

科技部補助
大專學生研究計畫研究成果報告

計畫名稱：基於DL-RMS的特徵抽取法與統計特徵篩選之乳房鈣化點良惡分析系統

報告類別：成果報告
執行計畫學生：張婷淇
學生計畫編號：MOST 110-2813-C-040-069-E
研究期間：110年07月01日至111年02月28日止，計8個月
指導教授：賴慶祥

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

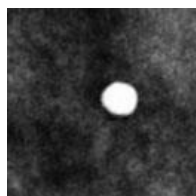
中華民國 111年03月29日

(一)摘要

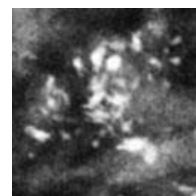
現今醫師主要透過乳房攝影影像觀察婦女乳房的狀況，但乳房鈣化點十分微小，一般需透過具侵入性的細針定位手術檢測方能判斷鈣化點群聚處中的鈣化點良惡，導致病患接受檢測的意願不高，耽誤乳癌的診治時間。此外，目前與乳癌相關的深度學習研究大多使用同一模型進行影像特徵抽取與選擇，鮮少針對鈣化點特徵選擇方法進行深入研究與改善。因此本計畫開發出「基於 DL-RMS 的特徵抽取法與統計特徵篩選之乳房鈣化點良惡分析系統」，首先將乳房攝影影像進行銳利化預處理，並透過 Gathering CNN 與 Mask R-CNN 找出鈣化點群聚處位置並進行鈣化點的標記與切割，接著使用影像組學(Radiomics)抽取大量影像特徵，以用於區分良惡鈣化點，但高維度特徵易導致模型產生過擬合(Overfitting)問題，所以分別使用統計學方法進行最佳特徵選取，以找出能代表鈣化點特徵的數值，最後將其分別輸入機器學習模型來分類乳房鈣化點的良惡。在實驗結果的部分，本計畫將有無經過特徵選擇步驟的機器學習模型分類結果進行比較，發現未經過特徵選擇步驟的模型分類準確率最高可達 72.34%，而經過特徵選擇步驟的模型分類準確率可達 85.11%，證明本計畫提出的特徵選擇方法有助於提升系統分類良惡鈣化點的準確率。期望本計畫所提出的方法能實際解決醫師診斷乳癌所面臨的問題，透過結合 Radiomics 與深度學習特徵值的特徵選擇方法，自動分析乳房攝影影像中的鈣化點良惡，縮短醫師診療時間的同時也減輕病患負擔。

(二)研究動機與研究問題

依據衛生福利部統計資料顯示，乳癌為我國婦女罹患率第一位之癌症，對於婦女健康造成極大的威脅，目前不僅政府積極推動乳癌篩檢政策，婦女與各醫療單位也相當重視乳癌篩檢與防治。而乳房攝影為目前最主要的乳房篩檢方法，常用於偵測乳房中的腫瘤與鈣化點，醫師能透過觀察乳房攝影影像中鈣化點群聚處的鈣化點大小、分布狀況及形狀，判斷為良性或惡性鈣化點，並依據惡性鈣化點進一步評估病患罹患乳癌的機率與乳癌期別。本計畫為了瞭解如何分辨鈣化點良惡以及醫師判讀乳房攝影影像的實際需求，與中山醫學大學附設醫院醫學影像部的醫師進行合作，依據醫師的臨床經驗指出，良性鈣化點在影像中的形狀多為圓形、茶杯狀或爆米花狀，且具有分散、均勻分布或對稱的特性，如圖 1(a)所示，惡性鈣化點的形狀多呈分岔、線形或不規則狀，且通常具有微小、聚集的特性，如圖 1(b)所示。



(a)



(b)

圖 1、(a)為良性鈣化點影像，(b)為惡性鈣化點影像

然而鈣化點於乳房攝影影像中非常微小且與乳腺組織與的特徵相近，在 X 光影像中皆為白色，醫師難以從乳房攝影影像中一眼觀察微小鈣化點群聚處中的鈣化點，一般需透過具侵入性的細針定位手術檢測放大觀察方可分辨是否為乳癌細胞壞死所導致的惡性鈣化點，但此檢測將會使患者感到疼痛，造成患者接受檢測的意願不高，而錯失治療最佳時機的問題。再者，醫師判讀影像並從中找出病灶位置的過程須經過些許時間方能進行診斷，且準確率也易受到醫師的疲勞程度影響。最後，本計畫參考多篇乳房鈣化點相關研究，發現其在分辨良性與惡性鈣化點的準確率仍有提升的空間。

近年來許多學者利用影像組學(Radiomics)高通量地從醫學影像中抽取大量不同類型的特徵，再透過統計學的演算法提取用來描述病變的定量特徵，以進行疾病的分析與預測。此外，目前人工智慧已廣泛應用於醫學影像的分析與判讀，透過大量的醫學影像資料，將能讓機器或模型針對輸入的資料進行分類或預測。因此本計畫期望利用影像組學、機器學習、深度學習與統計學方法，開發出「基於 DL-RMS 的特徵抽取法與統計特徵篩選之乳房鈣化點良惡分析系統」來解決上述問題，提供醫師輔助判斷婦女乳房中的良性與惡性鈣化點的系統，以實現自動分析乳房攝影影像來分辨群聚鈣化點中的良惡鈣化點之目標，降低醫師診療過程的時間，同時給予病患最佳的就醫體驗。

(三) 文獻回顧與探討

在乳房鈣化點的判斷中，影像的特徵抽取與特徵選擇的結果將會對鈣化點的分辨結果產生影響，因此正確地從乳房攝影影像中抽取微小鈣化點的特徵，並從大量特徵中篩選出具代表性特徵的步驟是十分重要的。本計畫參考多篇文獻的概念與方法，並針對其問題進行改善或加強其方法，以實現分辨乳房中良性與惡性鈣化點的目標。本計畫將探討的文獻分為四部份，以下將逐一進行介紹。

1. 影像組學應用於醫學影像上的相關研究

在 2019 年，Losurdo 學者等人利用影像組學從 CESM (Contrast-Enhanced Spectral Mammography) 影像中判斷出乳腺癌[1]，此研究證明 Radiomics 所提取的紋理特徵有助於區分良性乳腺疾病或是惡性乳腺癌。同一年，Ibrahim 學者等人針對 Radiomics 應用於臨床醫學領域所面臨的挑戰，以及在臨床決策支持系統中的應用潛力進行回顧與統整[2]，提出使用 Radiomics 提取的特徵未必能獲得高準確率結果的觀點，因此須針對特徵依賴性和相關性進行詳盡的分析。

