

長庚大學資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

Chang Gung University

Master Thesis

應用機器學習方法分析尿流速圖判斷排尿功能的進退
步

Using Machine Learning Methods to Interpret the
Uroflowmetry to Determine the Progress of Urination
Function

研究生：鍾蕙如

Graduate Student: Huei-Ru, Chung

指導教授：吳世琳 博士

林俊淵 博士

Advisor: Shih-Lin Wu, Ph.D.

Chun-Yuan Lin, Ph.D.

中華民國 110 年 10 月

October 2021

資 長
訊 庚
工 大
程 學
學 系

碩士論文

應用機器學習方法分析尿流速圖判斷排尿功能

鍾蕙如

110.
10.



長庚大學碩（博）士學位論文
指導教授推薦書

資訊工程 系（所） 鍾蕙如 君所提之論文

應用機器學習方法分析尿流速圖判斷排尿功能的進退步 (題目)

係由本人指導撰述，同意提付審查。

指導教授 吳世雄 林俊淵 (簽名)

110 年 10 月 25 日

長庚大學碩（博）士學位論文
口試委員會審定書

資訊工程學系 研究所 鍾蕙如 君所提之論文

應用機器學習方法分析尿流速圖判斷排尿功能的進退步

Using Machine Learning Methods to Interpret the Uroflowmetry to
Determine the Progress of Urination Function

係本委員會審議，認為符合碩（博）士資格標準。

論文口試委員會

召集人 趙志民 (簽名)

委員 林俊瑋 _____

林友翔 _____

吳世村 _____

系主任、所長 陳仁暉

中華民國 110 年 10 月 25 日

致謝

首先感謝指導老師吳世琳教授，在攻讀碩士班期間指引我正確的研究方向、並在課業上與論文上耐心指導，真的非常感謝老師，此外，也感謝共同指導的林俊淵教授、口試委員林友翔醫師和趙志民教授在口試期間給予的意見，讓論文內容更加趨於完善。

在論文撰寫期間受到許多同學還有學長的幫助，感謝研究室的各個夥伴在我學習過程中的幫忙，特別感謝將這個研究交接給我的楊俊學長，分享給我很多有關機器學習的經驗和知識，也謝謝資工系的大家，在這兩年的碩士生活，為我帶來許多美好的回憶。最後要感謝一直支持我的家人和朋友，因為有你們的默默支持與鼓勵，我才能夠完成學業，朝向人生的下個階段邁進。

中文摘要

目前下泌尿道問題是常見的疾病，例如最常見的前列腺肥大，進一步導致排尿困難。頻尿、排尿緩慢、尿量過小、急迫性尿失禁、殘留排尿、膀胱炎症...等等。這些疾病大多可以通過尿流速圖來診斷病因。

然而，對於同一位多次回診的病患，醫師會需要一個系統去判斷該病患在治療過後是否有改善。這是我們解決這個問題的研究動機，為了客觀地解釋這些數據並給出合理的進步分數。我們提出了一種基於人工智慧的解決方案來解決這個問題。我們使用自動編碼器對所有的尿流速圖進行特徵提取，然後用非監督式學習層次聚類對資料進行分群。並藉由醫師的協助確立各群與各群之間的變化分數，通過與泌尿科醫師評分的比較，我們的系統達到了 90% 的準確率。

關鍵詞——自動編碼器，卷積神經網絡，無監督學習，尿流速圖，層次聚類

Abstract

At present, lower urinary tract problems are common diseases, such as the most common prostatic hypertrophy, which further leads to dysuria. Frequent urination, slow urination, too little urine output, urge incontinence, residual urination, bladder inflammation... etc. Most of these diseases can be diagnosed by urine flow chart.

However, for the same patient who has returned for multiple visits, the physician will need a system to determine whether the patient has improved after treatment. This is our research motivation to solve this problem, in order to objectively explain these data and give reasonable progress scores. We propose a solution based on artificial intelligence to solve this problem. We use an autoencoder to extract features from all urine flow velocity maps, and then use unsupervised learning hierarchical clustering to group them. And with the help of physicians to establish the change score between each group and each group. Then calculate the pairs before and after treatment, find the most frequent paired categories and mark scores. By comparing with the scores of urologists, our system achieves an accuracy rate of 90%.

Keywords - Autoencoder, Convolutional Neural Network, Unsupervised Learning, Uroflowmetry, Hierarchical Clustering

目錄

指導教授推薦書.....	
口試委員會審定書.....	
致謝	iii
中文摘要	iv
Abstract.....	v
圖目錄.....	ix
表目錄	xi
第一章 緒論.....	- 1 -
1.1 研究背景與動機.....	- 1 -
1.2 研究目的.....	- 2 -
1.3 全文架構.....	- 3 -
第二章 相關研究	- 4 -
2.1 層次聚類(Hierarchical Clustering)	- 5 -
2.2 自動編碼器(Autoencoder)	- 8 -
第三章 研究方法	- 10 -
3.1 設計過方法.....	- 12 -

3.2 資料取得.....	- 14 -
3.3 資料重建(資料預處理).....	- 15 -
3.4 泌尿科醫師標注資料	- 16 -
3.5 圖片二值化處理	- 17 -
3.6 位移尿流速圖.....	- 18 -
3.7 自動編碼器(特徵提取).....	- 20 -
第四章 系統評估與比較.....	- 27 -
4.1 無監督式模型效能評估	- 28 -
4.1.1 外部資料介紹	- 29 -
4.1.2 外部資料預處理	- 30 -
4.1.3 統計相似度準確率計算	- 30 -
4.1.4 kappa 值與準確率計算	- 31 -
4.1.5 準確率 (Accuracy)	- 32 -
4.1.6 層次聚類鏈接方式效能評估	- 33 -
4.1.7 DBSCAN 與層次聚類效能比較.....	- 35 -
4.2 系統評估.....	- 36 -

4.2.1 圖片比例與解析度對於自動編碼器特徵提取的影響..	- 37 -
4.2.2 最適特徵維度	- 39 -
第五章 結論	- 42 -



圖目錄

圖 1、層次聚類結構圖	-5-
圖 2、自動編碼器結構	-8-
圖 3、為 SEISMIC SIGNALS 時頻圖	-9-
圖 4、系統架構圖	-10-
圖 5、尿流速圖 Q-MAX 示意圖	-13-
圖 6、尿流速圖 Q-MAX 示意圖	-13-
圖 7、MMS 操作介面	-14-
圖 8、尿流速圖重建結果	-15-
圖 9、原始圖片	-17-
圖 10、經過二值化的預處理後圖片	-17-
圖 11、原始圖片	-18-
圖 12、經過位移圖片	-18-
圖 13、模型一訓練流程圖	-21-
圖 14、模型二訓練流程圖	-22-
圖 15(A)、各類別間是示意圖與範例	-23-
圖 15(B)、各類別進退步範例	-23-
圖 16、二維特徵示意圖	-25-
圖 17、模型測試方法流程圖	-26-

