

科技部補助

大專學生研究計畫研究成果報告

* ***** *
* 計畫 : 結合多紋理特徵融合與 Fuzzy Adaboost 之乳房鈣化點 *
* 名稱 : 良劣分析系統 *
* ***** *

執行計畫學生： 吳怡儒
學生計畫編號： MOST 104-2815-C-040-057-E
研究期間： 104年07月01日至105年02月28日止，計8個月
指導教授： 秦群立

處理方式： 本計畫涉及專利或其他智慧財產權，2年後可公開查詢

執行單位： 中山醫學大學醫學資訊學系

中華民國 105年03月30日

二、內容：

(一)摘要

目前台灣乳癌是女性好發癌症的第一位，死亡率為第四位。國內婦女隨著生活型態及飲食習慣西化的改變，台灣乳癌患者一方面平均年齡逐漸年長。根據統計，民國93至98年間，每年約20萬人篩檢，9.4%女性須回診，因此發現的乳癌患者每年約200例，政府致力預防醫學用心值得肯定。篩檢工具除可檢出早期癌症，也會檢查出良性病灶及無法立即明確診斷的病灶。以乳房攝影篩檢為例，只有千分之四最後診斷為乳癌，其餘30%至40%為良性鈣化點或良性腫瘤等。

分辨良性或惡性乳房鈣化，可從乳房攝影(Mammography)的X光片中鈣化點的形狀、分佈情形及是否合併鈣化點周圍乳腺組織密度變化等多項因素加以分析。因此本計畫的目的是利用GLCM以特徵抽取方法，再結3種威斯康辛州乳腺癌資料集(Wisconsin Breast Cancer Database)的特徵，並使用我們提出之特徵融合的方法，再結合SONFIN Adaboot分類演算法提出一套應用於乳房鈣化點良劣分析之系統；選用SONFIN Adaboot為本計畫之分類演算法，是由於SONFIN是以模糊集合(fuzzy set)為基礎架構之演算法，分類效能較單純Fuzzy更為精確，此系統能夠更準確且更自動化的分析鈣化點的良劣。過程中我們建構出一個自動化分析的架構，利用乳房鈣化點的切割、特徵抽取、特徵融合、SONFIN Adaboot。經由這個簡單且更快速的自動化分析架構，使得這項研究突破以往的限制，並且能保有手動選擇鈣化點的準確性，則是若模糊理論若運用於此則可分類出良劣之程度，例如：該乳房鈣化點其為惡性鈣化點的程度為何，因此本計畫的分析結果就不單單只是良與劣的分類，而有程度上的區別，使得此系統能更方便及有效率的協助醫生做醫學相關研究。

(二)研究動機與研究問題

乳癌之所以可能會有密集的微小鈣化點，主要是因為乳癌組織生長時，細胞增加的速度快，部分搶不到養分的乳癌細胞就會產生壞死、鈣化現象，一般多是位於中心處的組織較容易缺氧而壞死，導致鈣化而聚集一處；又因為鈣化的時間先後不同而出現外觀大小不規則的鈣化點。

乳房鈣化是乳癌的機會僅一成五，而乳癌也不一定都會產生鈣化點，如圖 1 所示，約僅有 30~40%的乳癌患者會發生微小鈣化，但奇妙的是，許多摸不到的腫塊，經由乳房影像學檢驗而查出有微小鈣化點的乳癌，大多屬於零期的原位癌，因此乳房微小鈣化點可說是乳癌早期發現的重要指標之一。

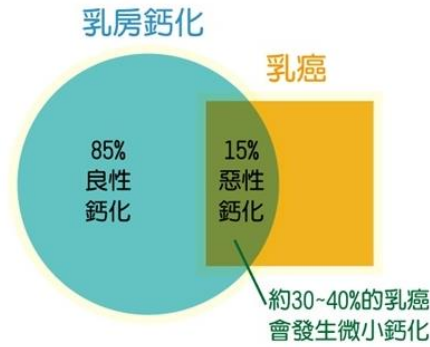


圖 1、乳房鈣化點良劣與乳癌間關係之示意圖

由於一般「良性乳房鈣化點」，形狀多為圓形、爆米花狀、茶杯形，分佈上也比較均勻或對稱；「惡性乳房鈣化點」則常具有不規則的外型，如呈現分岔形狀、線形等，在分佈上常呈現成簇聚集或延著乳管形成線狀分佈，有時可能合併出現鈣化點附近的乳腺產生密度變化。

因此本計劃的目的是利用這些鈣化點的特徵，經過我們所提出之特徵融合方法與 SONFIN Adaboot 的訓練，去分辨其所選取的鈣化點為良性亦或是惡性的鈣化點之程度，最後將結果提供醫生參考與做醫學上的相關研究。