在 2020 年，Caballo 學者等人利用 Unet 深度學習演算法從乳房 CT 影像中分割出良性與惡性腫瘤，再利用 Radiomics 抽取出乳房 CT 影像中的腫瘤特徵，最後利用統計學的 MANOVA 同時比較 2 個以上的應變數之間的差異性[3]，以篩選出主要的特徵。由於本計畫目的為分辨乳房攝影影像中的微小鈣化點，因此參考其特徵抽取的架構及使用統計學方法進行特徵篩選的概念，透過 Radiomics 提取群聚鈣化點影像中的大量特徵，以便後續利用統計學方法進行特徵篩選。

2. 機器學習與深度學習應用於提取醫學影像中特徵的相關研究

在 2020 年，魏同學提出「基於 BI-CNN 進行婦女乳房鈣化點群聚偵測及良惡預測系統」，期望運用深度學習的方法偵測出群聚鈣化點的位置並分析其良惡 [4]，有效輔助醫生進行診斷，但在準確率方面，僅採用深度學習方法來找出良性與惡性的乳房鈣化點的架構仍有改善空間，而特徵選擇的部份，此方法所使用的深度學習方法會依據專業醫師標記的乳房攝影影像自動找出鈣化點特徵，但尚未運用統計學方法作為衡量指標。因此本計畫延續與改良此計畫的方法，先將輸入影像進行前處理，之後運用深度學習方法切割出鈣化點的群聚處，再利用本計畫提出的方法進行後續的良性與惡性鈣化點分類。

在 2017 年，Macías-García 學者等人提出利用自動編碼器(AutoEncoder)深度學習演算法，針對經過免疫組織化學染色的組織切片數據進行預處理[5]，以降低數據中雜訊的影響，再透過存活分析(Survival analysis)和 Schoenfeld residuals 獲得乳腺癌中主要指標的相關性，證明使用自動編碼器可獲得具有乳腺癌預後價值的基因組特徵。在 2020 年，Toğaçar 學者等人使用自動編碼器重建資料集的影像，再透過 CNN 抽取浸潤性乳腺癌影像中的特徵，最後利用統計學方法分類出浸潤性導管癌[6]。此研究針對使用與不使用自動編碼器處理的數據進行浸潤性導管癌的分類測試，實驗結果證明經過自動編碼器處理的影像在最後的分類結果有較良好的表現。因此本研究參考上述文獻所使用的自動編碼器方法，用於抽取乳房攝影影像中的鈣化點特徵，希望能找尋到具有意義的乳房鈣化點特徵。

3. 現今統計學應用於選擇醫學影像中特徵的相關研究

在 2018 年，Hsu 學者等人為了準確區分 FDG-avid 組織為腫瘤或正常組織，將 Radiomics 的特徵利用獨立樣本 t 檢定方法選擇出具有意義的 Radiomics 特徵 [7]。在 2017 年，Ling 學者等人提出將 global shape descriptor、Hu moment invariant 和 Fourier descriptor 方法分別用於抽取脊柱 MRI 影像中的胸部、腰椎與骶骨特徵，再運用 ANOVA 統計學方法進行特徵選擇[8]，其準確率高達 91.1%，證明 ANOVA 作為特徵選擇的方法可有效篩選出影像中的重要特徵。在 2020 年，Das 學者等人提出將統計學的 MANOVA 方法用於選擇活體組織切片影像中特徵的研究[9]，利用 MANOVA 方法檢驗兒童髓母細胞瘤(Childhood Medulloblastoma, CMB)的活體組織切片中，是否存在具有統計學意義的特徵，以便後續對正常樣本及其各種亞型中的 CMB 進行分類。此研究證明經過特徵選擇再進行 CMB 分類的準確率相對較高，因此 MANOVA 可以有效地作為機器學習的特徵選擇方法。本計畫參考上述文獻的研究方法，希望使用獨立樣本 t 檢定、ANOVA 以及 MANOVA 方法選擇 Radiomics 提取出的大量乳房攝影影像特徵，並依據三者的實驗結果，找出具有代表性的乳房微小鈣化點特徵。

4. 現今應用於分類及預測醫學領域病灶的相關研究

醫學數據分類是非常複雜且具有挑戰性的，在 2019 年，Alam 學者等人提出基於特徵排名與特徵選擇策略的醫學數據分類方法[10]，再透過隨機森林分類器區分資料集排名較高的特徵，以預測出具有類似特徵的疾病。此研究在 10 個不

同疾病的資料集進行實驗，證實隨機森林分類器在具有相似特徵的疾病分類方面有不錯的成果。

在 2018 年，Makwana 學者等人提出用於辨識乳房惡性腫瘤的分割與分類方法[11]，首先利用灰階共生矩陣(GLCM)抽取紋理特徵，再透過 SVM 分類器針對抽取的特徵進行分類，其分類惡性腫瘤的準確率可高達 85%以上。在 2019 年，Vijayarajeswari 學者等人為了分類正常與異常的乳房攝影影像，利用霍夫變換偵測並提取影像的特徵，再將其用於訓練 SVM 分類乳房攝影影像[12]。

在 2018 年，Zemouri 學者等人提出基於 DNN 的架構，用於預測 Oncotype DX (ODX)乳腺癌檢測管腔 B 型乳癌的復發風險程度[13]，並在測試結果有良好表現。因此本計畫希望透過隨機森林、SVM 與 DNN 方法用於分類鈣化點良惡。

綜合上述的文獻回顧及探討，本計畫發現僅使用深度學習方法來抽取與篩選醫學影像中的特徵，在疾病分類與預測的準確率方面，仍具有提升的空間，而且目前研究大多是運用 Radiomics 進行腫瘤分割與抽取腫瘤特徵，鮮少有學者將 Radiomics 應用於抽取微小鈣化點特徵，因此本計畫希望結合 Radiomics 以及深度學習的方法抽取出乳房攝影影像中的微小鈣化點特徵，之後透過統計學方法找出具代表性的鈣化點特徵，最後運用機器學習與深度學習方法辨識良性與惡性鈣化點，以此來輔助醫師診斷。