圖 18、外部資料驗證層次具類鏈接方式方法流程.....	-28-
圖 19、為二維資料密度示意圖.....	-35-
圖 20、圖片比例改變與否與各種不同的解析度的比較.....	-38-
圖 21、準確率與 Kappa 在各個特徵維度的實驗結果.....	-40-
圖 22、540 特徵維度的收斂狀況.....	-40-
圖 23、844 特徵維度的收斂狀況.....	-41-



表目錄

表 1、各類別間進退步對應表格.....	-21-
表 2、KAPPA 算法範例.....	-31-
表 3、比較 MNIST 資料集在不同鏈接方式下的準確率.....	-34-
表 4、比較 ECG 資料集在不同鏈接方式下的準確率.....	-34-
表 5、尿流速圖層次聚類與 DBSCAN 的準確率比較.....	-35-
表 6、是否改變照圖片比例.....	-37-
表 7、比較本研究在不同特徵維度的準確率.....	-39-



第一章 緒論

1.1 研究背景與動機

醫師需要一個系統去判斷病患在治療後是否有改善。這是我們的研究動機，製作這個系統並給出的進退步分數(1分:表示退步；2分:表示不變；3分:表示進步)。

近年來，醫學進步很快。出現了許多新的尿道和膀胱治療方法，每位患者的治療方法也有很大不同。由於本研究主要是使用男性尿流速圖的資料，這邊介紹常見的男性泌尿系統疾病。由於組織增殖導致前列腺肥大，進一步導致排尿困難。頻尿、排尿緩慢、尿量過小、急迫性尿失禁、殘留排尿、膀胱炎症、夜尿增多...等等[1]；晚期可能出現膀胱炎症、血尿、尿滯留、排尿。尿滯留可導致腎功能障礙甚至尿毒症[2][3]。

“膀胱無力”是在長時間排尿困難、用力排尿和膀胱過度勞累後發生。可經由尿流速圖或尿流率檢查充分了解症狀[4][5]。目前，泌尿科醫師通常使用尿流速圖來確定病情是否改善。本篇論文的主要貢獻為，許多尿流速圖分類相關論文在分類形狀時常常會發生相似的圖形有些被分在鐘形 (Bell) 有些分在壓縮型 (Compression) 等等[6]，而本篇論文的解決方案就是使用無監督學習，將相似形狀分在同一個類別，以利之後協助醫師判斷病人治療後進退步的情形。

1.2 研究目的

對於同一位多次回診的病患，醫師會需要一個系統去判斷該病患在治療過後是否有改善。這是我們解決這個問題的研究動機，為了客觀地解釋這些數據並給出合理的進步分數。在進行相似圖形的尿流速圖分類時，一些醫師重視 Q-MAX(峰值)認為峰值相似就應該歸於同類型例如，峰值>60 歸類到鐘形，有些重視尿液總體積[7]，一些甚至會參考兩個以上的特徵[8][9]，例如，同時參考體積以及峰值...等等，來進行尿流速圖型的分類。由於我們缺乏醫師的專業知識，權重調整上有困難，因此使用這些特徵分類的結果並不理想，這對我們的研究是一個很大的挑戰。在我們的解決方案中，我們使用形狀做為分類特徵，因為形狀其實也包含了體積以及峰值的特徵，利用無監督式學習，將相似形狀分在同一個類別，以利之後協助醫師判斷病人治療後進步、退步或是不變的情形。

1.3 全文架構

本文架構編排如下，第一章說明研究動機和目的，第二章為相關方法的探討，第三章為本論文的研究方法，介紹本研究的系統架構，以及本研究所使用的預處理方式及所使用的模型。四章則是描述系統評估與測試，第五章為本篇論文之結論。



第二章 相關研究

在本節中，我們主要討論無監督學習、自動編碼器的相關方法，我們可以將這些研究的方法統整為以下流程：收集尿流速圖、圖片預處理、特徵擷取、依據特徵分群、利用分群結果計分。而本研究之步驟及方法會在第三章有詳細的描述。



2.1 層次聚類(Hierarchical Clustering)

本研究使用無監督式學習分群，而無監督式學習中有許多算法，例如 K-Means[10]、N-cuts[11]、譜分群 (Spectral Clustering) [12]、混合模型 (Mixture Model)[13]...等等，我們之所以使用層次聚類[14]原因為不知道要分成幾群比較合適，也沒有辦法預先計算要分成幾群，例如 K-Means 就必須要先知道 K 值才能夠做計算。而除了層次聚類之外 DBSCAN[15]也是不需要事先知道群數的無監督式學習算法，我們在實驗歷程中也曾經使用過，由於 DBSCAN 是一個基於密度的演算法，由於我們的資料密度差異大，因此效果較差。第四章會有 DBSCAN 與層次聚類效能的比較。

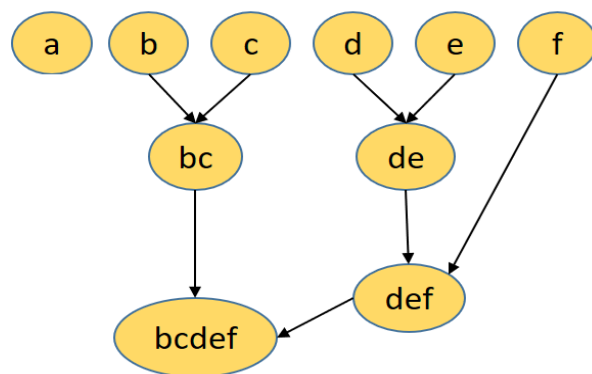


圖 1、層次聚類結構圖

本研究使用的層次聚類是一種自下而上的聚類方法，其中最初所有的數據點都是它自己的一個聚類。並應用隨著指定合併距離參數，距離近的群較早合併，隨著層次結構的向下移動，將集群配對並合併，如圖 1 所示。

而層次聚類有 4 種常用的鏈接方式分別是 AC-Single Linkage、AC-Complete Linkage、AC-Average linkage、AC-Ward linkage，

Single Linkage 群聚與群聚間的距離可以定義為不同群聚中最接近兩點間的距離，Complete Linkage 群聚間的距離定義為不同群聚中最遠兩點間的距離，Average linkage 群聚間的距離則定義為不同群聚間各點與各點間距離總和的平均，Ward linkage 群聚間的距離定義為在將兩群合併後，各點到合併後的群中心的歐幾里得距離平方和 [14]。