(三)文獻回顧與探討

近幾年來，數位女性名流罹患乳癌的新聞，使得女性同胞感受到乳癌魅影的惶恐，根據衛生署統計，乳癌發生率快速上升。為了能早期發現、早期治療，提高乳癌的治癒率，定期乳癌篩檢是不可或缺的重要方法。其中，乳房攝影[1]是最廣為人使用的乳癌篩檢工具之一。在乳房攝影中，X光可以穿透的部分會呈現灰黑色，X光無法穿透的部分，會呈現白色。我們可以利用乳房攝影，來尋找乳房中的腫瘤及惡性鈣化。但如果乳腺組織較緻密，乳房攝影片子看起來就是白白一片，要在其中找同樣白白的腫塊，相對上較困難。年紀大的婦女，隨著年齡增長，乳腺退化，黑色的脂肪組織會慢慢取代白色的乳腺，腫塊也會更易辨認，也因此，乳房攝影比較適合年紀大的婦女。

談到偵測鈣化[2]則不受此限制，因為乳房鈣化點[3-4]比乳腺更白。但是，並非所有的鈣化都有問題。鈣化的原因非常多，絕大多數都屬良性，據統計 50 歲以上的婦女，約有 50% 會有或多或少的良性鈣化。根據臨床統計，高達 85% 的乳房鈣化屬於良性，僅有 15% 的微小鈣化是乳癌的表現，醫師必須根據鈣化點的形狀、大小、聚集情況來判斷是否就是惡性腫瘤所導致。若經影像學甚至是病理切片確診為惡性，就必須針對病灶進行切除等處理；如果是良性鈣化，有時候根本不需要治療，因此當發現有鈣化點時不用太過悲觀，先諮詢專業醫師判斷是否需進一步檢查並定期追蹤即可。

惡性鈣化點有其特殊型態，分佈上大多是成團聚集，型態上較不規則、大小不一、呈線狀或樹枝分岔狀，每一鈣化點小於 0.5 毫米，又稱微小鈣化。這種成團聚集、型態不一的微小鈣化，非常具有診斷價值，它可以幫助我們偵測出許多

早期的乳癌及原位癌。然而，有部分的微小鈣化型態，無法被明確判斷為良性或惡性，此時，便需作切片化驗。

乳癌之所以可能會有密集的微小鈣化點，主要是因為乳癌組織生長時，細胞增加的速度快，部分搶不到養分的乳癌細胞就會產生壞死、鈣化現象，一般多是位於中心處的組織較容易缺氧而壞死，導致鈣化而聚集一處；又因為鈣化的時間先後不同而出現外觀大小不規則的鈣化點。

綜合上面的回顧，本計畫將利用 GLCM[5-7]的紋理特徵抽取方法進行乳房鈣化點影像特徵的抽抽取，再結 3 種威斯康辛州乳腺癌資料集(Wisconsin Breast Cancer Database)的特徵，並參考多篇特徵融合[8-12]的文獻做為我們提出的特徵融合方法之參考基礎，並結合 SONFIN Adaboost 進行乳房鈣化點良劣之分析。

(四)研究方法及步驟

本計畫進行自動化乳房鈣化點良劣分析其流程如圖2所示，當一張乳房攝影進入系統後，會先做影像處理，並手動切割出乳房鈣化點區域，接下來將所選取的乳房鈣化點區域進行GLCM的紋理特徵抽取，接著利用特徵正規化的方法將所抽取的特徵做值域的統一，在第四步驟，將正規劃後的GLCM特徵與Wisconsin Breast Cancer Database的特徵進行融合，我們參考多項文獻中特徵融合方法提出我們的方法進行融合，最後利用SONFIN Adaboost來辨識其鈣化點為良性或是惡性，而結果部份我們將與合作醫院的醫師所分類出的手動結果進行驗證，接下來會依序作詳細的介紹。

