行政院國家科學委員會補助

大專學生研究計畫研究成果報告

＊ ********* ** ＊
＊ 計　畫 ＊
＊ ： 乳房腫瘤偵測系統使用多特徵值與模糊線性判別式 ＊
＊ 名　稱 ＊
＊ ********* ** ＊

執行計畫學生： 鄭子寬
學生計畫編號： NSC 101-2815-C-040-010-E
研 究 期 間 ： 101 年 07 月 01 日至 102 年 02 月 28 日止，計 8 個月
指 導 教 授 ： 徐麗蘋

處理方式： 本計畫可公開查詢

執 行 單 位： 中山醫學大學醫學資訊學系

中華民國 102 年 03 月 28 日

(一)摘要

乳癌可以說是女性癌症中最常見的一種，根據世界衛生組織指出，每年全球年齡在 40-55 歲之間的女性會有至少 120 萬個被診斷出罹患乳癌，近年來越來越重視癌症的診斷，期望未來能在診斷及更有效的診斷方式下來提高診斷正確率。

由於乳癌腫瘤影像，在灰階值亮度會比其它正常組織為高，且紋理組織非常明顯的兩個特徵，所以本計畫針對這兩個特徵提出一套的演算法。在前處理的部分我們首先利用影像投影法對乳房的 CT 影像去除其中不必要的影像區域，接著將影像延著乳腺分不進行許多大小相同的分割區塊，每個區塊的大小皆符合特徵抽取時所需要的區塊大小，這些我們稱為 ROI 區塊，ROI 區塊將會做為特徵抽取的區塊。在特徵值擷取的部分我們採用 Local Binary Pattern 的特徵用來偵測乳房腫瘤的特徵，我們針對每一個 ROI 影像區塊去抽取出這些 Local Binary Pattern 特徵，並且用這些抽取出來的特徵來進行類神經網路及模糊線性判斷分析來進行機器學習，最後利用多次學習後得到的最佳結果來對乳房影像進行分類，判斷出該影像是否有腫瘤。

在實驗結果的部分我們的資料是使用中山醫學院的乳房腫瘤資料庫，使用 Local Binary Pattern 的方法來進行乳房腫瘤的特徵抽取，最後以模糊線性判別式來進行乳房腫瘤的特徵分類，而系統判斷的成功率大約為 60%，同時也有請合作醫院的放射師及醫師做評估驗證，如此更可以得到更正確的系統辨識成功率。我們計畫的成功實踐將可以改善現有偵測方法的準確度，產生一個真正可以使用在輔助醫療診斷系統的乳癌偵測系統，並能有效的減少醫師判斷的時間。

(二)研究動機與研究問題

乳癌（Breast cancer）是醫療上不算少見的急症，根據衛生署的統計，每年在台灣將近每十萬人口中有 34.91 位女性罹患乳癌，每十萬女性人口中有 11.36 人因乳癌死亡病例，在歐洲的統計，女性乳癌的初發生率約為每十萬人口有 105 人，女性乳癌癌症死亡率則約為每十萬女性人口有 40 人，而美國病例去是台灣的 4.5 倍新病例發生。它是從乳腺的上皮細胞或小葉生長出來的一種惡性瘤，因癌細胞的生長失去控制，會侵入並破壞鄰近的組織及器官，或經由血液或淋巴系統轉移到其它器官，正確而及時的診斷在早期乳癌，沒有侵犯附近的組織，病人通常接受乳房保留手術及放射線治療或行乳房全切除手術，此方法對早期的乳癌相當有療效，某些病人則需接受術後輔助性化學治療或荷爾蒙治療，以預防乳癌的復發。一般偵測的方法可以方為三類: 乳腺 X 光攝影檢查術(Mammography），核磁共振造影（Magnetic resonance imaging, MRI）與超音波技術（Sonograms, Ultrasound Images）等，首先在乳房 X 光攝影的優點是對於早期乳癌的鈣化診斷比超音波佳，正確性比較高，而缺點是對年輕、緻密型乳房，不易診斷、隆乳者不宜、有輻射線，乳房 MRI 的優點 一次掃描可完整檢察兩側乳房、胸壁及腋下淋巴之變化、快速擷取影像，而缺點是大部分醫師對乳房 MRI 尚未熟悉，故不普遍、費用較昂貴，乳房超音波的優點是費用便宜、操作方便，而缺點是操作醫師需要極佳經驗才不容易誤診、儀器好壞影響判讀，若以醫師常用不易誤判的情況下，因此我們採用了乳腺 X 光攝影檢查術(Mammography)的方式來完成本計畫的計畫，傳統的偵測都必須藉重醫師們的專業知識，如此會使得醫師的疲勞感增加而產生判斷錯誤的發生，因此我們提出這個計畫的目的是要發展出一套成功率高且精確度高的乳房腫瘤偵測系統，此系統可以輔助醫師的檢測減少醫師的疲勞和錯誤率的發生。