由於我們的資料中有離群值(outlier)，離群值尋找的部份嘗試應用統計方式與 DBSCAN 尋找，但是 DBSCAN 的效果並不好，原因為 DBSCAN 是運用密度的方式進行計算，而根據我們第四章的說明我們的資料密度差異大，運用 DNSCAN 來做分群或是離群值的尋找，結果都不理想因此決定使用統計方式尋找離群值，首先會先對資料進行常態性檢定，確認資料來自於常態分布之後，再計算第一四分位數 (Q1)、第三四分位數 (Q3) 與四分位距 (IQR)，若資料超出 $Q1 - 1.5 * IQR$ 到 $Q3 + 1.5 * IQR$ 這個範圍，我們就將其定義為離群值。而 Complete linkage 和 Average linkage 比較容易產生「齊頭並進」的效果，對於離群值容易忽略，Single linkage 傾向產生較於緊緻(compact)的群聚，會在群聚的過程中產生「大者恆大」

的效果，較不適合我的資料，因此本研究使用的是 Ward linkage，而點對點的計算方式是使用歐幾里得距離來做計算，由於沒有去指定分成幾群，因此需要用 distance_threshold 參數來區分距離多遠就不能進行群與群的合併，因此並不需要指定 K，就能完成分群。



2.2 自動編碼器(Autoencoder)

本文採用自動編碼器以類神經網路建立模型[13]，建模過程中的神經網路的網路結構去提取圖形特徵在以無監督學習的方式學習有效的編碼特徵。自動編碼器是深度學習裡面的一種演算法，我們在做 Training 的時候會提供給神經網路一個 Ground Truth 的答案去算 Loss，在訓練自動編碼器的時候有別於訓練其他類神經網路需要使用人工的方式來對資料進行標記做為 Ground Truth，自動編碼器則是直接使用資料本身來當它的 Ground Truth。

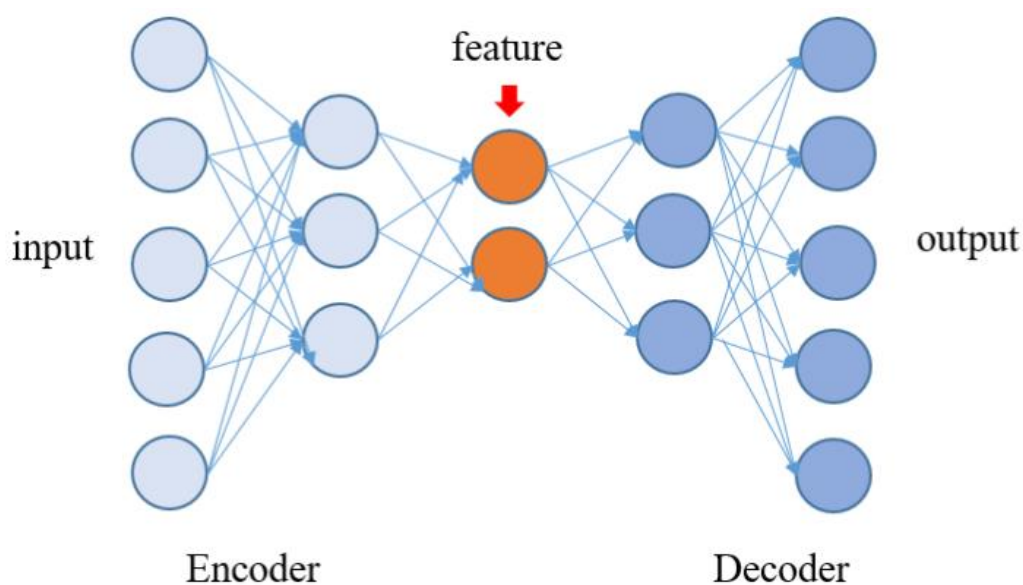


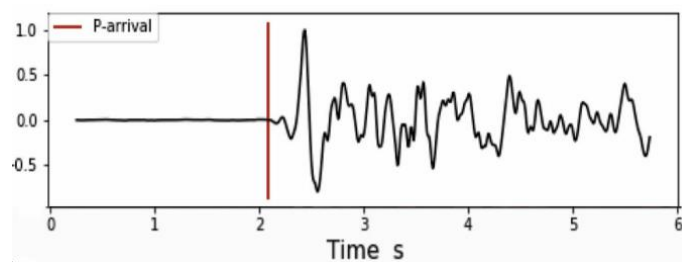
圖 2、自動編碼器結構

如圖 2 所示自動編碼器的輸入與輸出會呈現出一個鏡像對稱的結構，這樣當資料傳遞到網路中最壓縮層 (Most Compression Layer)

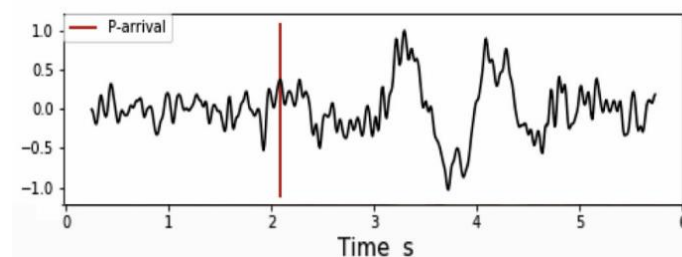
時，必然有大量的資訊會遺失；然而網路的後半部又能用從最壓縮層得到的少量資訊來還原出原始資料。

於是前半部的網路就是一個編碼器 (Encoder)；後半部的網路就是解碼器 (Decoder)，而整個網路其實就是自動在做一個編碼與解碼的動作。

參考文獻[16]也是使用深度神經網絡自動編碼器對地震數據 Seismic Signals 提取適量特徵維度後放入無監督學習模型中訓練，與本研究結構類似，資料也是曲線的型態，圖 3 為該篇論文輸入資料 Seismic Signals 時頻圖。