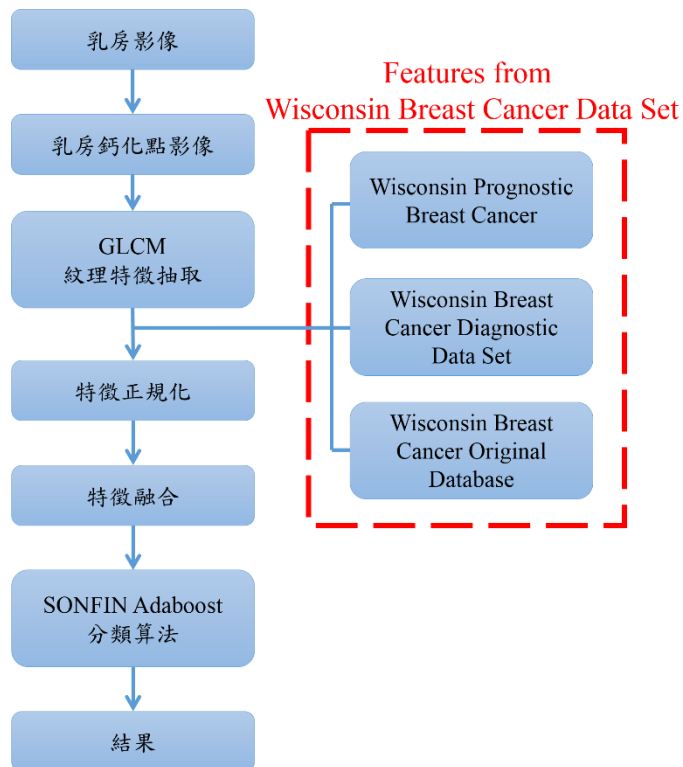


圖2、乳房鈣化點良劣分析流程圖

就上述的流程，研究方法的步驟將分述如下：

一、乳房鈣化點影像：

在乳房鈣化點影像擷取的部分我們需要將原乳房攝影之影像，如圖 3(a)所示，手動選取其鈣化點之影像進行特徵的抽取，由於如圖 3(b)與(c)所示，我們可以得知乳房鈣化點可以大略分為單一顆乳房鈣化點與多顆重疊之乳房鈣化點，但他們及其相似，且多顆重疊之乳房鈣化點不利於切割，因此目前本計畫將這兩類分開進行乳房鈣化點的良劣的分析。

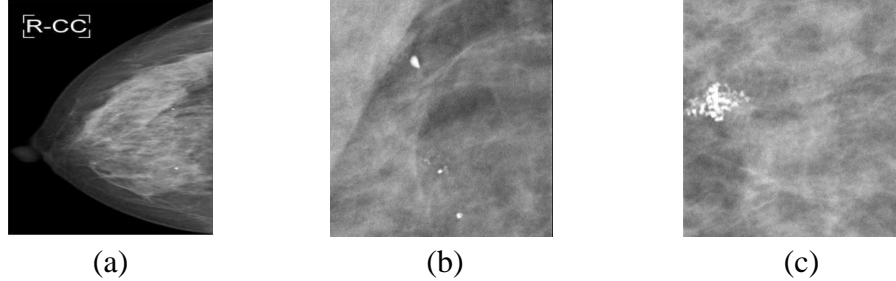


圖3、(a)為乳房攝影之影像，(b)單一顆乳房鈣化點，和(c)多顆重疊乳房鈣化點

二、紋理特徵抽取：

再取得切割後之乳房鈣化點影像後，我們將進行 GLCM 的紋理特徵抽取來擷取特徵向量，並逐一紀錄下來，做為後續特徵融合的基礎。

本計畫所使用的實驗特徵之一為 GLCM 灰階共生矩陣，其為圖像空間中相隔某距離的兩像素之間會存在一定的灰度關係，即圖像中灰度的空間相關特性。灰度共生矩陣即是一種通過研究灰階度的空間相關特性來描述紋理的常用特徵。藉由 GLCM 特徵來辨別兩類影像的依據。GLCM 的量化方法分別有以下五種特徵：對比度、相關性、能量值、同質性和熵值。其計算的方式如下：

$$\text{對比度(Contrast)} : \sum_{i,j} |i - j|^2 p(i, j) \quad (1)$$

$$\text{相關性(Corrlation)} : \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j} \quad (2)$$

$$\text{能量值(Energy)} : \sum_{i,j} p(i, j)^2 \quad (3)$$

$$\text{同質性(Homogeneity)} : \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad (4)$$

$$\text{熵值(Entropy)} : \sum_{i,j} p(i, j) \log p(i, j) \quad (5)$$

其中 $p(i, j)$ 為 GLCM 之灰階共生矩陣， μ_i 為平均 i 值， μ_j 為平均 j 值。

威斯康辛州乳腺癌資料集(Wisconsin Breast Cancer Data Set)

1. 威斯康辛州乳腺癌預後資料集(Wisconsin Prognostic Breast Cancer)：

總共提供了 194 筆案例資料，每筆資料有 34 項特徵，其中良性 148 筆 (76.3%)、惡性 46 筆 (23.7%)。包含十一至三十一項不等的特徵及一個類別，分別為:ID 編號(ID number)、半徑(Radius)、質地(Texture)、周長(Perimeter)、面積(Area)、平滑度(Smoothness)、密實度(Compactness)、凹性(Concavity)、凹點數量(Concave Points)、對稱性(Symmetry)、分形維數(Fractal Dimension)以及類別: 復發/無(Class : recur / nonrecur)。此資料

