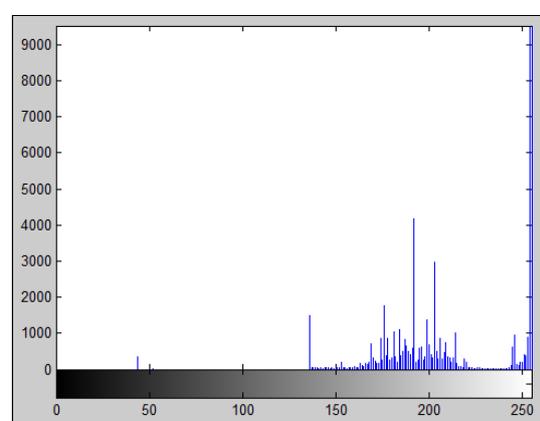
本研究將子宮頸細胞圖像分割分為細胞質分割與細胞核分割兩大步驟，細胞質分割的部分是使用k-means分群演算法[6]分割出細胞質區域，細胞核分割的部分是使用動態門檻值[8]來分割出細胞核區域。分割完後會使用一套計分方法，來判斷哪部分的區域才是我們分割的目標區域。本研究所針對的子宮頸細胞圖像為單一細胞，不考慮重疊細胞的情況。本章節會清楚描述研究方法與流程。

在進行細胞質與細胞核的分割前，都會先

將子宮頸細胞圖像轉為灰階色彩，再使用形態學的雙構模型[5]，目的是將部分的雜訊去除，以利做為後續的目標區域分割，雙構模型的步驟是使用結構元素進行形態學的灰階侵蝕，經過上下兩次的灰階侵蝕並且重建後，即可達到雙構模型的結果。圖3(a)為圖2(a)未使用雙構模型的灰階直方圖，圖3(b)為圖2(a)使用雙構模型後的灰階直方圖，在一般的情況下，在圖像中最深的區域為細胞核，因為細胞核是呈現黑色，由圖3可以發現灰階的分布較為發散，無法決定一個門檻值來代表整個細胞核的區域，但是經過雙構模型的灰階分布圖，從圖3(b)可以發現，灰階的分布會較為整齊，可以很容易的分割出細胞核、細胞質與背景的区域，使用雙構模型後的圖形進行分割，不但能夠去除雜訊，更能準確的分割出目標區域，如果不先使用雙構模型，將無法正確的使用k-means分群演算法或使用動態門檻值切割出完整的區域。對子宮頸細胞圖像進行形態學雙構模型時，分割細胞質與細胞核所使用的結構元素並不相同，因為目標的特徵不相同，在分割細胞質時，會使用較大的結構元素，目的在去除更多的深色色彩，希望圖像最後只剩下背景與細胞質，這樣的圖像在進行k-means分群演算法時，會更容易分類出兩群，也就是背景與細胞質這兩群。而在分割細胞核時，會使用較小的



(a)未使用雙構模型

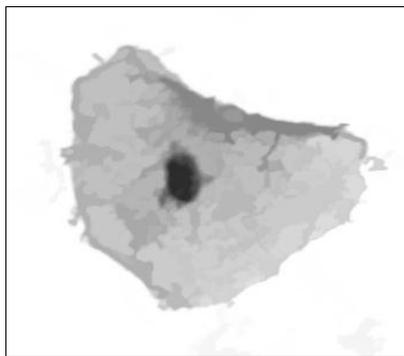


(b)使用雙構模型

圖3 灰階分布圖



(a)



(b)

圖4 使用不同的結構元素所呈現的
雙構模型結果

結構元素，目的與細胞質相反，為了保留較深色的區域，在這部分雖然也會同時保留部分雜訊，例如白血球，但是當細胞質區域被分割出來後，可以肯定的是細胞核一定在細胞質內，依照這樣的定理，可以排除位於背景的雜訊，可以較正確的分割出細胞核的區域。以圖2(a)為例子，假設目前為分割細胞質，使用較大的結構元素進行雙構模型，得到的結果為圖4(a)，從圖中可以發現絕大部分的雜訊已經被清除，中間的細胞核也淡化，由此可證明較深色的白血球也能被淡化，但這部分的分割目標為細胞質，所以只須保留細胞質的區域即可。反之當分割目標為細胞核時，所使用的結構元素就會比較小，以圖2(a)為例子，其雙構模型的結果為圖4(b)，從圖中可以發現使用較小的結構元素進行雙構模型，結果圖像將會保留較為深色的區域，目的就是保留細胞核，進而去除部分雜訊，以利後續的目標分割。