(a)為震源地震時頻圖，紅線則為 P 波抵達的時間



(b)為地震波經過遠距離傳導的地震時頻圖，紅線則為 P 波抵達的時間
圖 3、Seismic Signals 時頻圖。

第三章 研究方法

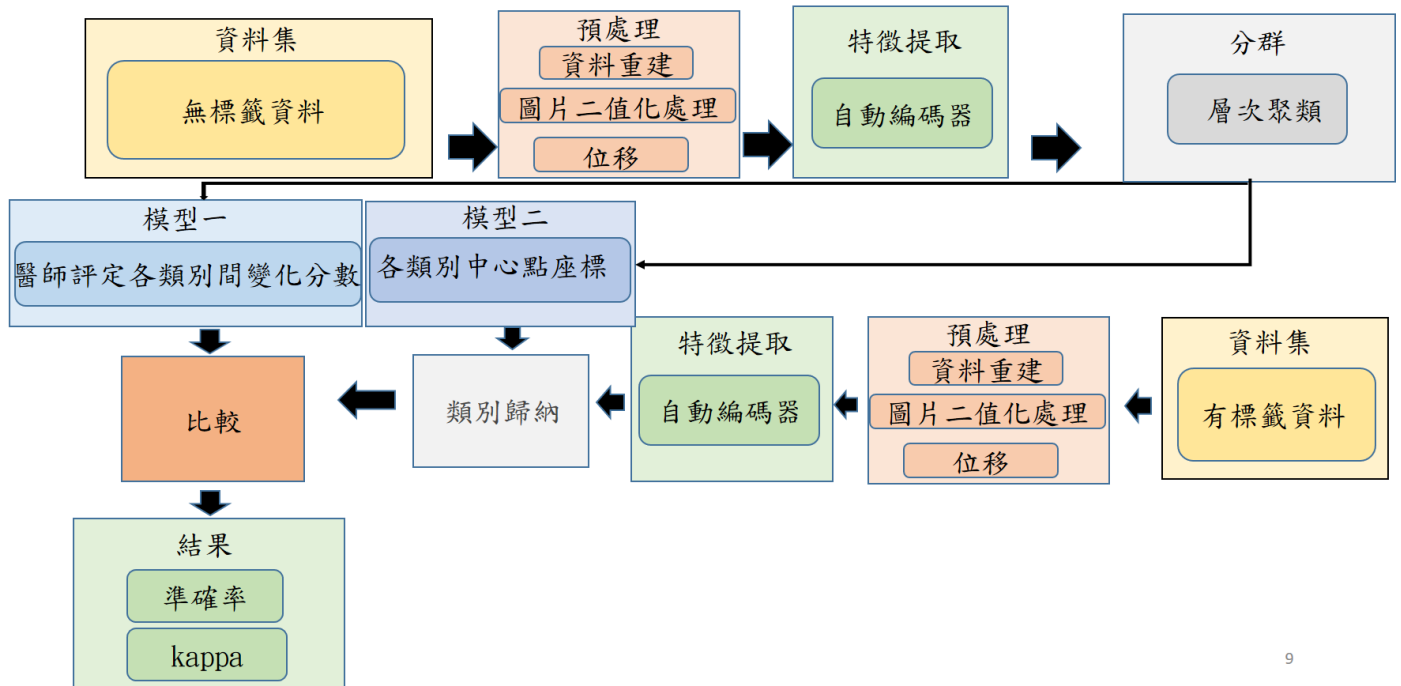


圖 4、系統架構圖

對於同一位多次回診的病患，醫師需要一個系統去判斷該病患在治療過後是否有改善。完成系統並以兩張圖能判斷進退步，這是我們解決的問題。

我們為了將大量的無標註資料分類因此採用無監督學習的方式進行分群，根據任意兩分群間請醫師標註分數。計算每群的中心點並儲存。新資料進入模型時就不需要再進行分群只需要依據距離大小對應到所屬的群中心點，這是一個十分節省運算資源與人力的方法。

第三章架構編排如下，先說明我們遇到的問題及解決方法與特色，並說明設計過的方法，接著說明圖 4 的本研究的方法流程與介紹本

研究的系統的架構，以及本研究所使用的資料取得方法與預處理方式及所使用的模型。



3.1 設計過方法

本研究方法經過不斷地摸索改良，經歷了一些歷程才定下目前的流程，之前有試過使用孿生類神經網路 (Siamese Neural Networks) [17]去比較 2 張圖片的相似度去做訓練，但是結果並不理想，原因可能是因為有標注的資料太少以及應用方式不適合所導致，孿生類神經網路大多都是應用在判別圖片相似度上，經常應用在人臉辨識 [18]，因此可能不適合用在曲線資料的變化判別。由於我們的特徵眾多，之前也有嘗試運用 PCA、LDA [19]...等等方式進行降維，但是這類型的線性降維，並不適用於圖片資料，而我們的尿流速圖是圖片資料，所以結果並不好，因此本研究應用自動編碼器進行特徵提取。

我們也有嘗試過監督式學習做分類，當時是使用 SVM (Support Vector Machine) [20]與隨機森林(Random Forest)[21]作為分類模型並進行特徵選擇，依據醫師兩張圖比較變好或是變差的標籤，使用 2 筆資料的特徵去做相乘相除或是相加減...等等方式作為特徵去做分類，當時使用了很多特徵，Q-Max (峰值)、體積、開始到峰值的時間、排尿時間...等等。使用過很多特徵，並依據”商”的大小去判斷每個特徵的重要性，最後結果不太理想，因為我們沒有醫師的專業知識，無法得知在眾多特徵中，哪些特徵是比較重要的，是否還有

其他特徵沒被發現。後來在與醫師開會時了解到醫師通常較看重形狀當作主要參考特徵，因此之後轉而使用尿流速圖進行分類。

另外，使用 Q-Max 做為特徵時我們也發現會出現如圖 5、6 的問題，觀察這兩張圖的排尿狀況，由越像鐘形的排尿狀況越好[7]的定義得知，圖 6 的排尿狀況較佳，但是圖 5 的峰值卻比較高。若是依據這個特徵去做分類，這兩張圖片可能會被分在同一個類別，這樣便會有這個特徵較不適合直接拿來分類的疑慮，因此我們選用類神經網路自動編碼器來提取特徵，期望能夠找到一些適合分類的特徵，同時又能夠避免上述的問題。