集由於年代較早且資料筆數較少，而且是針對乳癌是否復發為其類別，因此所得之特徵結果以本計畫使用之角度則皆為惡性。

2. **威斯康辛州乳腺癌診斷資料集(Wisconsin Breast Cancer Diagnostic Data Set)**: 總共提供了 569 筆案例資料，每筆資料有 32 項特徵，其中良性 359 筆 (63.1%)、惡性 210 筆 (36.9%)。包含十至三十項不等的特徵及一個類別，分別為:ID 編號(ID number)、半徑(Radius)、質地(Texture)、周長(Perimeter)、面積(Area)、平滑度(Smoothness)、密實度(Compactness)、凹性(Concavity)、凹點數量(Concave Points)、對稱性(Symmetry)、分形維數(Fractal Dimension)以及診斷結果:良/惡(Diagnosis : benign / malignant)。此資料集為醫師診斷後所判定之數據，與威斯康辛州乳腺癌原始資料集(Wisconsin Breast Cancer Original Database)不同在於醫生有進行篩選及數據的調整。
3. **威斯康辛州乳腺癌原始資料集(Wisconsin Breast Cancer Original Database)**: 總共提供了 699 筆案例資料，其中良性 458 筆 (65.5%)、惡性 241 筆 (34.5%)。其中包含十項特徵及一個類別，分別為:ID 編號(Sample code number)、叢生厚度(Clump Thickness)、細胞大小均勻性(Uniformity of Cell Size)、細胞形狀的一致性(Uniformity of Cell Shape)、邊際附著力(Marginal Adhesion)、單上皮細胞尺寸(Single Epithelial Cell Size)、裸核(Bare Nuclei)、布蘭德染色質(Bland Chromatin)、正常核仁(Normal Nucleoli)、有絲分裂(Mitoses)以及類別:良/惡(Class : benign / malignant)。此資料集為病理測得之所有原始數據及資料。

三、特徵正規化：

在得到上一個步驟所抽取出的特徵後，由於特徵的單位主要可分為連續的數值或名目，當其單位為數值時，則各個特徵之數值範圍大小可能不同。若直接使用各特徵之原值進行特徵融合時，則結果易因數值較大的特徵而有偏差，亦即數值範圍較大特徵易支配數值範圍較小特徵的影響。例如GLCM之數值範圍較Wisconsin Breast Cancer Data Set之數值範圍大，則分類結果易受GLCM之影響。在此，本計畫在此階段使用的方法為極值正規化(min-max normalization)，其適合用於需要將資料規範在某一個指定範圍內的情況因此選用之。假設欲將特徵A之數值範圍轉換至 $[new_min_A, new_max_A]$ ，其中 min_A 為特徵A轉換前最小值， max_A 轉換後最大值，若 v' 為數值 v 經正規化轉換的值，則方法如下：

$$v' = \frac{v - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A \quad (11)$$

經正規化轉換後之範圍為 $[-1,1]$ ，轉換後之數值在特徵融合時各特徵之權重 (weight) 將均等，所建立的方法亦不會因此產生偏差。

四、特徵融合：

在這個階段，本計畫以目前使用之GLCM的紋理特徵抽取並以完成特徵正規化後的結果，與三項威斯康辛州乳腺癌資料集(Wisconsin Breast Cancer

Data Set)的進行特徵融合後再交與後續的SONFIN Adaboot分類演算法做訓練。

由於乳房鈣化點的良劣與其形狀、分佈、密度有密切的關係，因此本計畫選用全部的特徵，並將性質相近或相同的特徵進行融合，以GLCM及Wisconsin Breast Cancer Diagnostic Data Set(WBCDDS)為例，由於GLCM之能量值(Energy)為度量其紋理之一致性，同質性(Homogeneity)為度量影像的均調性，而與WBCDDS之質地(Texture)、平滑度(Smoothness)相似，都與其紋理之灰階值分佈均勻與否有關；但皆與GLCM之對比度(Contrast)用來衡量影像中對比的強烈程度以及WBCDDS之凹性(Concavity)、凹點數量(Concave Points)來看其紋理之灰階值差異程度恰恰相反。再者兩種紋理特徵抽取之特徵均有與紋理雜亂程度有關的特徵，分別為GLCM之熵值(Entropy)與WBCDDS之分形維數(Fractal Dimension)。而本計畫採取之融合方法為平均值以及相加後特徵正規化成原本的值域兩種融合方式，而得多種新的融合特徵值，且正規化轉換後之範圍為亦為[-1,1]，如下所示：

$$a = \text{GLCM之能量值(Energy)} + \text{GLCM之同質性(Homogeneity)} \\ + \text{WBCDDS之質地(Texture)} \\ + \text{WBCDDS之平滑度(Smoothness)}$$

$$b = \text{GLCM之對比度(Contrast)} + \text{WBCDDS之凹性(Concavity)} \\ + \text{WBCDDS之凹點數量(Concave Points)}$$