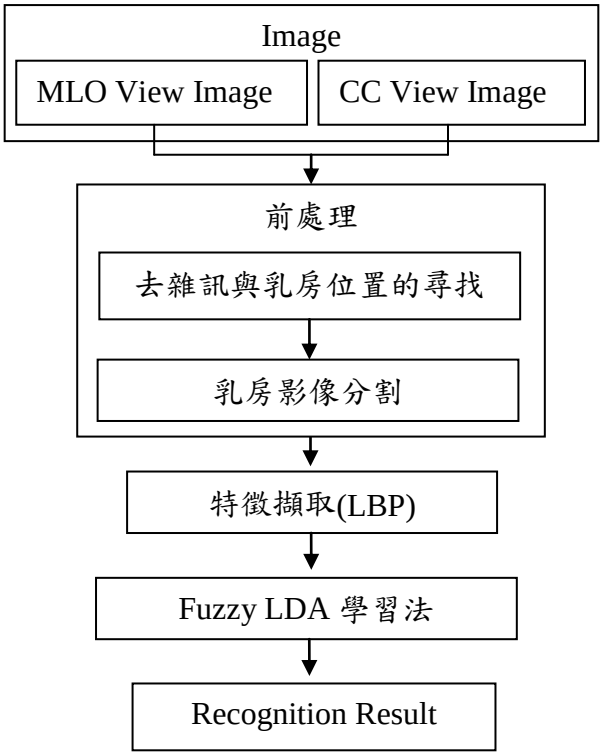
(三)文獻回顧與探討

在過去針對乳房腫瘤的相關研究中，研究的方向可以分為兩大類，第一是特徵擷取[1-3]，第二是分類器[1-3]的使用。首先在特徵擷取的部分，特徵擷取可以區分為統計型和結構型二種。統計型的特徵是使用灰階直方圖統計或是灰階共發生矩陣來計算不同區域的特性，如影像的mean或standard deviation運用灰階值方圖來獲得影像中亮度的特徵，或使用影像的Average distance來做為影像對比增強與否的指標，讓較低的對比度有效的提升至較高型態[4]，或是應用在原圖影像計算冠狀動脈中心，並且求出血管與動脈的最短路徑[5]。在結構型的特徵部分，一般是使用原始影像的基本紋理元素，利用這些綜合的紋理規則來表示紋理的型態， 如[6-7]的專家利用Laws' energy measure，將它結合多階小波轉換來做影像紋理擷取，並且應用在大型的影像資料庫檢索上，或是應用在頸動脈粥樣硬化的3D超音波影像紋理分析，在[8-9]則是利用曲波轉換與小波轉換進行乳房影像的良性與惡性腫瘤分類，結果顯示曲波轉換較於小波轉換來的好且在統計學上也有明顯的差異。在第[10]中，使用Local Binary Pattern來進行圖像特徵的擷取的方法。接下來是分類器的部分，近年來最常被使用的分類器為SVM[11-15]，它可以搭配Genetic Algorithm、Gaussian或polynomial等不同的Kernel提高SVM分類上面的結果，在[16]中提到使用模糊非線性轉換做資料筆數少且高維度的特徵，最後做T檢定的結果都優於PCA或KPCA等，在[17]中提到使用Fuzzy C-means clustering 進行腫瘤輪廓特徵分類，而最後做P檢定的結果是有顯著性差異。