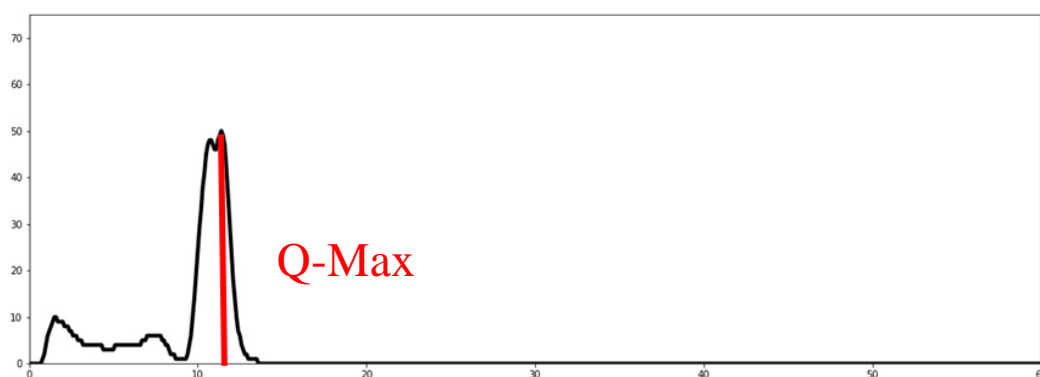


圖 5、尿流速圖 Q-Max 示意圖，排尿狀況差

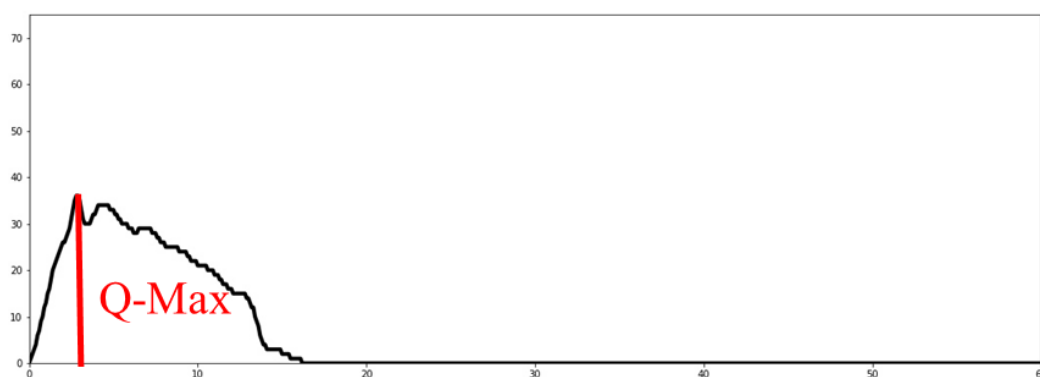


圖 6、尿流速圖 Q-Max 示意圖，排尿狀況佳

3.2 資料取得

尿流速與病患資料均是由 **MMS(Medical Measurement Systems)** 系統所提取出來，該系統為醫學資料庫，醫師做完病人的泌尿道各種檢查之後，可以在系統輸入，患者基本資料、檢查結果、檢測參數、報告和註記。系統介面如圖 7。

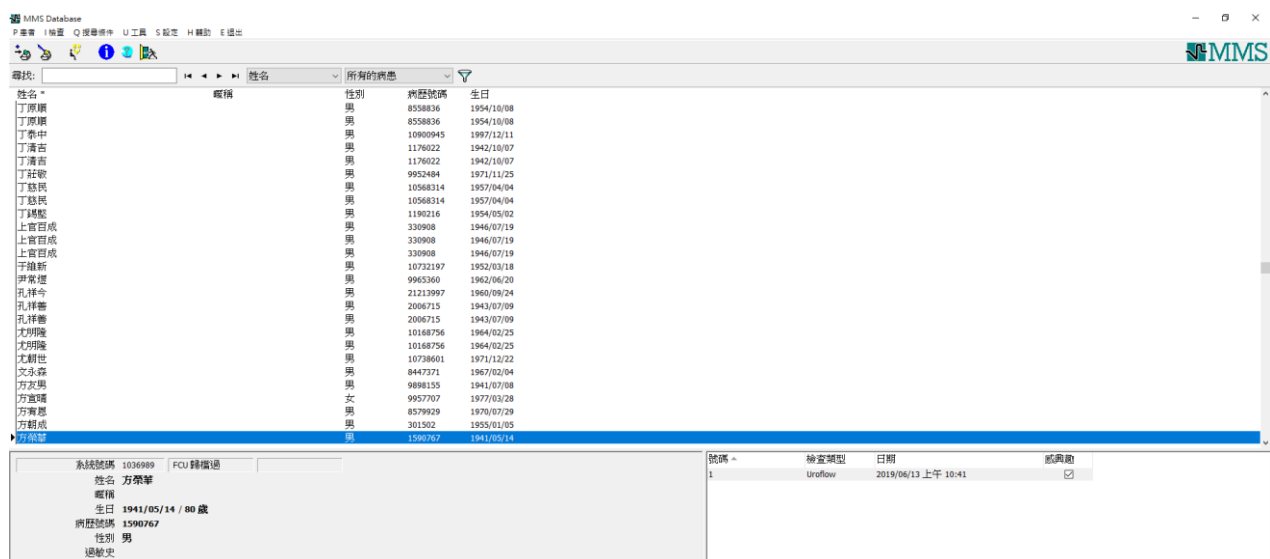


圖 7、MMS 操作介面

而我則從該系統提取出，我所需的患者基本資料以及尿流速圖的資料。

該系統內雖然可以藉由測量數值重建尿流速圖，但是會有比例失真的問題，進而導致醫師的判斷困難，因此提取出尿流速的數值後使用 Python 將其重建。

3.3 資料重建(資料預處理)

MMS 系統提取出來的檔案，資訊有 60 秒內 900 個值(每個值代表當下時間的尿流速)，再利用 Python 程式進行圖片重建，總共重建出 11892 張圖片，大小為 1500*550 像素的尿流速圖，如圖 8。

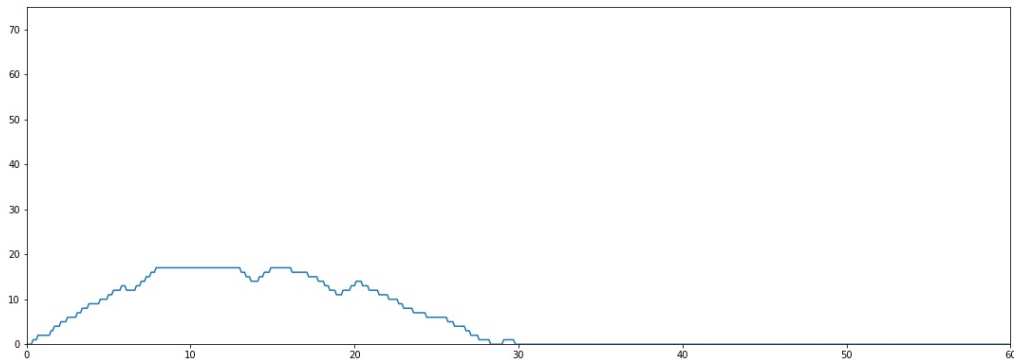


圖 8、尿流速圖重建結果



3.4 泌尿科醫師標注資料

完成重建後再將圖片交給泌尿科醫師進行評分，總共有 486 個病人，泌尿科醫師會去比較同一位病人每次回診測量的尿流速圖(第一次看診與第二次看診相比，第二次看診與第三次看診相比)觀察是否有進步並依據變化給出分數(1 分:表示退步；2 分:表示不變；3 分:表示進步)，泌尿科醫師總共提供 786 對 1186 張尿流速圖的進退步的標注。



3.5 圖片二值化處理

將泌尿科醫師判讀完資料後，將圖片進行二值化處理，更利於模型分類，將所有圖片讀入後遍歷每一張圖的每一個像素，將所有不是白色的像素改為黑色，完成二值化處理。圖 9、10 分別為預處理前與預處理後。