$$c = \text{GLCM之熵值(Entropy)} \\ + \text{WBCDDS之分形維數(Fractal Dimension)}$$

$$f_1 = \frac{a}{4}$$

$$f_1' = a|_{[-1,1]}$$

$$f_2 = \frac{b}{2}$$

$$f_2' = b|_{[-1,1]}$$

$$f_3 = \frac{c}{2}$$

$$f_3' = c|_{[-1,1]}$$

以及假設另外三種我們選擇獨立使用的特徵，如下所示：

$$f_4 = \text{GLCM之相關性(Corrlation)}$$

$$f_5 = \text{Law's Mask之平均值(Mean)}$$

$$f_6 = \text{Law's Mask之偏度(Skewness)}$$

則本計畫所提出之八種特徵融合後之特徵向量，如下所示：

$$F_1 [f_1, f_2, f_3, f_4, f_5, f_6]$$

$$F_2 [f_1, f_2', f_3, f_4, f_5, f_6]$$

$$F_3 [f_1, f_2, f_3', f_4, f_5, f_6]$$

$$F_4 [f_1, f_2', f_3', f_4, f_5, f_6]$$

$$F_5 [f_1', f_2, f_3, f_4, f_5, f_6]$$

$$F_6 [f_1', f_2', f_3, f_4, f_5, f_6]$$

$$F_7 [f_1', f_2, f_3', f_4, f_5, f_6]$$

$$F_8 [f_1', f_2', f_3', f_4, f_5, f_6]$$

並分別進行SONFIN Adaboot分類演算法的訓練與後續的實驗來證實本計畫所提出之特徵融合方式何種特徵向量最適合乳房鈣化點的良劣分析，且更為精準與良善。

五、SONFIN Adaboot分類演算法：

為了要驗證我們所提出的特徵融合方法是否可以有效的分析乳房鈣化點的良劣，因此本計畫選擇以SONFIN Adaboot分類演算法做為分類器，是由於SONFIN是以模糊集合(fuzzy set)為基礎，其基本精神是接受模糊性現象存在的事實，而以處理概念模糊不確定的事物為其研究目標。

AdaBoost[12-14]方法則是一種迭代演算法，在每一輪中加入一個新的弱分類器，該階段分類器會判斷出其效能，而該階段分錯的樣本會被用來訓練並調整下輪的分類器，同時每次計算錯誤率和結合係數並持續迭代，直到達到某個預定的足夠小的錯誤率。每一個訓練樣本都被賦予一個權重，表明它被某個分類器選入訓練集的機率。如果某個樣本已經被準確地分類，那麼在構造下一個訓練集中，它被選中的機率就被降低；相反，若某個樣本沒有被準確地分類，那它的權重就會被提高。

因此本計畫在AdaBoost演算法中所作迭代的弱分類器即是SONFIN。首先作樣本權重的初始化建置，再來到AdaBoost階段作運行，在AdaBoost演算法中所作迭代的弱分類器即是SONFIN所產生出的模糊推論集合。在每一次迭代中即呼叫一個SONFIN作一次執行得到一個弱分類器，並判斷出其效能，來為下次訓練模式的架構作調整，同時每次計算錯誤率和結合係數並繼續下來，直到停止條件成立為止。並將所有SONFIN訓練模式產生出的模糊集合輸出組合成一最終分類器。參考之方法如下所示：

$$H(x) = \text{sign}\left(\sum_{k=1}^T \alpha_k h_k(x)\right) \quad (12)$$

其中 $H(x)$ 為最終得出之強分類器， T 為總共迭代次數， α_k 為迭代至第 k 次之弱分類器 $h_k(x)$ 之權重值，而 $h_k(x)$ 則為我們選用之SONFIN迭代至第 k 次，其計算方式為：

$$h_k(x) = \mu_k(x^p) \quad (13)$$

本方法使用SONFIN的原因是由於SONFIN是以Fuzzy為模糊理論為基礎，能夠產生程度上的區別，使得模糊理論系統除了具有對於不確定性及不精確性的處理能力外，亦能夠透過組織訓練來自我調整參數，達到最佳效能；而應於乳房鈣化點良劣分析上則可以分類出他良劣的程度，例如：該乳房鈣化點為惡性鈣化點，其為惡性鈣化點的程度又是如何，以此方法則可以辨別之，所以本計畫的分析結果就不單單只是良與劣這種分類，而有程度上的區別，此為本計畫選用之重要因素。

以下為本計畫使用之SONFIN Adaboost之流程圖，如圖5所示，初使化階段，先行建立訓練集以及其樣本權重的建置，經過SONFIN來進行訓練模式，產生出一個弱分類器，進而效能評估是否達到門檻值條件，否則將進行下一輪的訓練模式並進行錯誤率的計算以及樣本權重的調整，反反覆覆地直到門檻條件符合，最後將每次訓練模式的結果輸出組合成一個最終分類器，即達成本計畫提出方法的產出分類器。