從文獻回顧中，我們發現了乳房腫瘤偵測有2個需要改進的地方，首先是，文獻使用了許多的特徵來描述乳房腫瘤，這些特徵抽取方法包含有影像的灰階直方統計圖和頻率域特徵抽取等，太多數量的特徵會讓系統的計算複雜度提高，本計畫便思考是否需要使用大量的特徵才能夠有效的且正確的辨識出乳房的腫瘤。接著是評估成功率方面所使用的資料庫，大多數的論文都只使用1種已知的乳房資料庫來測試其所提出方法的成功與否，這會讓所提出的偵測演算法無法明顯的顯示出方法的優勝之處。為了要改善上述兩個地方，我們提出了一種有效、簡單且高辨識的乳房腫瘤偵測系統，減少醫師因為疲勞而產生的誤判或失誤。

(四)研究方法及步驟

本計畫進行乳房腫瘤辨識系統其流程如圖一所示，當一張CT影像進入系統後，會先做影像前處理，保留影像中較亮的部分，接下來是特徵抽取，第三步驟則是學習方法的介紹，在這裡我們利用類神經網路及模糊線性判別分析的方法來完成辨識乳房腫瘤的工作，而結果部份我們請兩位合作醫院的醫師與放射師做驗證，接下來會依序作詳細的介紹。



圖一：乳房腫瘤偵測流程

(1)影像格式

在乳房 X 光攝影中，我們使用的影像資料庫分為三種，第一種是由中山醫學大學附設醫院（CSMU）所提供的 CSMU Mammographic Database、第二種是美國南佛羅里達大學的乳房攝影影像資料庫(Digital Database for Screening Mammography, DDSM)及最後一種是歐洲 The Mammographic Image Analysis Society（MIAS）所提供的 MIAS MiniMammographic Database，而每個資料庫都可以分別為:兩側斜位向(由側邊照)Bilateral MLO view 與兩側頭尾向(由上往下照)Bilateral CCview 的影像，而乳房影像的性質可分為三類，首先是緻密腺體 (Dense-glandular)，其次是脂肪 (Fatty)最後是脂肪腺體(Fatty-glandular)。緊密(緻密腺體代表)的乳房組織，在 X 光片中就以白色顯現，一點也不緊密的脂肪，看起來就像是灰色偏透明。

乳房中大多數是乳房組織，並且非常緊密，隨著年齡的增長，乳房就像皮膚一樣老化，乳房組織逐漸減少，脂肪逐漸增多。三、四十歲時，兩者的比例約各佔一半，

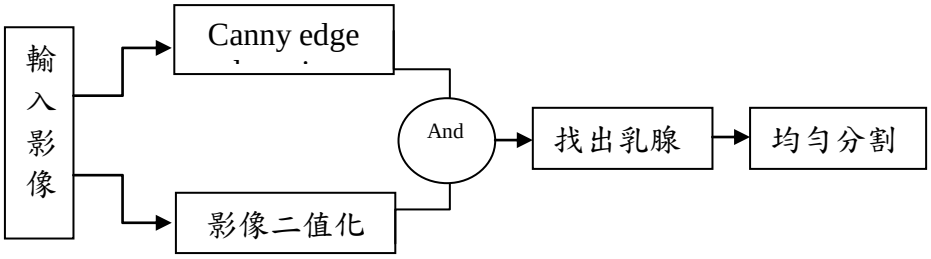
此時所拍攝的 X 光片就以脂肪腺體代表。當五、六十歲以上，乳房大多為脂肪，此時所拍攝的 X 光片就以脂肪為代表，而年輕婦女的乳房組織較為緊密，多有緊實的纖維囊腫，導致在乳房 X 光影像中十分困難找出這類型腫瘤。