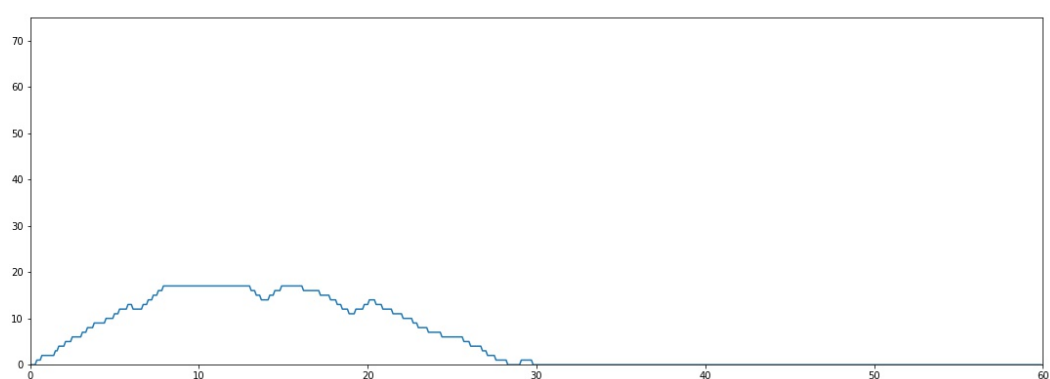


圖 9、原始圖片

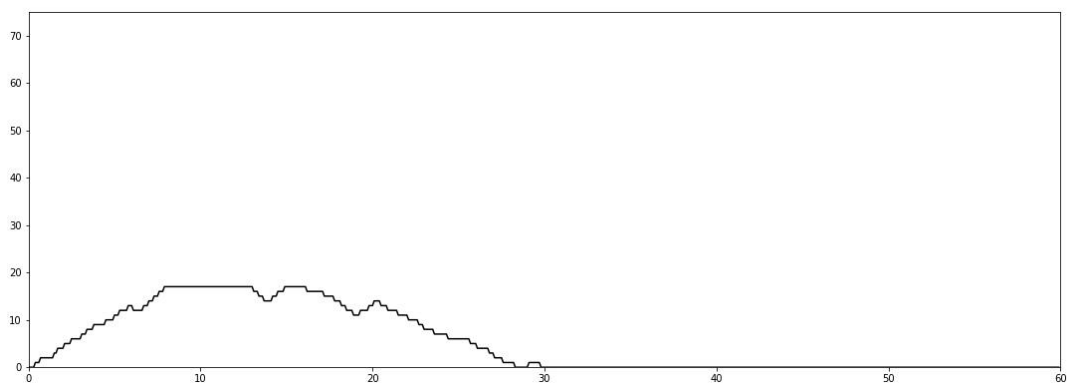


圖 10、經過二值化的預處理後圖片

3.6 位移尿流速圖

所有資料之中有些圖片有以下情形，如下圖在主要圖型前還有排出少量的尿，經過與泌尿科醫師討論後決定保留前面紅圈處的尿的體積大於總體積 2% 的圖形，因為大於總體積 2% 代表前面先排出的尿也是重要的特徵不能忽視，其餘的都做主圖型位移到貼近 y 軸的預處理。圖 11、12 分別為預處理前與預處理後。

會出現這樣的誤差是因為蒐集資料時，醫護按下開始測量的按鈕時

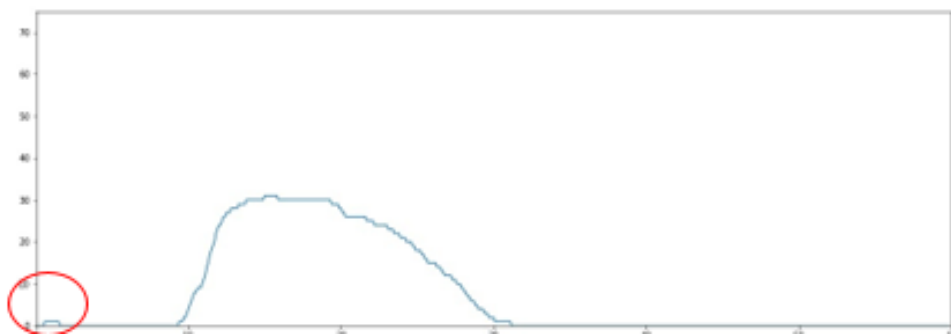


圖 11、原始圖片

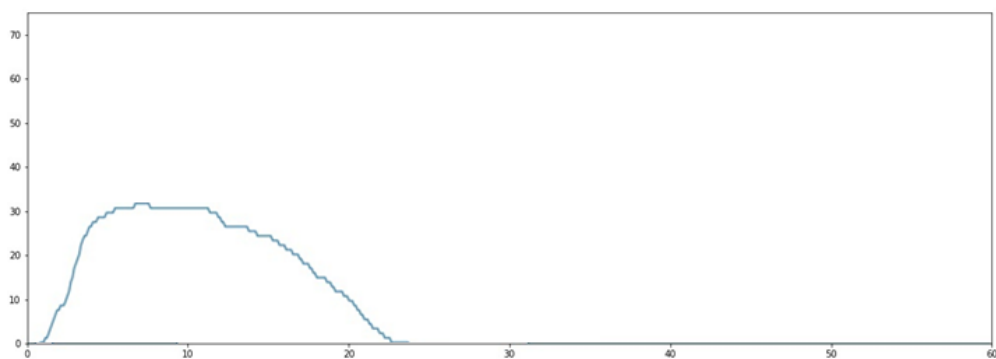


圖 12、經過位移圖片

病人還沒有準備好而造成了這樣的誤差，或是病人對於醫院設置的

排尿環境感到緊張，而造成這樣的排尿結果，對於這樣的資料，都進行了向 y 軸方向的位移直到貼近 y 軸，以排除測量誤差。



3.7 自動編碼器(特徵提取)

在建立自動編碼器參考[22]的 ECG 分類的網路結構，因為 ECG 的圖形與尿流速圖一樣都是曲線資料，所以參考該篇論文的神經網路模型，再依據尿流速圖的資料與實驗結果做參數上的調整，調整了層數與 filter 的數量，kernel size 的大小與提取出來的不同特徵維度的大小。