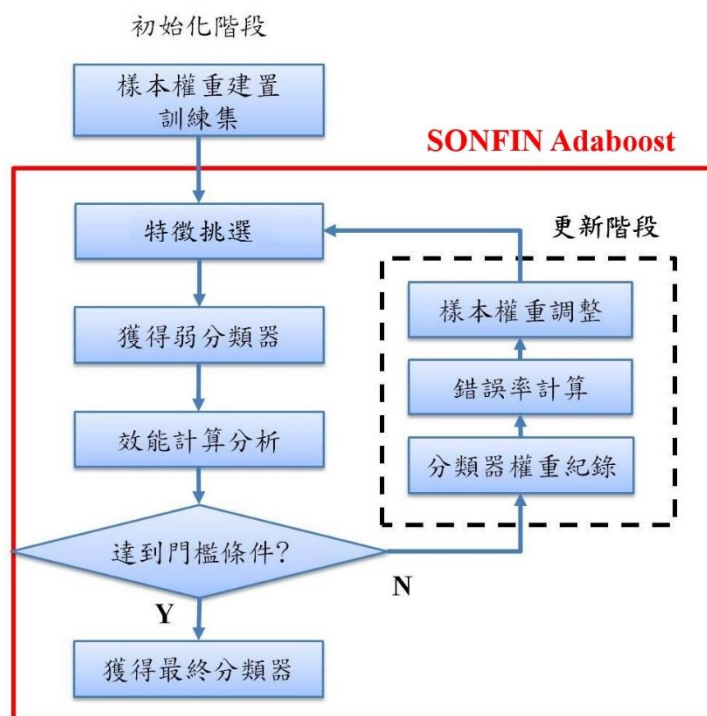


圖5、SONFIN Adaboost之流程圖

六、實驗樣本：

1. 威斯康辛州乳腺癌預後資料集(Wisconsin Prognostic Breast Cancer)：

本計畫使用175筆做為訓練集(90%)，剩下19筆為測試集(10%)，由於此資料集的特性，因此本計畫將所有筆數視為惡性類別的特徵進行後續的訓練與分類。第二階段實驗結合其他資料集時，由於此資料及比數數量較少，因此將以全部194筆與其他資料集合並做為訓練集使用，而測試集部分皆為全部資料集隨機選擇10%的資料進行測試。

2. 威斯康辛州乳腺癌診斷資料集(Wisconsin Breast Cancer Diagnostic Data Set)：

本計畫使用512筆做為訓練集(90%)，其中良性為323筆、惡性為189筆，剩下57筆為測試集(10%)，其中良性為36筆、惡性為21筆。第二階段實驗結合其他資料集時，將以455筆(80%)與其他資料集合並做為訓練集使用，而測試集部分皆為全部資料集隨機選擇10%的資料進行測試。

3. 威斯康辛州乳腺癌原始資料集(Wisconsin Breast Cancer Original Database)：

本計畫使用630筆做為訓練集(90%)，其中良性為412筆、惡性為218筆，剩下69筆為測試集(10%)，其中良性為46筆、惡性為23筆。第二階段實驗

結合其他資料集時，將以560筆(80%)與其他資料集合並做為訓練集使用，而測試集部分皆為全部資料集隨機選擇10%的資料進行測試。

4. 中山醫學大學附設醫院(Chung Shan Medical University Hospital)：

本計畫也使用合作醫院之醫師提供的乳房攝影影像進行擷取，擷取出兩千張不等的影像來為本次實驗所使用的影像，影像格式為DICOM檔並擷取出50×50大小格式，並同時剔除極端值影像或者是無法使用的影像來確保實驗的結果完整性。最後以訓練集跟測試集各1000筆乳房鈣化點影像特徵數據作為實驗樣本。第二階段實驗結合其他資料集時，將以隨機1000筆與其他資料集合並做為訓練集使用，而測試集部分皆為全部資料集隨機選擇10%的資料進行測試。

(五)結果

由於乳癌為我國女性好發癌症排名第一位，相較於歐美國家，台灣乳癌好發年齡比歐美國家早。在醫學技術方面，乳房攝影在偵測乳癌初期的微鈣化點症狀，尤其是在無法以觸診診斷或一些乳癌症狀較不明顯的狀況下有非常好的辨識率。本計畫預期結果為利用這些鈣化點的特徵，透過不同的特徵融合方式，結合SONFIN Adaboot的訓練，去處理大量乳房影像的問題，並且能夠利用所設計實驗之方法，更準確的分類出乳房鈣化點的良劣，並將最後結果提供醫生參考與做醫學上的相關研究。

以下將展示我們的實驗成果，如圖6所示為我們所設計之Matlab操作版面及流程，可以手動調整所要擷取之鈣化點影像的大小，以及更進階的切割出所要分析之鈣化點影像。並且使用GLCM進行的特徵的抽取以及使用SONFIN以及SONFIN Adaboot進訓練與行分類，而實驗結果的辨識率達96%以上。