(2)前處理(Preprocessing):

A. 去雜訊與乳房位置的尋找

第一步驟為針對整張乳房影像進行中值濾波去除雜訊，同時加強影像的特徵資訊，而獲得一張清晰可用的影像，然後將去除雜訊後的乳房影像進行投影，接著將投影過後的影像進行二值化的個別影像元素處理動作，目的是為了去除較暗的非腫瘤區域，因此得到較亮的腫瘤區域。

B. 乳房影像分割



圖二、影像分割流程圖

第二步驟為決定影像分割區的區域，將所取得的二值化乳房影像與 Canny 邊緣偵測演算法的結果進行邏輯預算，將有交集的部分留下來並進行等範圍的影像切割，分割後的影像包含乳線等組織，此區塊我們稱為感興趣區域(Regions Of Interest, ROI)，而 ROI 的數量會隨著影像的長與寬而有所差異，接著經由特徵擷取的方法抽取可以得知特徵影像區塊的能量大小，並做篩選無相關的區塊，以下將做詳細的特徵擷取的介紹。

圖三、 (a)原始影像、(b)二值化的影像、(c)Canny 邊緣偵測、(d)交集後乳腺區域

(3) 特徵值擷取(Feature Extraction):

當將影像分割完成後，接下來我們會對影像中的每個區塊的特徵擷取，在這裡我們使用的方法是 LBP ， LBP 是一種用來描述圖像局部紋理特徵的演算法，優點為其統計直方圖（Histogram）之資料不隨影像灰階值改變與影像旋轉影響。且對影像進行運算後，其特徵值亦不會改變太多。

LBP 的計算方法是藉由比較影像中每一個像素與其鄰近像素的差異，取得一個以二進位數值表示的結果。其局部的處理方式為以 3x3 的區塊將影像標記，首先將彩

色影像轉為灰階影像，假設影像中某區塊的灰階值如圖四，那麼 3x3 區塊中間的值即為閥值（threshold）。

68	79	38
56	55	43
48	39	59

A. 圖四：3×3 區塊範例

在圖四的閥值為 55，周圍八個點若大於或等於閥值，就將此點設為 1；反之，則將點的值設為 0，如圖五所示。

1	1	0
1		0
0	0	1

B. 圖五：經過 LBP 運算結果

再將圖五計算結果週圍給予權重值，如圖六所示。

2^0	2^1	2^2
2^7		2^3
2^6	2^5	2^4

C. 圖六：3×3 區塊之權重

將圖四與圖五計算之結果相加計算，得出此 3×3 區域之 LBP 值。所以圖四的 3×3 區塊 LBP 值為 $1*2^0+ 1* 2^1+ 1*2^4+ 1*2^7= 147$。對於影像中每一個像素進行迭代，即可累加統計出 IR 圖形影像的區域二元特徵直方圖（LBP histogram）。依據上述之 LBP 計算方法，即可統計各別紅外線光譜圖形之二元特徵直方圖（LBP histogram），本文採用 3×3 區塊的方式建立各別像素之特徵值，因此建立之直方圖中將存在 256 個 bin，也就是一個 256 階的陣列。

(4)學習方法(Learning Method)

在學習方法的部分，我們採用模糊線性判別分析，模糊線性判斷分析可以將多維度參數，有效的表示每個維度所隸屬於各類別上的程度。

模糊線性判斷分析（Fuzzy Linear Discriminant Analysis）簡稱 Fuzzy LDA，乃是一種根據 Linear Discriminant Analysis algorithm 衍生而來的分類法，透過模糊邏輯的概念，希望能進一步提升分群的效果。