(四)研究方法及步驟

本計畫利用 Radiomics、統計學、機器學習與深度學習方法建置出「基於 DL-RMS 的特徵抽取法與統計特徵篩選之乳房鈣化點良惡分析系統」,DL-RMS(Deep Learning Radiomics)分為 Radiomics 及 AutoEncoder 方法抽取乳房鈣化點特徵；ANOVA、MANOVA 及獨立樣本 t 檢定篩選具有代表性的特徵和 DNN、SVM 及隨機森林分類鈣化點良惡，系統流程圖如圖 2 所示，主要流程可分為 6 部份，以下將針對各部分詳細說明。

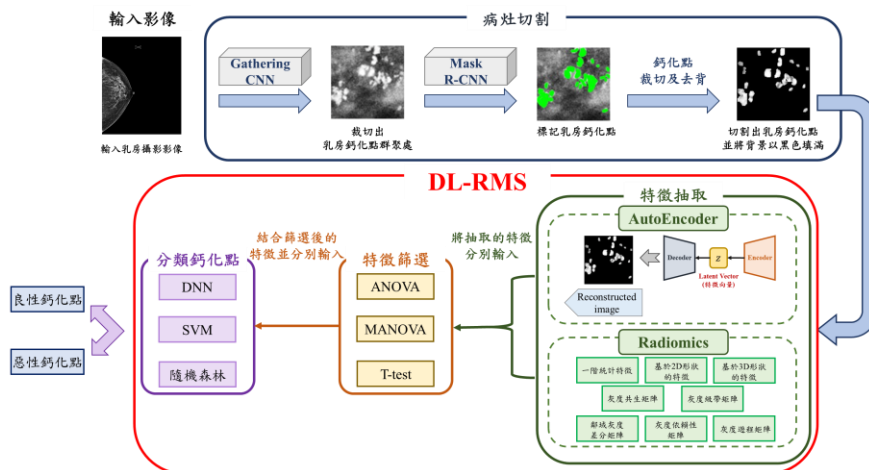


圖 2、系統流程圖

1. 輸入乳房攝影影像

本計畫與中山醫學大學附設醫院(Chung Shan Medical University Hospital)醫學影像部合作，由專業醫師提供 1586 張乳房攝影影像作為影像資料集，其中包含每位病患的乳房上下側影像(CC view)，如圖 3(a)所示，及左右側影像(MLO view)，如圖 3(b)所示。本計畫將此資料集中 80%的影像資料作為訓練影像，而 20%的影像資料作為測試影像。



圖 3、(a)為上下側影像(CC View)，(b)為左右側影像(MLO View)

2. Gathering CNN 偵測鈣化點群聚處

Gathering CNN 分為兩部分，分別為全卷積神經網路(Fully Convolutional Neural Networks, FCN)及 ISODATA(Iterative Self-Organizing Data Analysis Techniques Algorithm)。為了自動找出乳房攝影影像中的鈣化點群聚處，本計畫會先透過醫師提供的細針定位影像得知確切的鈣化點群聚位置，以作為模型分類群聚與非群聚鈣化點影像的依據，將乳房攝影影像以 128x128 的大小擷取出鈣化點群聚處。接著為了因應不同設備拍攝影像導致影像對比度不佳，以及乳房鈣化點輪廓不明顯而影響模型辨識結果的問題，本計畫將乳房攝影影像進行銳利化處理，如圖 4 所示。此外，為了區分鈣化點群聚與非群聚處，本計畫將擷取出的群聚鈣化點影像數量作為標準，再以相同數量擷取影像中非群聚處，如：乳腺、乳頭及標籤等，接著將影像及其群聚鈣化點座標範圍輸入 FCN 模型進行分類群聚與非群聚的訓練，以產生數個候選框(Candidate Box)。最後，由於 ISODATA 分群法擁有自動調整群數的能力，因此本計畫使用 ISODATA 對 FCN 所產生的候選框進行分群，並將相同群聚的候選框融合，以找出真正的群聚鈣化點數量及其在影像中的確切位置。

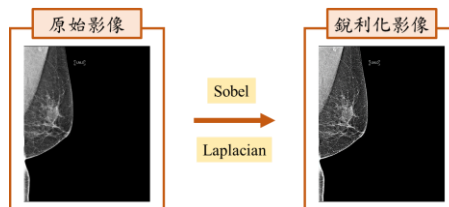


圖 4、乳房攝影影像銳利化處理之示意圖

3. Mask R-CNN 切割乳房鈣化點影像

在擷取出鈣化點群聚處影像後，本計畫首先會將 Gathering CNN 切出的群聚鈣化點影像作為 Mask R-CNN 訓練及測試的影像，如圖 5(a)所示，接著對鈣化點群聚影像中的微小鈣化點，依據其輪廓進行標記，作為 Mask R-CNN 訓練時的目標影像，如圖 5(b)所示，最後在 Mask R-CNN 訓練完成並且能成功切割出群聚

鈣化點影像中的鈣化點後，為避免背景干擾造成後續良惡判斷準確率不高的問題，本計畫將背景的部分以黑色填滿，如圖 5(c)所示。

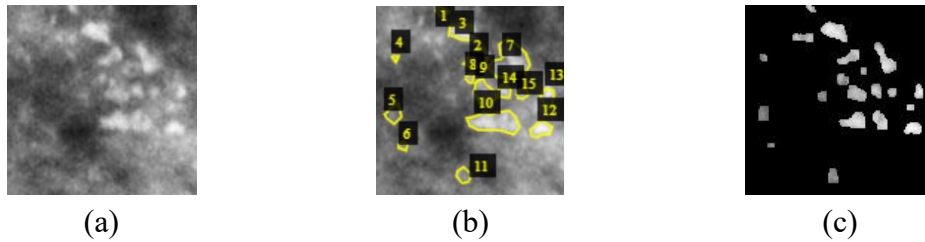


圖 5、(a)為乳房鈣化點群聚處影像，(b)為 Mask R-CNN 訓練時的目標影像及 (c)為將未標記鈣化點處的背景填為黑色之影像

4. 運用 Radiomics 與自動編碼器(AutoEncoder)抽取鈣化點群聚處特徵

在 Mask R-CNN 切割出影像中所有的鈣化點，並將背景填為黑色後，本計畫分別運用 Radiomics 與 AutoEncoder，抽取影像中的鈣化點特徵。Radiomics 是高通量的特徵抽取方法，可從每張鈣化點影像中快速提取大量的資訊，將感興趣區域(Region of Interest, RoI)影像轉為多筆具有高分辨率的特徵空間數據，可反映潛在病理及生理特徵，常應用於疾病檢測、腫瘤分型與評估疾病預後等方向，因此現今已有多篇研究使用 Radiomics 特徵抽取法從醫學影像中抽取大量特徵如腫瘤與乳腺癌特徵等[1-3]。AutoEncoder 是一種非監督式學習演算法，常用於降低高維度資料及移除影像中的雜訊，並保留原始資料中的重要特徵，且在文獻探討部份提到預先使用 AutoEncoder 對影像進行預處理將能有效提高分類或預測模型的準確率[5-6]。