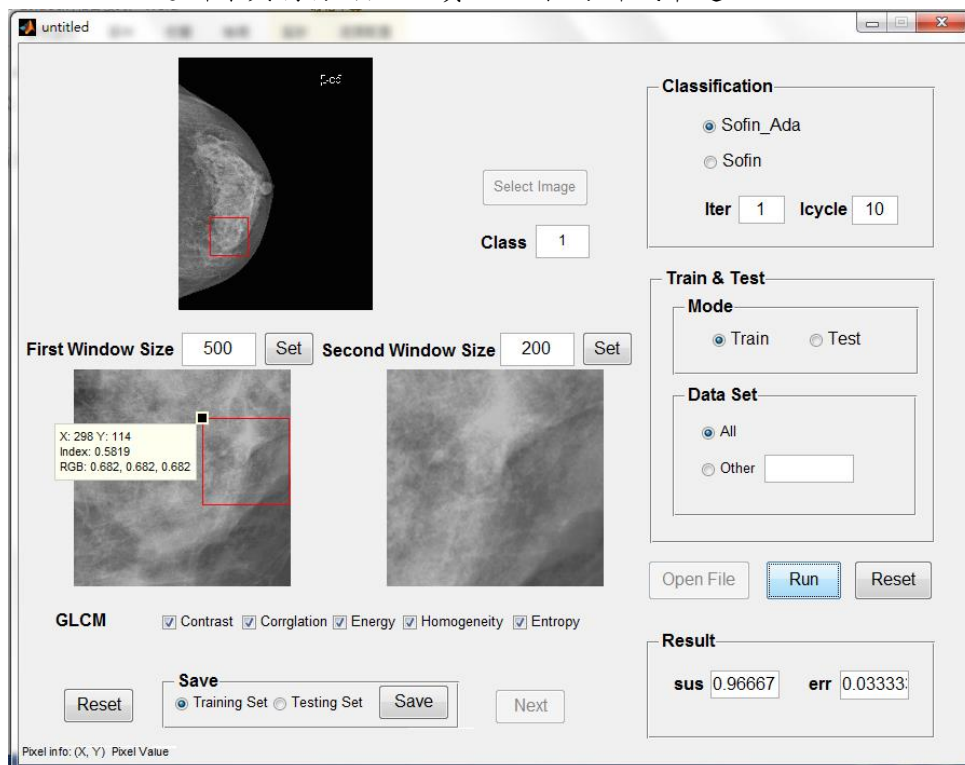


圖6、本系統之Matlab操作版面及流程

單一資料集實驗結果

本階段僅使用單一資料集分別進行 SONFIN Adaboost 與 SONFIN 之訓練，再個別以該資料集剩下之筆數進行測試，我們分別以 5 個與 10 個特徵進行兩次訓練與測試，並將兩次的成功率平均，所得之實驗結果如表一所示。

表一、使用單一資料集進行 SONFIN Adaboost 與 SONFIN 之成功率。

使用資料集	SONFIN Adaboost	SONFIN
Wisconsin Prognostic Breast Cancer(WPBC)	0.981	0.981
Wisconsin Breast Cancer Diagnostic Data Set(WBCD)	0.960	0.960
Wisconsin Breast Cancer Original Database(WBCO)	0.973	0.972
Chung Shan Medical University Hospital(CSMU)	0.987	0.985

雙資料集實驗結果

本階段將個別資料集資料取 80%的筆數後，進行兩兩交叉組合成為新的訓練集，分別進行 SONFIN Adaboost 與 SONFIN 之訓練，而測試集部分皆為該新資料集之全部資料集隨機選擇 10%的資料進行測試，我們分別以 5 個與 10 個特徵進行兩次訓練與測試，並將兩次的成功率平均，所得之實驗結果如表二所示。

表二、使用單一資料集進行SONFIN Adaboost與SONFIN之成功率。

使用資料集	SONFIN Adaboost	SONFIN
WPBC+WBCD	0.965	0.964
WPBC+WBCO	0.975	0.972
WPBC+CSMU	0.978	0.978
WBCD+WBCO	0.961	0.960
WBCD+CSMU	0.975	0.977
WBCO+CSMU	0.977	0.970

多資料集實驗結果

本階段同上一實驗步驟相同，也是將個別資料集資料取 80%的筆數後，取其中三種資料集的排列組合與全部資料集組合成為新的訓練集，分別進行 SONFIN Adaboost 與 SONFIN 之訓練，而測試集部分皆為該新資料集之全部資料集隨機選擇 10%的資料進行測試，我們分別以 5 個與 10 個特徵進行兩次訓練與測試，並將兩次的成功率平均，所得之實驗結果如表三所示。

表三、使用單一資料集進行SONFIN Adaboost與SONFIN之成功率。

使用資料集	SONFIN Adaboost	SONFIN
WPBC+WBCD+WBCO	0.965	0.964
WPBC+WBCD +CSMU	0.958	0.953
WPBC +WBCO+CSMU	0.966	0.962
WBCD+WBCO+CSMU	0.978	0.977
WPBC+WBCD+WBCO+CSMU	0.960	0.960