假設我們現在預期的分類數目為 w（w_1, w_2,...），整組維度包含 n 個（x_1, x_2 , ..., x_n），則我們可以以一個 l×n 的矩陣 U 來表示每個維度隸屬於每個類別的程度。倘若我們針對其中一點 x_j 而言，他隸屬於各個類別的程度總和應該正好等於 1，接著根據矩陣 U，我們可以定義出我們的目標函數 W，其公式如(1)。

$$J_{Fuzzy\ LDA}(U,w_1 W_2 \cdots)=\sum_{i=0}^{c} J_i=\left|\frac{W^T FS_b W}{W^T FS_w W}\right|+\sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c}\mu_{ij}^{p}-1\right) \tag{1}$$

在方程式中，λ為機率係數，可以是介於[1, ∞]之間的任何數；p 為常數，為影響控制模糊歸屬函式。為了求得最佳化的$J_{Fuzzy\ LDA}$，我們針對各傳入參數分別進行微分，便可歸納出下列結論，其公式如下(2)。

$$W=C\cdot FS_W^{-1}\cdot\mu_{ij}^{p}(M_i-M') \tag{2}$$

其中μ_{ij}^{p}可以依據計算後得到如(3)。

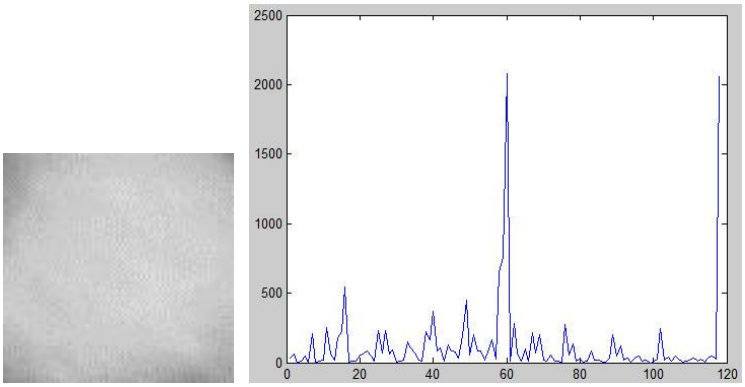
$$J_{Fuzzy\ LDA}(W)=\sum_{y=1}^{n} p\lambda\sum_{i=1}^{c}\mu_{ij}^{p}+\left(W^T S_B WP\mu_{ij}^{p-1}\right)\left(W^T\mu_{ij}^{p}S_W W\right)^{-1}-\left(W^T P\mu_{ij}^{p}S_B W\right)\left(W^T\mu_{ij}^{p}S_W WP\right)^{-2}=0 \tag{3}$$

(5)結論

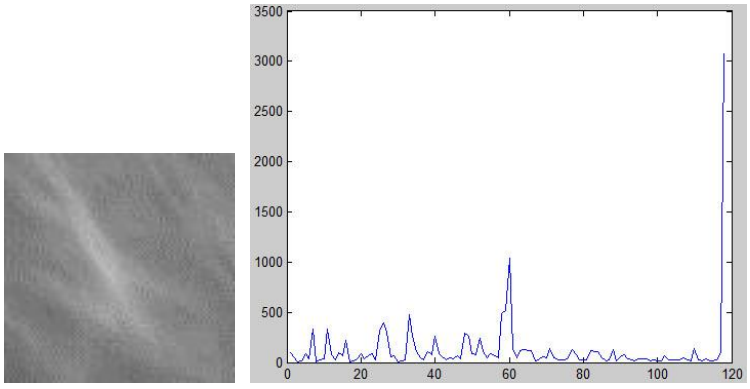
最後經由模糊線性判斷分析後的結果，我們使用邏輯分析中的邏輯閘來判斷最後輸出分類的結果，其邏輯閘的特性為當兩者只要一個是腫瘤(True)則為腫瘤，當兩則為非腫瘤(False)則為非腫瘤，而最後輸出的結果是請合作醫院的醫師與放射師做驗證並比對我們運算的結果是否合宜。