2.1 細胞質分割

在分割目標為細胞質的部分，所使用的方法為k-means分群演算法，其目的為將以使用雙構模型過後的子宮頸細胞圖像，使用k-means分群演算法分類成兩群，一群為細胞質像素，另一群為背景像素，即可達到細胞質分割的效果。k-means分群演算法是由J. B. MacQueen於1967年所提出的演算法，k-means分群演算法的步驟，必須先自訂群集的數量k群，這部分設定為兩群，分別為背景與細胞質，演算法會隨機指派群集中心，隨機在訓練集中指派k個初始中心點，接下來會進行初始群集，計算每一筆紀錄到各個初始中心點的距離，比較到每個中心點的距離，並取最小距離的中心點為群集，在計算每筆紀錄後，就會形成群集間的邊界，產生初始的群集。為了讓分群更加的準確，k-means分群演算法會不斷的進行迭代，其迭代的方式，是將初始群集的每一個群集之中心點，作為新的中心點，再次的計算每一筆紀錄到各個新中心點的距離，取最小距離的中心點為群集，計算完每一筆紀錄後，得到新的群集，也重新取得群集間的新邊界。在不斷迭代的情況下，也必須有收斂的方法，其收斂的方式為，當群集成員與群集中心點不再變動時，就代表已經為最適合的分群結果。一般使用分群演算法，都是為了訓練某個訓練集，以針對某類型的測試集進行分類，但是本研究將分群演算法應用於分割圖像上，因為背景與目標區域的像素強度並不相同，所以可以藉由分群演算法，將背景與目標區域進行分割，以達到本研究的目的。

以圖4(a)為例子，使用k-means分群演算法後的分群結果為圖5，由圖5可以發現k-means分群演算法可以將背景與細胞質區域成功的分群，以達到分割的效果，但是每次分群的結果會產生不同的分群編號，分群編號為1與2，所以細胞質的區域可能會編號為1，也可能編



圖5 使用k-means分群結果

號成2，以圖5為例子，黑色區域為2；白色區域為1，因此本研究設計了一個簡單的細胞質評分方法來決定細胞質區域為哪一個分類群。

本研究的細胞質評分方法為，將分群結果圖像的四個角及中心點的分群結果紀錄下來，將四個角的群集，每一個角加上負1分，最多負4分，並且將中心點的分群加上正3分，最後取分數最高的群集作為目標區域，且將圖像轉換為二值圖像，以利後續的處理。其原因是，本研究所針對的圖像為子宮頸單一細胞圖像，目標細胞的位置應位於圖像中心，所以位於角落的分群結果，將不是分割的目標，以分數的計算，能快速且簡易的決定目標區域。

當分群結束後，可能會將部分的雜訊分類到同一群集，其發生的原因是，較接近細胞的白血球或是位於背景的白血球在使用雙構模型時，還是會留下較淺色的色彩，就有可能被分類到同一群，所以必須排除其可能性，才能正確的分割出目標細胞的細胞質區域。本研究的方法，首先將分群後的圖像進行二值圖像的opening，也就是先侵蝕再擴張，其目的為分割部分與細胞質所連結小區塊，也能將目標細胞質區域處理得更平滑，最後選擇範圍最大的物件作為目標區域，就完成了細胞質的分割，以圖5為例子，最後的分割結果為圖6，圖6分割出正確的細胞質區域。