由實驗結果我們可以發現，並非資料集的數量越多成功率就越高，是由於資料集的建置條件以及資料正確性，以及他們所包含的特徵是否會影響分類成功率，亦或是特徵的同質性太高需要再更進一步的進行特徵挑選，這些都可能影響最後分類的成功率；使用之特徵數量也會影響到分類的時間複雜度，因此我們採用 5 個與 10 個特徵分別進行訓練，平均訓練所需時間為 30s，並會根據訓練集之筆數增加。

(六)結論

目前乳房鈣化點的學術研究有很多，但是皆難以清楚地區分出其良劣，更不用提其良劣程度，使用的方法也不夠多元化而使辨識的成功率下降，其中所消耗的時間也需耗費許多，因此，本計畫的完成能夠有效的提供一個乳房鈣化點良劣分析之診斷指標，幫助更正確地診斷病患的健康情況，並能及早做出相關的預防措施及治療方法，也使得研究人員在做相關研究時，能更快速及更清楚的了解到最佳的紋理特徵抽取，以及最適合乳房鈣化點良劣分析之特徵融合方法。且本計畫所提出利用簡單的影像處理工具、GLCM 特徵抽取法及 SONFIN Adaboost 所構成的架構，它改善了目前研究中所存在的問題，能夠精準與快速的分析出乳房鈣化點良劣之程度。

而在未來我們更計畫做更詳細的研究，例如將更完整的區分出單一顆乳房鈣化點與多顆重疊之乳房鈣化點之間的區別，以及不同的影像大小切割出之影像是否因為所抽出之紋理特徵的不同而影響分類的結果，亦或是增加紋理特徵抽取的方法進行相關的研究。

(七)參考文獻

- [1] Isabelle Thomassin-Naggara, Isabelle Thomassin-Naggara, Sophie Dechoux, "Added value of one-view breast tomosynthesis combined with digital mammography according to reader experience," *European Journal of Radiology*, vol.84, pp. 235-241, 2015.
- [2] Rahimeh Rouhi, Mehdi Jafari, Shohreh Kasaei, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, pp. 990-1002, 2015.
- [3] Nahid Sedighi, Amir Reza Radmard, Ali Radmehr, "Breast arterial calcification and risk of carotid atherosclerosis: Focusing on the preferentially affected layer of the vessel wall," *European Journal of Radiology*, vol. 79, pp. 250-256, 2011.
- [4] R.F. Brem, J.A. Rapelyea, J. Torrente, "Evaluation of breast amorphous calcifications by a computer-aided detection system in full-field digital mammography," *Breast Diseases: A Year Book Quarterly*, vol. 24, pp. 39-40, 2013.
- [5] Guang-ming Xian, "An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM," *Expert Systems with Applications*, vol. 37, pp. 6737-6741, 2010.
- [6] Xiang Ou, Wei Pan, Perry Xiao, "In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM)," *International Journal of Pharmaceutics*, vol. 460, pp. 28-32, 2014.

- [7] Qinggang Wu, Yong Gan, Yong Gan, "An active contour model based on fused texture features for image segmentation," *Neurocomputing*, vol. 151, pp. 1133-1141, 2015.
- [8] U. Rajendra Acharya, S. Vinitha Sree, M. Muthu Rama Krishnan, "Atherosclerotic Risk Stratification Strategy for Carotid Arteries Using Texture-Based Features," *Ultrasound in Medicine & Biology*, vol. 38, pp. 899-915, 2012.
- [9] Y.J. Chin, T.S. Ong, M.K.O. Goh, "Integrating palmprint and fingerprint for identity verification," *Third International Conference on Network and System, Security*, pp. 437-442, 2009.
- [10] Y.J. Chin, T.S. Ong, A.B.J. Teoh, "Integrated biometrics template protection technique based on fingerprint and palmprint feature-level fusion," *Information Fusion*, vol. 18, pp. 161-174, 2014.
- [11] Jaeik Jo, Sung Joo Lee, Kang Ryoung Park, "Detecting driver drowsiness using feature-level fusion and user-specific classification," *Expert Systems with Applications*, vol. 41, pp. 1139-1152, 2014.
- [12] Shao-Zi Li, Shao-Zi Li, Wei Wu, "Feature learning based on SAE-PCA network for human gesture recognition in RGBD images," *Neurocomputing*, vol. 151, pp. 565-573, 2015.
- [13] D.G. Stavrakoudis, J.B. Theocharis, G.C. Zalidis, "A Boosted Genetic Fuzzy Classifier for land cover classification of remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 529-544, 2011.
- [14] T. Sunil Kumar, Vivek Kanhangad, Ram Bilas Pachori, Shao-Zi Li, "Classification of seizure and seizure-free EEG signals using local binary patterns," *Biomedical Signal Processing and Control*, vol. 15, pp. 33-40, 2015.