Radiomics 抽取的特徵可分為 8 類，分別為一階統計特徵、基於 3D 形狀的特徵、基於 2D 形狀的特徵、灰度共生矩陣(Gray Level Co-Occurrence Matrix)特徵、灰度遊程矩陣(Gray Level Run-Length Matrix)特徵、灰度級帶矩陣(Gray Level Size Zone Matrix)特徵、灰度依賴性矩陣(Gray Level Dependence Matrix)特徵與鄰域灰度差分矩陣(Neighbor Gray Tone Difference Matrix)，並將獲取的 8 類特徵皆作為描述鈣化點的定量特徵，以獲取鈣化點群聚處的良好鈣化點資訊，為後續的鈣化點分類提供更多思路。

本計畫除了運用 Radiomics 方法大量獲取乳房鈣化點特徵外，還使用 AutoEncoder 深度學習方法自動提取乳房鈣化點的特徵，AutoEncoder 架構中包含編碼器(Encoder)和解碼器(Decoder)兩部份，其架構圖如圖 6 所示，編碼器與解碼器主要分別進行影像壓縮與解壓縮的動作，將高維度的原始群聚鈣化點影像經過編碼器後，會以低維度的特徵向量保留輸入影像中具代表性的鈣化點特徵，方便解碼器重建出相近於原始影像的輸出影像。

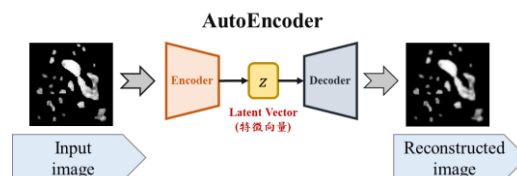


圖 6、AutoEncoder 架構圖

AutoEncoder 模型需使用大量影像進行訓練，方能準確地從影像中自動進行鈣化點特徵抽取與選擇，而 Radiomics 方法無需訓練階段便能抽取每張影像中的大量特徵數值，但一般還需透過特徵工程的方法來選擇出具有意義的鈣化點特徵。因此本計畫使用 Radiomics 與深度學習兩種方法進行特徵抽取，以確保後續的特徵選擇或分類乳房鈣化點良惡的結果具有穩定與可靠性，並獲得更多存在乳房鈣化點影像中的資訊。

5. 使用獨立樣本 t 檢定、ANOVA 與 MANOVA 進行特徵選擇

由於 Radiomics 特徵抽取法與 AutoEncoder 模型抽取的大量鈣化點特徵中，並非所有特徵皆能用於將惡性鈣化點分類為三類的代表性特徵，因此本計畫為了找出 Radiomics 與 AutoEncoder 中最具代表性的特徵，分別使用獨立樣本 t 檢定、單因子獨立變異數分析 (Analysis of variance, ANOVA) 及多變量變異數分析 (Multivariate analysis of variance, MANOVA) 進行最佳特徵選取，除去影像中不相關 (irrelevant) 或是冗餘 (redundant) 特徵，以獲得真正鈣化點良惡分類的特徵。

獨立樣本 t 檢定是醫學領域常用的統計學方法，常用於比較兩組特徵的平均數是否呈顯著差異，因此本計畫分別從 Radiomics 與 AutoEncoder 抽取出的多類別特徵中，透過排列組合的方式各選擇一類特徵，再透過統計學的假設檢定進行兩特徵類別的比較。在假設檢定中 H_0 是指虛無假設，而 H_1 則是指對立假設， H_0 表示抽取的特徵與 Radiomics 或 AutoEncoder 的特徵類別之間無差異，反之則表示有差異，接著透過公式(1)進行統計量 t 值的計算。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (1)$$

其中， \bar{X}_1 與 \bar{X}_2 分別代表兩組樣本的平均數， n_1 與 n_2 代表兩組樣本的數量，而 S_1 和 S_2 分別代表兩組樣本的標準差，可用來判斷兩特徵類別之間是否有顯著差異。在獨立樣本 t 檢定中，顯著水準是作為假設檢定判斷的依據，用來表示觸犯第一類型錯誤 (Type I Error) 機率，通常以 α 表示，可利用邦佛洛尼校正 (Bonferroni correction) 方法修正 α 。在統計學與許多研究領域通常認為 α 為 0.05，是指假設 H_0 為真，但檢定結果卻為拒絕 H_0 的機率。而 p 值則是用來衡量整體樣本資料拒絕 H_0 的機率，表示在分佈常態分布下，大於等於以及小於等於 t 值的機率密度值 (probability density value)，當 p 值越小，統計意義越大。最後自由度 (degree of freedom, df) 是指各特徵類別的所有特徵中有多少特徵可以用來評估特徵類別的變異數，可透過公式(2)計算 Radiomics 與 AutoEncoder 兩者特徵類別的自由度大小，並依據自由度與 p 值查詢「t 值表」，以找出 t 值的臨界值，判斷 Radiomics 與 AutoEncoder 兩組特徵類別之間是否具有顯著差異。

$$df = n_1 + n_2 - 2 \quad (2)$$

接著本計畫還會採用 ANOVA 與 MANOVA 方法進行鈣化點特徵選取。ANOVA 主要用於比較兩組以上的特徵類別間是否存在平均數差異，而 MANOVA 是 ANOVA 的延伸，可用於控制整體標準，瞭解多組特徵類別在兩個以上的應變