(五)實驗結果

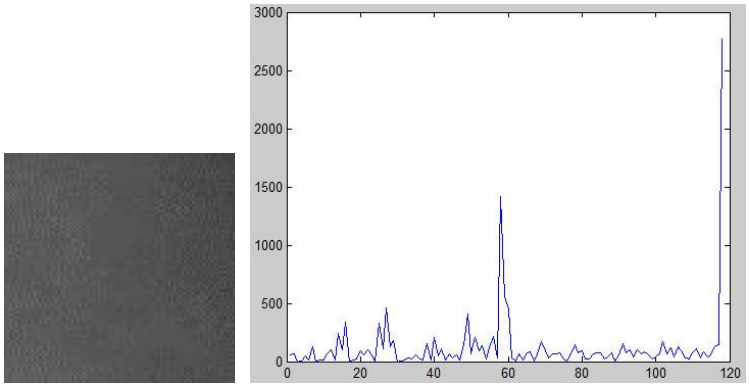
在這個計畫中，乳房影像的來源是採用中山醫學院的乳房攝影資料庫，影像的數量總共使用104張乳房攝影的影像來進行演算法的測試，在演算法運作之前，我們先將104張的乳房影像進行LBP特徵擷取，並將截取出來的特徵值進行資料正規化，每張乳房影像可以取得118個特徵值來進行機器訓練，最後再將這些訓練過後的資料與新輸入的乳房影像經由模糊線性判別式來進行特徵值的比對分類，進而判斷出該乳房影像中是否有腫瘤。下面附上有腫瘤與沒有腫瘤的特徵值直方圖。



圖七、腫瘤影像及腫瘤LBP特徵值



圖八、乳腺影像及乳腺LBP特徵值



圖九、脂肪影像及脂肪LBP特徵值

在醫學的診斷程序上，實驗的準確性必須以公正方式來評比才能令人信服。根據醫學影像技術普遍採用的四個統計值：腫瘤組織–正確辨識個數(True Positive Number, TPN)，正常組織–正確辨識個數(True Negative Number, TNN)，腫瘤組織-錯誤辨識個數(False Negative Number, FNN)，和正常組織–錯誤辨識個數(False Positive Number, FPN)，以及其所衍生的三個效能因子做比較，他們分別是偵測率(Detection Rate, DR)、錯誤警示率(False Alarm Rate, FAR)和正確分類率(Correction Classification Rate, CR)三個，它們的計算方式如下:

$$DR=\frac{TPN}{TPN+FNN} \tag{4}$$

$$FAR=\frac{FPN}{TNN+FPN} \tag{5}$$

$$CR=\frac{(TPN+TNN)}{TPN+FNN+TNN+FPN} \tag{6}$$

其中 Np 代表正樣本(腫瘤組織)的個數，Np =TPN+FNN；Nn 代表負樣本(正常組織)的個數，N n =TNN+FPN。全部樣本的個數 N = N p+ N n。

DR代表是腫瘤偵測的偵測率，若百分比越高則偵測率越高，FAR代表是腫瘤偵測的錯誤率，若百分比越低則錯誤率越低，CR代表是腫瘤偵測的正確分類率，若百分比越高則正確分類率越高。下表為我們系統值型後的成功率。

I Fuzzy LDA	Train	Test
DR	78	56
FAR	31	53
CR	74	63

表一、系統成功率

目前我們系統的成功率約六成，造成成功率不高的原因可能是因為我們乳房影像資料庫中同一個病人會從不同方向取得該乳房的影像，而造成特徵抽取的計算有所差異，亦或者是我們可能有抽取到不必要的特徵來進行訓練，將來可望使用不同的特徵抽取方法來進行多重的成功率比對，以便找出最佳的特徵值進行乳房腫瘤的分辨。

(六) 結論與建議

目前乳房腫瘤偵測的研究有很多，但是正確率卻不高，所得到的結果也不是非常地準確，研究中使用 LBP 演算法來抽取乳房影像的特徵，並且以模糊現性判別式來進行乳房腫瘤的偵測，其獲得了約 60%的成功率，在將來的研究中將尋找不同的特徵值來進行實驗，以提高腫瘤的偵測率。若這項研究能夠獲得較高的成功率，則可以有效地幫助醫師的診斷，節省醫師診斷的時間。近年來乳癌的罹患率已經逐漸提高，若真的實作出一套乳房腫瘤偵測系統，且擁有高正確率，將會為醫師帶來更多便利以及幫助，同時也能避免過勞檢測時所帶來的疏失。

(七) 參考資料

[1] M. M. Eltoukhy, I. Faye and B. B. Samir, “Breast cancer diagnosis in digital mammogram using multiscale curvelet transform,” *Journal of Computerized Medical Imaging and Graphics*, vol. 34, pp. 269–276, 2010.