2.2 細胞核分割



圖6 細胞質區域

在分割目標為細胞核的部分，使用的方法是動態門檻值的方式，從雙構模型後的子宮頸細胞圖像中分割出細胞核的區域。從圖3(a)觀察出，細胞核的灰階分布位置是容易變動的，分布的位置可能出現在0到100的像素強度之間，無法找出一個門檻值來分割整個細胞核的區域，為了解決細胞核的分布變動問題，使用動態門檻值的方式來設計分割方法，能夠滿足分割效果，如果不使用動態門檻值，使用的為靜態門檻值，其結果無法分割灰階分布不同的子宮頸抹片細胞圖像。以圖4(b)為例子，將圖4(b)轉換為灰階直方圖，顯示結果為圖3(b)，從圖3(b)可以發現細胞核、細胞質與背景的灰階分布是可以使用門檻值分割出來的，因為細胞核的像素強度位於0到100之間，其動態門檻值的設定，首先考慮雙構模型後的子宮頸細胞圖像內的最小像素強度值，選擇最小值是因為在細胞圖像內，細胞核應該為最深的像素強度，再加上一個最小值乘上小數值0.6，此數字為重複實驗過後，較能適用於各種情況，因為細胞核的像素強度分布位於這個範圍可以容易設定門檻值，門檻值設定後，將大於門檻值的像素點設為255；小於門檻值的像素點設為0，以達到分割細胞核區域與其他區域。

在前言的部分，文中有提到白血球很容易影響分割的結果，因為白血球與細胞核的像素強度非常相近，幾乎一樣，所以當子宮頸細胞圖像中存在著白血球，使用動態門檻值作為分割的工具，就會影響圖像分割的結果，結果會

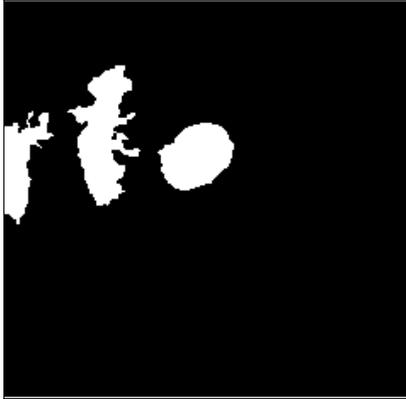


圖9 動態門檻值

將白血球的區域也作為目標區域之一，在圖像中就會存在許多的物件。以圖2(c)為例子，當子宮頸細胞圖像受到白血球的影響，使用動態門檻值的結果為圖9，從圖9可以發現，分割出來的區域也會包含白血球的區域，圖中會產生大於一個的物件。因此本研究設計了一套細胞核評分方法，其目的為決定哪一個物件才是目標細胞核的區域，以得到正確的結果。

本研究在進行細胞核評分方式之前，先將使用門檻值處理後的圖像中的每一個物件進行標記，以圖9為例子，標記的物件數為三件，首先將細胞質分割後的區域與動態門檻值處理的圖像做相乘，因為兩者的圖像都為二值化圖像，背景都為0的情況，相乘過後的結果只會保留兩張圖像交集的部分，其理由為，目標細胞的細胞核一定位於目標細胞的細胞質內，以相乘的方式，就可以排除位於背景的雜訊物件。

移除背景的雜訊後，剩下的物件就必須進行評分，來決定哪一個物件才是分割的目標，本研究的評分方式針對物件的完整性、大小以及與細胞質中心點的距離，這三項評分方法進行評分，使用這三項評分方法的理由為，細胞核會比雜訊來得完整，且來得大，從圖2(c)就可以發現，細胞核是大於周圍的白血球及細菌的，在正常的情形下，細胞核的位置是位於整顆細胞的中心，所以以上述三個評分方式進行不同的權重加分，計算出分數最高的物件，即



圖10 細胞核區域

為目標細胞的細胞核區域，就完成了細胞核之分割。圖10為圖9進行評分後的結果，對照圖2(c)可以發現，以本研究的評分方式，是可以有效的在多物件的情形下，找出正確目標的細胞核分割區域，排除門檻值因為像素強度相似的情形，而無法正確的分割。

3. 實驗

3.1 影像來源

本研究所使用的實驗圖像，提供來源取自台中榮民總醫院病理部。影像的取得過程是使用電腦輔助掃描子宮頸抹片，並使用10倍率的接目鏡與40倍率的接物鏡，進行子宮頸細胞圖像的擷取，擷取的圖像以單一細胞為目標，目標細胞應位於圖像的中心。在本實驗中，共取得321張子宮頸抹片細胞圖像，其中包含正常細胞200張細胞圖像、異常細胞121張細胞圖像。