數之間的相關性。兩者的差異在於 MANOVA 可以同時比較兩個以上的應變數，而 ANOVA 一次僅能分析一種應變數。

ANOVA 包含總離均差平方和(Sum of Squares Total, SST)，組內離均差平方和(Sum of Squares Within, SSW)與組間離均差平方和(Sum of Squares Between, SSB)三部分，SST 表示整體樣本在應變項的差異狀況，如公式(3)，其可細分為 SSW 與 SSB 兩部分，SSW 是用於衡量 Radiomics 與 AutoEncoder 的每個特徵類別內部樣本之間的差異，一般認為是隨機誤差如公式(4)，而 SSB 則是用於衡量不同特徵類別之間的差異，如公式(5)。

$$SST = \sum (X_i - \bar{X})^2 = SSW + SSB \quad (3)$$

$$SSW = \sum (X_i - \bar{X}_k)^2 \quad (4)$$

$$SSB = \sum n_k (\bar{X}_k - \bar{X})^2 \quad (5)$$

其中總特徵數為 n ，各特徵類別的特徵數為 n_k ， X_i 是指總特徵中的第 i 個特徵， \bar{X}_k 是指各特徵類別的平均數，而 \bar{X} 是總特徵值的平均數，比較 SSB 與 SSW 的比值是否高於臨界值，若數值高於臨界值則表示樣本特徵對於特徵類別有一定的區分度。而界定臨界值則要用到 ANOVA 檢定的統計量 F ，如公式(6)所示。

$$F = \frac{MSB}{MSE} = \frac{SSB/(k-1)}{SSW/(n-k)} \quad (6)$$

其中 n 為總特徵數， k 為特徵類別數，MSB 與 MSE 是分別由 SSB 與 SSW 除以其自由度計算而得，而 SSW 和 SSB 的自由度分別為 $n - k$ 和 $k - 1$ ，之後再對照「F 分布表」，依設定的顯著水準決定臨界值，判斷 Radiomics 與 AutoEncoder 各組特徵類別間是否呈顯著差異。此外，在大部分的統計推論研究中常會超過一個應變數，且應變數之間容易存在中等相關，因此僅採用 ANOVA 分析易違反 Type I Error，代表應該修正 α 方能獲得正確的顯著水準。本計畫為了確保所選擇的特徵能正確分類出良性與惡性的乳房鈣化點，還利用 MANOVA 評估 Radiomics 所獲得的大量特徵，MANOVA 與 ANOVA 的運算邏輯相似，是利用總體、組內與組間平方和與交叉相乘矩陣(Sum of Square and Cross Product)表示整體特徵在應變項的差異狀況，如公式(7)、(8)與(9)所示。

$$T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})(y_{ij} - \bar{y})' \quad (7)$$

$$W = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)' \quad (8)$$

$$B = \sum_{i=1}^k (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})' \quad (9)$$

其中 k 表示特徵類別數， n_i 代表各特徵類別的特徵數， y_{ij} 表示第 i 組特徵類別中的第 j 項， \bar{y}_i 為各特徵類別之平均數，而 \bar{y} 為總特徵值的平均數，接著同樣透過 F 檢定找出臨界值，判斷各特徵類別之間是否存在顯著差異，以統計學所獲得的數據為基礎，作為乳房鈣化點特徵選擇的依據，如公式(10)所示。

$$F = \frac{MSA}{MSE} = \frac{B/(k-1)}{W/(n-k)} = \frac{n-k}{k-1} W^{-1} B \quad (10)$$

6. 採用 DNN、SVM 及隨機森林分類鈣化點之良惡

本計畫將獨立樣本 t 檢定、ANOVA 及 MANOVA 篩選出的乳房鈣化點特徵

進行結合，並分別輸入深度神經網路(Deep Neural Networks, DNN)、支援向量機(Support Vector Machine, SVM)及隨機森林(Random Forest)三種常用於醫學影像的分類模型。首先 DNN 內部的神經網路層可分為輸入層，隱藏層和輸出層三類，且包含前向傳遞與反向傳遞兩階段，前向傳遞階段是從輸入層開始逐層往後計算，直到獲得鈣化點良惡的預測值。而反向傳遞階段會先設定目標函數來衡量實際與期望輸出值之間的誤差，再透過梯度下降法進行權重值的更新與訓練，將目標函數最小化，達到分類鈣化點良惡之目的。接著，SVM 是一種二元分類器(binary classifier)，當樣本數量不多時，分類的準確率高，且泛化能力強，適合用於乳房鈣化點影像中微鈣化點的良惡分類。最後，隨機森林可用於處理高維度的資料，且當部分特徵遺失時，仍能維持準確率，適合用於處理經過 Radiomics 抽取出的多特徵數據。本計畫預計使用上述 3 種分類演算法進行乳房鈣化點的良惡分類，若最終的分類結果優異，則能證明本計畫所提出的方法有助於提升分類模型的準確率，且獨立樣本 t 檢定、ANOVA 及 MANOVA 方法所篩選出的特徵為分類乳房鈣化點良惡的重要特徵。

(五) 實驗結果

本計畫的實驗結果將分為以下三部分進行探討，首先介紹系統開發環境與使用工具，接著介紹使用 ANOVA F 檢定計算出各 Radiomics features 的 F 值進行特徵選擇，最後介紹 SVM 模型之混淆矩陣及其相關評估指標，並討論是否經過統計學方法進行特徵選擇之 SVM 模型分類結果的差異。

1. 開發環境與工具

本計畫的訓練及測試階段皆在 Windows 系統中進行，利用 Anaconda3 搭配 sklearn 機器學習函式庫、Keras 深度學習函式庫、Pyradiomics 影像組學特徵抽取函式庫以及 Tkinter GUI Toolkit 實現 AI 環境的建置、影像組學特徵抽取以及 GUI 介面的設計，並透過 Python 程式語言進行開發。硬體設備使用 Intel Core i9-9820X、NVIDIA Geforce RTX 2080 Ti 和 12G RAM 的電腦，所使用的開發設備及環境如表 1 所示。

表 1、軟硬體設備

軟體	作業系統	Windows 10
	深度學習開發環境	Anaconda 3
	開發程式語言	Python
	AI 函式庫	sklearn、Keras
	GUI 套件	Tkinter
	影像組學特徵抽取函式庫	Pyradiomics
硬體	CPU 處理器	Intel Core i9-9820X
	GPU 顯卡	NVIDIA Geforce RTX 2080 Ti
	RAM 容量	12G