3.8 訓練流程介紹

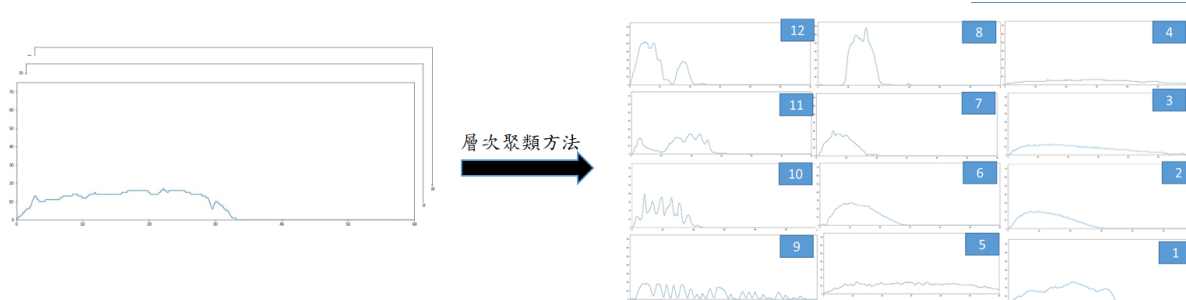
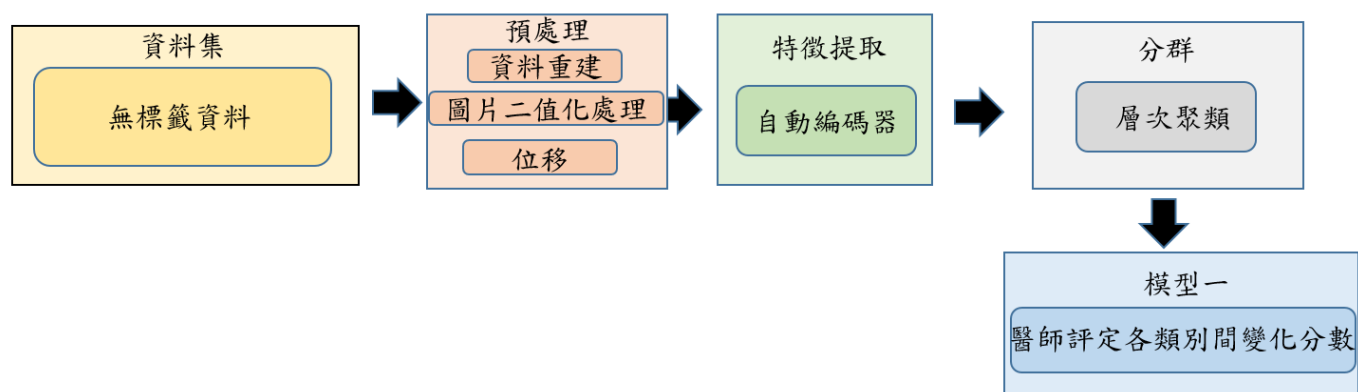


圖 13、模型一訓練流程圖

1→2: 退步	2→3: 退步	3→4: 退步	4→5: 進步	5→6: 退步	6→7: 退步	7→8: 退步	8→9: 退步	9→10: 進步	10→11: 退步	11→12: 進步
1→3: 退步	2→4: 退步	3→5: 進步	4→6: 進步	5→7: 退步	6→8: 退步	7→9: 退步	8→10: 進步	9→11: 進步	10→12: 進步	
1→4: 退步	2→5: 進步	3→6: 進步	4→7: 進步	5→8: 退步	6→9: 退步	7→10: 退步	8→11: 退步	9→12: 進步		
1→5: 進步	2→6: 進步	3→7: 進步	4→8: 進步	5→9: 退步	6→10: 退步	7→11: 退步	8→12: 進步			
1→6: 進步	2→7: 進步	3→8: 進步	4→9: 進步	5→10: 退步	6→11: 退步	7→12: 進步				
1→7: 進步	2→8: 進步	3→9: 退步	4→10: 進步	5→11: 退步	6→12: 進步					
1→8: 進步	2→9: 退步	3→10: 進步	4→11: 進步	5→12: 進步						
1→9: 退步	2→10: 進步	3→11: 進步	4→12: 進步							
1→10: 進步	2→11: 退步	3→12: 進步								
1→11: 退步	2→12: 進步									
1→12: 進步										

表 1、各類別間進退步對應表格

圖 4 為整個系統模型架構，而模型訓練中又分為模型一與模型二。如圖 13，在模型一中先將我們的訓練資料 11892 張無標籤資料，圖片經過預處理之後，使用自動編碼器提取特徵，再使用層次聚類分群，最後再去對應醫師評定各類別間變化分數。

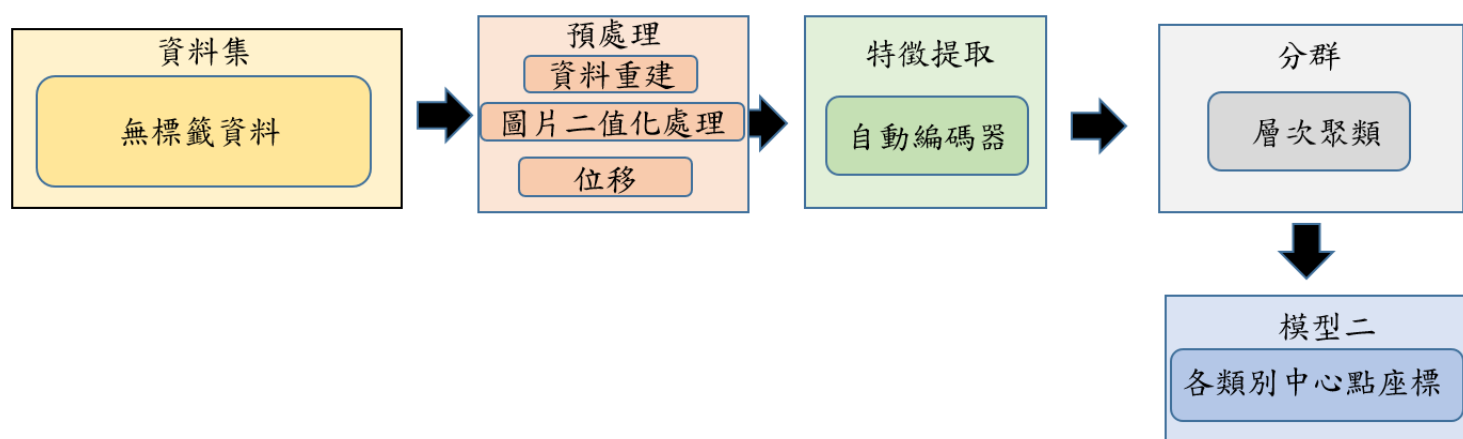


圖 14、模型二訓練流程圖

如圖 14，模型二將圖片經過預處理之後，使用自動編碼器提取特徵，再使用層次聚類分群，最後計算出各類別的中心點，以利模型測試圖 17 類別歸納計算時使用。對於有標籤資料，經過預處理之後使用自動編碼器提取特徵，便可以藉由模型二的各類別中心點座標，進行類別歸納，歸納到與其最相近的中心點就可以得知有標籤資料的類別如圖 15(a)。