3.2 實驗設備

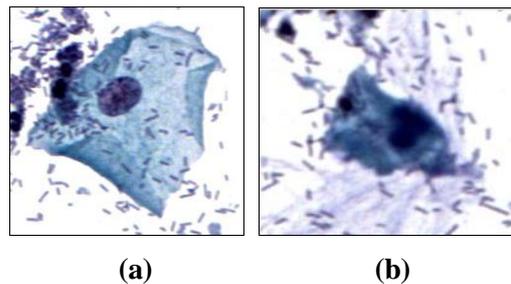
使用個人電腦(Intel(R) Core(TW) i7-2600 3.40GHz CPU, 4GB RAM)進行實作，軟體使用MATLAB(版本為R2012a)，取用軟體中提供的影像處理工具，撰寫本篇所提供的子宮頸抹片細胞圖像分割方法。

3.3 實驗結果與討論

本實驗是使用核質比作為特徵值，來進行分類器的訓練與測試，以分類出正常細胞與異常細胞，其中核質比為使用第二章節所提出的研究方法，將分割出來的細胞核區域與細胞質區域，計算面積後進行相除，即可得到此目標細胞的核質比，使用分類器進行訓練與測試，能得到分類的正確率，正確率越高，越能代表使用此分割方法計算的核質比，能夠正確分類正常細胞與異常細胞。

3.3.1 分割方法比較

除了由本篇所提出的分割方法之外，在實驗的部分也加入了兩項分割方法進行比較，分別是單純使用門檻值進行分割，以及使用雙構模型加上動態門檻值進行分割，以圖11中的兩個圖像進行分割比較，其圖11(a)分割結果為圖12，圖11(b)分割結果為圖13。



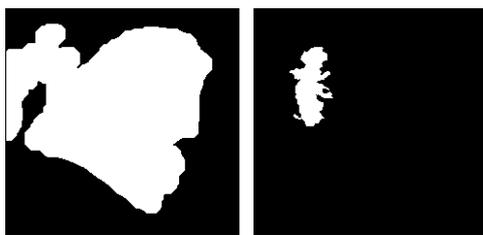
(a) (b)

圖11 欲分割細胞圖像

由圖12與圖13的分割結果，可以很明顯的發現細胞核與細胞質因為不同的分割方法，而有不同的表現。單純使用門檻值所分割的結果，可以明顯的發現，分割結果會包含非常多的雜訊，完全無法分辨出那個區域才是細胞質與細胞核的區域，所以可以證明使用形態學的雙構模型可以有效的去除雜訊，更重要的能夠將灰階分布更整齊，就可以更完整的分割出目標區域。而使用雙構模型與動態門檻值的方法，在細胞質的部分，其分割結果沒辦法分割出正確



(a)本研究的分割方法



(b)雙構模型與動態門檻值

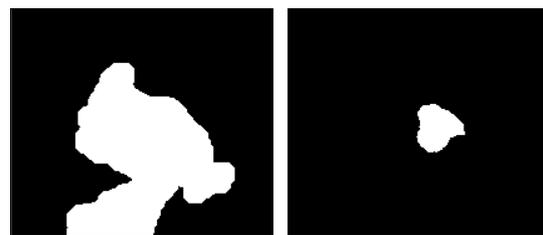


(c)門檻值

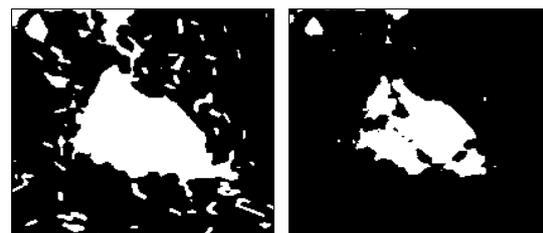
圖12 由圖11(a)之細胞質與細胞核分割結果



(a)本研究的分割方法



(b)雙構模型與動態門檻值



(c)門檻值

圖13 由圖11(b)之細胞質與細胞核分割結果

的細胞質區域，分割細胞質區域的結果是使用k-means分群演算法較為正確，而細胞核的部分，因為沒有使用評分方法，所以會將其他雜訊物件判斷為目標細胞的細胞核。因此使用目視的最佳分割結果為，本研究所提出的分割方法，本分割方法使用雙構模型排除了子宮頸抹片內可能存在的雜訊，並將灰階分布的情況達到整齊，以及使用了目標區域的評分方法，才能夠正確的分割出細胞核區域與細胞質區域，才能夠進行正確的核質比的計算，以利後續的分類器分類，接下的實驗，將使用分類器進行正確率的測量，以證明分割的結果與分類的結果息息相關。