2. ANOVA F 檢定計算出各 Radiomics features 的 F 值並進行特徵選擇

目前一般使用 Pyradiomics 影像組學特徵函式庫進行特徵抽取時，共會計算出 8 種不同類型的特徵值，其中包含一階統計特徵、基於 2D 形狀的特徵及基於 3D 形狀的特徵等，但由於本計畫所使用的乳房鈣化點影像是屬於 2D 影像，因此在特徵抽取階段並未計算 3D 與相關的特徵值。本計畫透過 Pyradiomics 影像組學特徵函式庫針對乳房微鈣化點的區域進行特徵抽取，共抽取 102 種特徵，接著採用 ANOVA 來評估多種特徵值之間的差異，透過 ANOVA 的 F 檢定統計量計算出各特徵值的 F 值，如表 2 所示，可用於判斷各特徵值是否適合用來分類良性與惡性乳房鈣化點。

表 2、各 Radiomics features 的 F 值

<i>First order statistics Features</i>			<i>Shape-based (2D) Features</i>		
0	10Percentile	0.295932	18	Elongation	1.969700
1	90Percentile	3.709242	19	Major Axis Length	17.316508
2	Energy	11.702807	20	Maximum 2D diameter	33.923449
3	Entropy	8.694413	21	Mesh Surface	17.029646
4	Interquartile Range	1.500421	22	Minor Axis Length	20.252432
5	Kurtosis	0.660259	23	Perimeter	31.779611
6	Maximum	12.038673	24	Perimeter to Surface ratio	9.465856
7	Mean Absolute Deviation (MAD)	2.295969	25	Pixel Surface	18.026436
8	Mean	1.306950	26	Sphericity	42.740264
9	Median	1.053885	<i>Gray Level Co-occurrence Matrix (GLCM) Features</i>		
10	Minimum	5.300412	27	Autocorrelation	8.109272
11	Range	20.887386	28	Cluster Prominence	2.769930
12	Robust Mean Absolute Deviation (rMAD)	1.622637	29	Cluster Shade	1.508821
13	Root Mean Squared (RMS)	1.455684	30	Cluster Tendency	3.121143
14	Skewness	0.050491	31	Contrast	3.721265
15	TotalEnergy	11.702807	32	Correlation	22.638504
16	Uniformity	9.166207	33	Difference Average	3.305123
17	Variance	1.589201	34	Difference Entropy	0.286059
<i>Gray Level Run Length Matrix Feature</i>			35	Difference Variance	1.329438

51	Gray Level Non-Uniformity (GLN)	23.926473	36	Inverse Difference (ID)	2.913301
52	Gray Level Non-Uniformity Normalized (GLNN)	8.454810	37	Inverse Difference Moment (IDM)	3.031177
53	Gray Level Variance (GLV)	1.464435	38	Inverse Difference Moment Normalized (IDMN)	23.260212
54	High Gray Level Run Emphasis (HGLRE)	10.128861	39	Inverse Difference Normalized (IDN)	15.359914
55	Long Run Emphasis (LRE)	0.001258	40	Informational Measure of Correlation (IMC) 1	6.994711
56	Long Run High Gray Level Emphasis (LRHGLE)	0.209777	41	Informational Measure of Correlation (IMC) 2	15.570703
57	Long Run Low Gray Level Emphasis (LRLGLE)	2.456071	42	Inverse Variance	1.880141
58	Low Gray Level Run Emphasis (LGLRE)	9.463949	43	Joint Average	8.635873
59	Run Entropy (RE)	25.638934	44	Joint Energy	6.295740
60	Run Length Non-Uniformity (RLN)	26.268580	45	Joint Entropy	7.356590
61	Run Length Non-Uniformity Normalized (RLNN)	8.125485	46	Maximal Correlation Coefficient (MCC)	8.731826
62	Run Percentage (RP)	4.729306	47	Maximum Probability	5.140978
63	Run Variance (RV)	0.029996	48	Sum Average	8.635708
64	Short Run Emphasis (SRE)	5.895822	49	Sum Entropy	16.934830
65	Short Run High Gray Level Emphasis (SRHGLE)	7.449175	50	Sum of Squares	2.307877
66	Short Run Low Gray Level Emphasis	12.006420	Gray Level Size Zone Matrix (GLSZM) Features		
Neighbouring Gray Tone Difference Matrix (NGTDM) Features			67	Gray Level Non-Uniformity (GLN)	23.926473

83	Busyness	6.070743	68	Gray Level Non-Uniformity Normalized (GLNN)	8.454810
84	Coarseness	0.589104	69	Gray Level Variance (GLV)	1.464435
85	Complexity	5.074996	70	High Gray Level Zone Emphasis (HGLZE)	10.128861
86	Contrast	5.977543	71	Large Area Emphasis (LAE)	0.001258
87	Strength	14.427952	72	Large Area High Gray Level Emphasis (LAHGLE)	0.209777
Gray Level Dependence Matrix (GLDM) Features			73	Large Area Low Gray Level Emphasis (LALGLE)	2.456071
88	Dependence Entropy (DE)	33.585059	74	Low Gray Level Zone Emphasis (LGLZE)	9.463949
89	Dependence Non-Uniformity (DN)	17.237819	75	Size-Zone Non-Uniformity (SZN)	26.268580
90	Dependence Non-Uniformity Normalized (DNN)	15.538451	76	Size-Zone Non-Uniformity Normalized (SZNN)	8.125485
91	Dependence Variance (DV)	9.703390	77	Small Area Emphasis (SAE)	5.895822
92	Gray Level Non-Uniformity (GLN)	9.291882	78	Small Area High Gray Level Emphasis (SAHGLE)	7.449175
93	Gray Level Variance (GLV)	1.395565	79	Small Area Low Gray Level Emphasis (SALGLE)	12.006420
94	High Gray Level Emphasis (HGLE)	8.582999	80	Zone Entropy (ZE)	25.638934
95	Large Dependence Emphasis (LDE)	3.944425	81	Zone Percentage (ZP)	4.729306
96	Large Dependence High Gray Level Emphasis (LDHGLE)	5.862991	82	Zone Variance (ZV)	0.029996

97	Large Dependence Low Gray Level Emphasis (LDLGLE)	3.553417			
98	Low Gray Level Emphasis (LGLE)	9.168657			
99	Small Dependence Emphasis (SDE)	6.578853			
100	Small Dependence High Gray Level Emphasis (SDHGLE)	5.610242			
101	Small Dependence Low Gray Level Emphasis (SDLGLE)	13.664867			