[2] D. R. Chen, Y. L. Huang and S. H. Lin, “Computer-aided diagnosis with textural features for breast lesions in sonograms,” *Journal of Computerized Medical Imaging and Graphics*, vol.35, pp. 220–226, 2011.

[3] M. Y. Lee and C. S. Yang, “Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardized thermograph images,” *Journal of Computer Methods And Programs In Biomedicine*, vol.100, pp. 269–282, 2010.

[4] Z. Chen, B. R. Abidi, D. L. Page and M. A. Abidi, “Gray-Level Grouping (GLG): An Automatic Method for Optimized Image Contrast Enhancement—Part I: The Basic Method,” *IEEE Transactions on Image Processing*, vol.15, 2006.

[5] V. Fono, A. C. Evans, K. Botteron and C. R. Almli**,** “Coronary Artery Centerline Extraction in Coronary CTA,” *Journal of Frontiers of Computer Science and Technology*, vol.11, pp. 999-1005, 2011.

[6] S. S. Kumar and L. Ganesan, “ Texture classification using wavelet based Laws energy measure,” *International Journal of Soft Computing*, vol.3, pp. 293–296, 2008.

[7] J. Awad , A. Krasinski , G. Parraga and A. Fenster, “Texture analysis of carotid artery atherosclerosis from three-dimensional ultrasound images,” *International Journal of Medical Physics Research and Practice* , vol. 37, 2010.

[8] M. M. Eltoukhy, I. Faye and B. B. Samir, “A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram,” *Journal of Computers in Biology and Medicine*, vol.40, pp.384-391, 2010.

[9] L. Dettori and L. Semler, “A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography,” *Journal of Pattern Recognition Letters*, vol.37, pp.486-498, 2007.

[10] Zhenhua Guo, Lei Zhang, Member, and David Zhang*,“A Completed Modeling of Local Binary Pattern Operator for Texture Classification, *Submitted to IEEE Transactions on Image Processing.*

[11] H. L. Chen, B. Yang, J. Liu and D. Y. Liu, “A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis,” *Journal of Expert Systems with Applications*, vol. 38, pp. 9014–9022, 2011.

[12] D. Wang, L. Shi and P. A. Heng, “Automatic detection of breast cancers in mammograms using structured support vector machines,” *Journal of Neurocomputing*, vol.72, pp. 3296–3302, 2009.

[13] M. Karabatak and M. C. Ince, “An expert system for detection of breast cancer based on association rules and neural network,” *Journal of Expert Systems with Applications,* vol.36, pp.3465-3469, 2009.

[14] K. Pola and S. Günes, “Breast cancer diagnosis using least square support vector machine,” *Journal of Digital Signal Processing,* vol.17, pp.694-701, 2007.

[15] T. S. Subashini, V. Ramalingam and S. Palanivel, “Breast mass classification based on cytological patterns using RBFNN and SVM,” *Journal of Expert Systems with Applications,* vol.36, pp.5284-5290, 2009.

[16] D. C. Li, C. W. Liu and S. C. Hu, “A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets,” *Journal of Artificial Intelligence in Medicine,* vol. 52, pp. 45–52, 2011.

[17] W. K. Moon, S. C. Chang, C. S. Huang and R. F. Chang, “Breast tumor classification using fuzzy clustering for breast elastography ,” *Journal of Ultrasound in Medicine and Biology,* vol.37, pp.700-708, 2011.