3.3.2 分類器

本實驗使用了兩種分類器進行分類，分別為support vector machine(SVM)[11]分類器與k-nearest neighbor(k-NN)[3]分類器。個別使用leave one out的方式進行訓練與測試，所以一共使用了兩種分類方式來檢測正確率。除了實驗本研究方法的分類結果，也使用了另外上述兩個分割方法的分類結果來進行比較，比較不同的分割方法的核質比，所分類的正確率。

3.3.3 分類結果

表1為各分割方法與各分類器之分類正確率，由表1中可以發現本研究方法在各項分類器的正確率都是高於另外兩項分割方法，由數據可以證明，一項良好且穩定的分割方法所提供的特徵值，能夠影響分類的結果，可以有效的提高分類器的正確率。

表1 各分割方法與各分類器之分類正確率

| | 本研究方法 | 雙構模型與動態門檻值 | 門檻值 |
|------|--------|------------|--------|
| SVM | 87.22% | 86.91% | 68.22% |
| k-NN | 86.29% | 83.48% | 69.47% |

4. 結論

使用電腦輔助自動化的進行細胞的正常與異常判斷，可以增加作業效率，也降低人為誤判的結果，判斷的依據是使用病理師最常使用的核質比進行檢測，其結果也能輔助醫生做出更精確的診斷。本研究所提出的分割方法，是將細胞質分割與細胞核分割分開處理，在分割之前，使用形態學雙構模型，將子宮頸細胞圖像的雜訊排除，並重建出較為整齊的灰階分布狀態，讓分割區域更為明顯，在分割細胞質的部分使用k-means分群演算法，可以比動態門檻值有更好的分割效果，而分割細胞核的部分使用動態門檻值，並加入了評分方式，不會因為白血球等雜訊的影響，而分割錯誤的目標，分割出正確的目標區域，才能計算正確的核質比，利於分類器的分類，使用SVM分類器可以達到87.22%的正確率。本研究對於重疊細胞圖像、雙核細胞圖像沒辦法進行正確的分割，目前只能針對單一細胞進行分割，本研究的細胞質分割方法並不能非常準確的將其細胞質輪廓分割出來，期望未來可以使用輪廓偵測或更合適的分割方法，來解決尚未解決的問題。

參考文獻

- [1] Ashidi, N. and Isa, M., "Automated edge detection technique for pap smear images using moving k-means clustering and modified seed based region growing algorithm," *International Journal of the Computer*, Vol. 13, No. 3, pp. 45-59, 2005.
- [2] Cole, H. M., *Human Papillomavirus DNA Testing in the Management of Cervical Neoplasia*, JAMA, pp. 2975-2981, 1993.
- [3] Denoeux, T., "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 25, No. 5, pp. 804-813, 1995.
- [4] Dadeshidze, V., Olsson, L. J., and Domanik, R. A., "Segmentation of nuclear images in automated cervical cancer screening," *Proceedings of the International Society for Optical Engineering*, Vol. 2622, pp. 723-727, 1995.

- [5] Haralick, R. M., Sternberg, S. R., and Zhuang, X., "Image analysis using mathematical morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 4, pp. 532-550, 1987.
- [6] MacQueen, J., "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-297, 1967.
- [7] Mat-Isa, N. A., Mashor, M. Y., and Othman, N. H., "Seeded region growing features extraction algorithm; its potential use in improving screening for cervical cancer," *International Journal of the Computer, the Internet and Management*, Vol. 13, No. 1, pp. 61-70, 2005.
- [8] Otsu, N., "A threshold selection method from gray-level histograms," *Automatica*, Vol. 11, No. 285-296, pp. 23-27, 1975.
- [9] Paley, P. J., "Screening for the major malignancies affecting women: current guidelines," *American Journal of Obstetrics and Gynecology*, Vol. 184, No. 5, pp. 1021-1030, 2001.
- [10] 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 noplasia," *J. Biomed. Opt.*, Vol. 13, No. 1, pp. 014029-014029-10, 2008.
- [11] Vapnik, V., *The Nature of Statistical Learning Theory*, springer, 1999.