F 值能用來評估特徵值間的組間差異，當 F 值越大，則代表特徵值之間的差異性越顯著，能用來區分良性與惡性鈣化點。圖 7 為各 Radiomics feature 所計算出的 F 值直方圖，可發現圖中的峰值大約界於 20 至 40 左右，而高於 35 的特徵值只剩 1 種，因此本計畫則分別篩選出 F 值大於 20、F 值大於 25 以及 F 值大於 30 的特徵值輸入至 SVM 模型進行乳房鈣化點的良好分類。

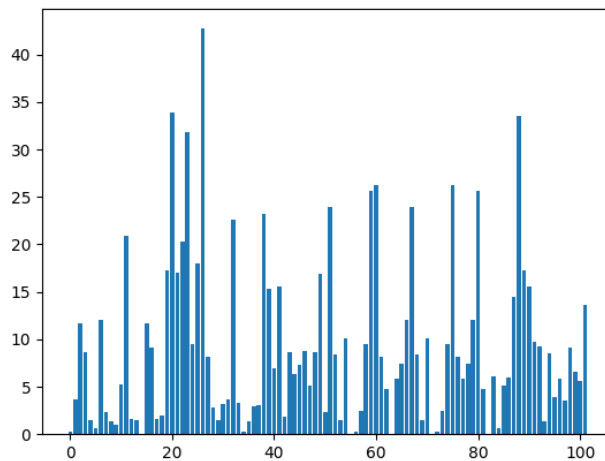


圖 7、各 Radiomics feature F 值的直方圖

將抽取出的 Radiomics features 分別以 F 值大於 20、F 值大於 25 以及 F 值大於 30 進行篩選的結果如表 3 所示，從表中可得知共有 14 個特徵 F 值大於 20，而這 14 個特徵分別由 6 種不同類型的特徵值所組成；又有 8 個特徵 F 值大於 25，且分別由 4 種不同類型的特徵值所組成；最後，剩下 4 個特徵 F 值大於 30，由 2 種不同類型的特徵值所組成，本計畫將分別以上述不同的三個

條件輸入至 SVM 模型進行實驗，以此得出對於分類乳房鈣化點良惡具影響力的特徵。

表 3、根據不同 F 值所篩選出的 Radiomics features

	特徵數量	篩選後的特徵
F 值>20	14	<ul style="list-style-type: none"> ✓ First order statistics Features: Range ✓ Shape-based (2D) Features: Maximum 2D diameter, Minor Axis Length, Perimeter, Sphericity ✓ Gray Level Co-occurrence Matrix (GLCM) Features: Correlation, Inverse Difference Moment Normalized (IDMN) ✓ Gray Level Run Length Matrix (GLRLM) Feature: Gray Level Non-Uniformity (GLN), Run Entropy (RE), Run Length Non-Uniformity (RLN) ✓ Gray Level Size Zone Matrix (GLSZM) Features: Gray Level Non-Uniformity (GLN), Size-Zone Non-Uniformity (SZN), Zone Entropy (ZE) ✓ Gray Level Dependence Matrix (GLDM) Features: Dependence Entropy (DE)
F 值>25	8	<ul style="list-style-type: none"> ✓ Shape-based (2D) Features: Maximum 2D diameter, Perimeter, Sphericity ✓ Gray Level Run Length Matrix (GLRLM) Feature: Run Entropy (RE), Run Length Non-Uniformity (RLN) ✓ Gray Level Size Zone Matrix (GLSZM) Features: Size-Zone Non-Uniformity (SZN), Zone Entropy (ZE) ✓ Gray Level Dependence Matrix (GLDM) Features: Dependence Entropy (DE)
F 值>30	4	<ul style="list-style-type: none"> ✓ Shape-based (2D) Features: Maximum 2D diameter, Perimeter, Sphericity ✓ Gray Level Dependence Matrix (GLDM) Features: Dependence Entropy (DE)

3. 有無進行特徵選擇之 SVM 模型分類結果與討論

為了證明採用統計學方法進行特徵選擇是能夠有效提升 SVM 模型的分類準確率，本計畫一共包含 4 種實驗，分別將 3 種不同特徵選擇方法後篩選出的特徵值及全部特徵值分別輸入至 SVM 模型進行訓練。另外，由於 SVM 模型可選擇搭配不同 kernel 函數來提高模型分類效能，因此本計畫選擇了 4 種較為常見的 kernel 作為 SVM 模型的 kernel 函數，分別為 Linear kernel、Polynomial kernel、

Gaussian Radial Basis Function kernel (RBF)及 Sigmoid kernel，經由非線性投影後，在更高維度空間中區分出良性與惡性乳房鈣化點。

表 4 是將全部 Radiomics features 輸入至 SVM 模型分類乳房鈣化點良惡的測試結果，本計畫透過混淆矩陣計算出準確率、精準度、敏感度及特異度等 4 項指標來評估 SVM 模型分類良性與惡性鈣化點的狀況。從表 4 中可知，未經過特徵選擇步驟便直接使用 SVM 模型分類良性與惡性鈣化點的準確率約介在 60%至 75%之間，其中使用 Polynomial kernel 的 SVM 模型準確率、精準度、敏感度及特異度分別為 72.34%、53.33%、57.14%、及 78.79%，在準確率及特異度部分是 4 種 kernel 之間最高的，表示 SVM 搭配 Polynomial kernel 在分類良性與惡性微鈣化點的效能較佳，因此後續主要會以 SVM 搭配 Polynomial kernel 的實驗結果來與經過特徵選擇的實驗結果進行比較。

表 4、未經過特徵選擇時，SVM 分類乳房鈣化點良惡之測試結果

	測試結果			
	準確率	精確度	敏感度	特異度
Linear kernel	68.08%	50.00%	60.00%	71.88%
Polynomial kernel	72.34%	53.33%	57.14%	78.79%
RBF kernel	63.83%	52.63%	55.56%	68.97%
Sigmoid kernel	61.70%	57.90%	52.38%	69.23%

當 F 值大於 20 時，會從原先的 102 種特徵值中篩選出 14 種特徵值，其實驗結果如表 5 所示。而此時 SVM 搭配 Polynomial kernel 的準確率、精準度、敏感度及特異度分別為 85.11%、72.22%、81.25%及 84.38%，實驗結果皆優於未經過特徵篩選的實驗結果，證明使用統計學方法進行特徵選擇，找出具代表性特徵的步驟確實能提高 SVM 模型分類鈣化點良惡的準確率。

表 5、當 F 值>20 時，SVM 分類乳房鈣化點良惡之測試結果

	測試結果			
	準確率	精確度	敏感度	特異度
Linear kernel	82.98%	66.67%	76.92%	85.29%
Polynomial kernel	85.11%	73.68%	82.35%	83.87%
RBF kernel	78.72%	66.67%	75.00%	80.65%
Sigmoid kernel	78.72%	61.11%	78.57%	78.79%

當 F 值大於 25 時，會從原先的 102 種特徵值中篩選出 8 種特徵值，其實驗結果如表 6 所示。而此時 SVM 搭配 Polynomial kernel 的準確率、精準度、敏感度及特異度分別為 80.85%、68.75%、78.57%及 84.38%，表示篩選出具代表性的 8 個特徵值是能提升 SVM 分類良性與惡性鈣化點的準確率，但僅採用 8 種特徵值的實驗結果略差於採用 14 種特徵值的實驗結果，推測可能是因為具代表性特徵值的所屬類別只有 4 種，所以對於 SVM 的分類準確率有所影響。

表 6、當 F 值>25 時，1SVM 分類乳房鈣化點良惡之測試結果

	測試結果			
	準確率	精確度	敏感度	特異度
Linear kernel	76.60%	64.71%	68.75%	80.65%
Polynomial kernel	80.85%	68.75%	78.57%	84.38%
RBF kernel	78.72%	72.22%	72.22%	82.76%
Sigmoid kernel	72.34%	75.00%	72.00%	72.73%

當 F 值大於 30 時，會從原先的 102 種特徵值中篩選出 4 種特徵值，其實驗結果如表 7 所示。而此時 SVM 搭配 Polynomial kernel 的準確率、精確度、敏感度及特異度分別為 76.60%、61.11%、73.33%及 78.13%。從實驗結果可知，採用 14 種或是 8 種具代表性特徵值進行鈣化點良惡分類結果皆優於採用 4 種具代表性特徵值的實驗結果。

表 7、當 F 值>30 時，SVM 分類乳房鈣化點良惡之測試結果

	測試結果			
	準確率	精確度	敏感度	特異度
Linear kernel	74.47%	61.11%	68.75%	77.42%
Polynomial kernel	76.60%	61.11%	73.33%	78.13%
RBF kernel	70.21%	55.55%	62.50%	74.19%
Sigmoid kernel	63.83%	50.00%	52.94%	70.00%

從上述的實驗結果與討論可得知透過 Radiomics 抽取出大量不同類型的微鈣化點特徵值之後，再經過特徵選擇步驟找出具代表性的特徵確實是能有效提高 SVM 模型分類準確率、精確度、敏感度及特異度，但目前實驗結果的測試準確率仍未達 90%以上，推論可能的原因是進行特徵選擇時，未能考量到所有的特徵類別，導致訓練出的模型泛化能力不佳，因此後續若能針對特徵類別的多樣性做調整，相信能大大改善準確率。

(六) 結論

在臨床上，由於乳房鈣化點微小不易觀察，一般會透過具侵入性的細針定位手術檢測來確定鈣化點群聚處中的良性與惡性鈣化點，容易讓病患產生不適感。因此本計畫致力於研究鈣化點的特徵抽取及特徵選擇方法，使用自動抽取乳房攝影影像中鈣化點群聚處的鈣化點特徵，再使用統計學方法找出具代表性的鈣化點特徵，以達到自動分析良性與惡性群聚鈣化點之目標。本計畫在分類鈣化點良惡部分的準確率可達到 85%，期望未來能有效輔助醫師診斷疾病以及減輕病患負擔。

(七) 參考文獻

- [1] L. Losurdo, A. Fanizzi et al, "Radiomics Analysis on Contrast-Enhanced Spectral Mammography Images for Breast Cancer Diagnosis: A Pilot Study," *Entropy*, vol. 21, pp. 11, 2019.
- [2] A. Ibrahim, M. Vallières et al, "Radiomics Analysis for Clinical Decision Support in Nuclear Medicine," *Seminars in Nuclear Medicine*, vol. 45, pp. 438-449, 2019.
- [3] Marco Caballo et al, "Deep learning-based segmentation of breast masses in dedicated breast CT imaging: Radiomic feature stability between radiologists and artificial intelligence," *Computers in Biology and Medicine*, vol. 118, pp. 2020.
- [4] 魏嘉芯。基於 BI-CNN 進行婦女乳房鈣化點群聚偵測及良惡預測。科技部大專學生研究計畫(計畫編號：109-2813-C-040-014-E)
- [5] Laura Macías-García et al, "A study of the suitability of autoencoders for preprocessing data in breast cancer experimentation," *Journal of Biomedical Informatics*, vol. 72, pp. 33-44, 2017.
- [6] Mesut Toğaçar et al, "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders," *Medical Hypotheses*, vol. 135, 2020.
- [7] Chih-Yang Hsu et al, "Radiomics Features Differentiate Between Normal and Tumoral High-Fdg Uptake," *Scientific Reports*, vol. 8, 2018.
- [8] C. S. Ling et al, "Shape Based Image Retrieval System for MRI Spine," 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI), pp. 1-6, 2017.
- [9] Daisy Das et al, "A study on MANOVA as an effective feature reduction technique in classification of childhood medulloblastoma and its subtypes," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, pp. 16, 2020.
- [10] M. Z. Alam et al, "A Random Forest based predictor for medical data classification using feature ranking," *Informatics in Medicine Unlocked*, vol. 15, 2019.
- [11] Gaurav Makwana et al, "Identification of Micro-calcification in Mammogram for Breast Cancer Analysis using SVM Classifier," *International Journal of Engineering & Technology*, vol. 7, no. 2.16, pp. 29-32, 2018.
- [12] R. Vijayarajeswari et al, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement*, vol. 146, pp. 800-805, 2019.
- [13] R. Zemouri et al, "Constructive Deep Neural Network for Breast Cancer Diagnosis," *IFAC-PapersOnLine*, vol. 51, pp. 98-103, 2018.