再經過模型一的醫師評定個類別變化分數如圖 13。藉由上述步驟就

可以得知有標籤資料的進退步分數如圖 15(b)，進而計算準確率。

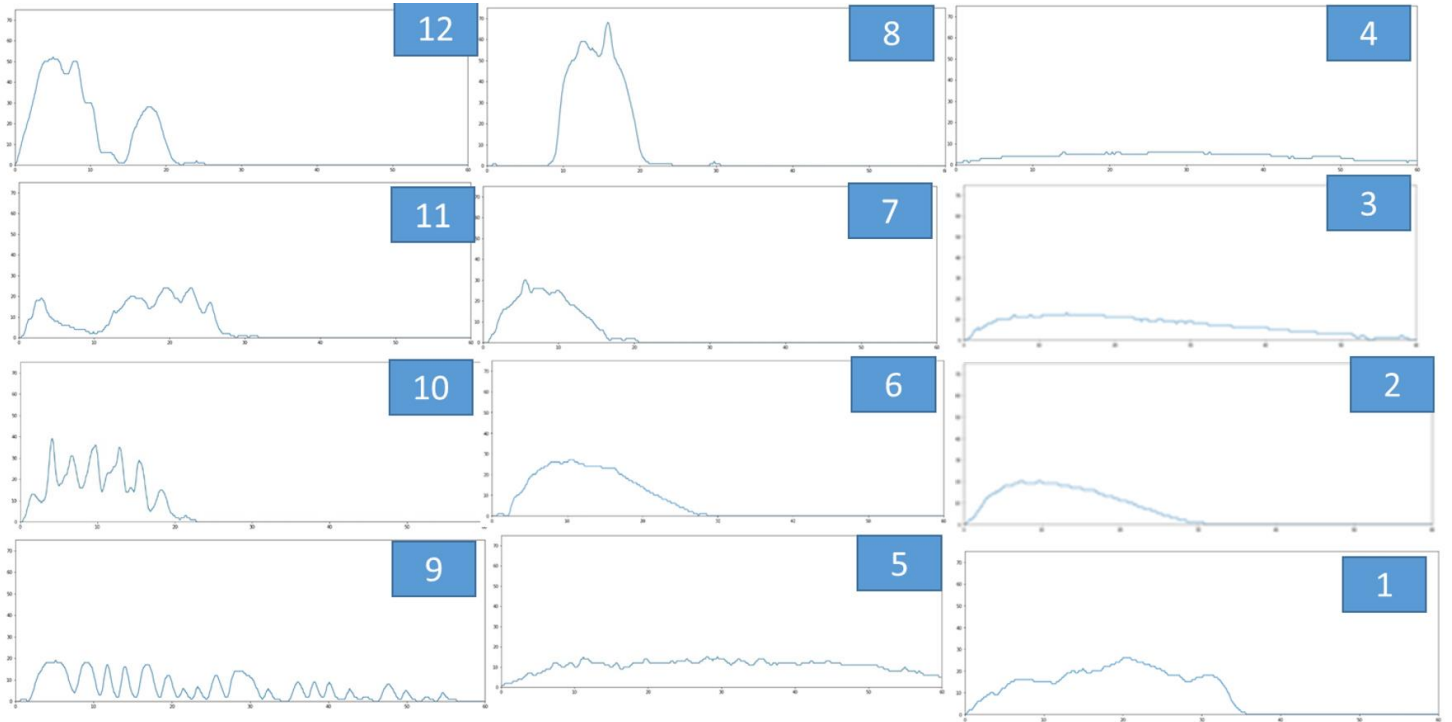


圖 15(a)、各類別示意圖

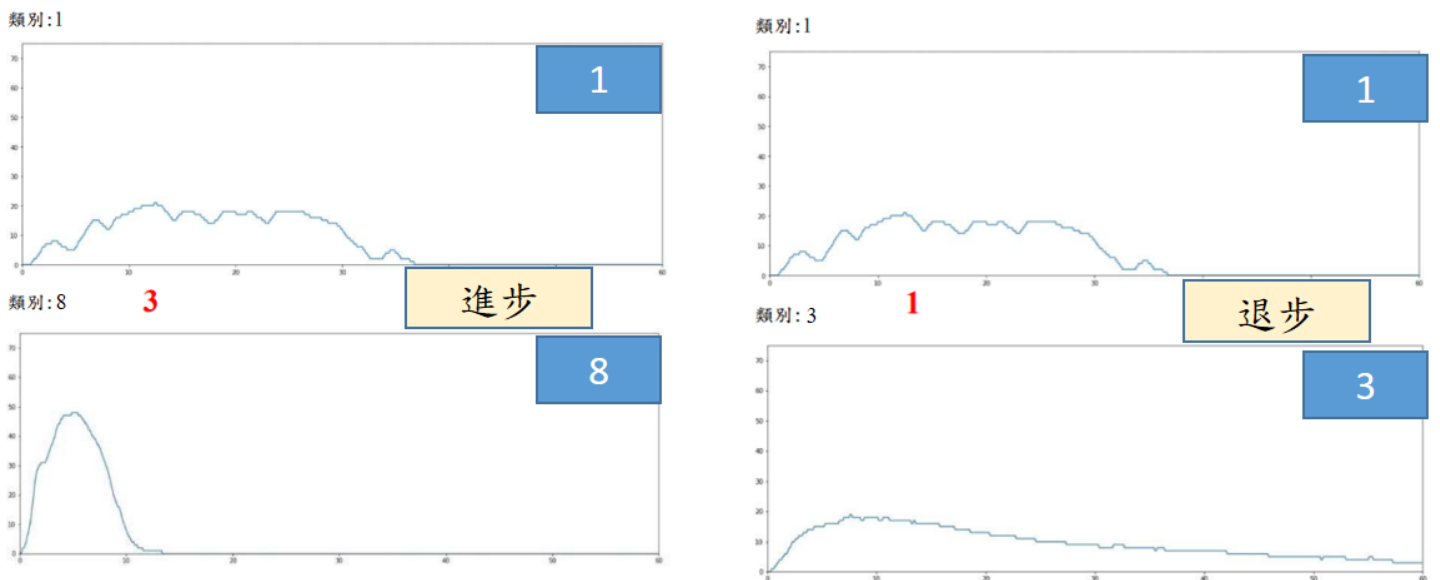


圖 15(b)、各類別進退步範例

對於無標籤資料，考慮到資料量龐大無法全部進行標註，我們決定用自動編碼器對每張圖提取適當的特徵維度 m ，經過實驗我們發現當特徵維度為 540 的情況下有最好的效能，再將這些特徵向 $X_i = (x_1^i, x_2^i, \dots, x_m^i)$ 以無監督式分群的方式，將相似的尿流速圖分在同一個類別，由於使用機器學習分群因此分群結果是客觀的，我們認為將客觀的分群結果給醫師評定分數最為合適。

在模型二中先將圖片經過預處理之後，使用自動編碼器提取特徵，再使用層次聚類分群，接著利用公式(1) 計算各群中心點座標 $Mean_j$ 。

$$Mean_j = \frac{\sum_{i=1}^{N_j} (x_{ij})}{N_j} \quad (1)$$

其中， N_j 為群 j 總個數， x_{ij} 為群 j 中第 i 個特徵向量。圖 16 為分群視覺化二維的示意圖，每個顏色分別代表不同群，而黑點則是各類別之中心點。

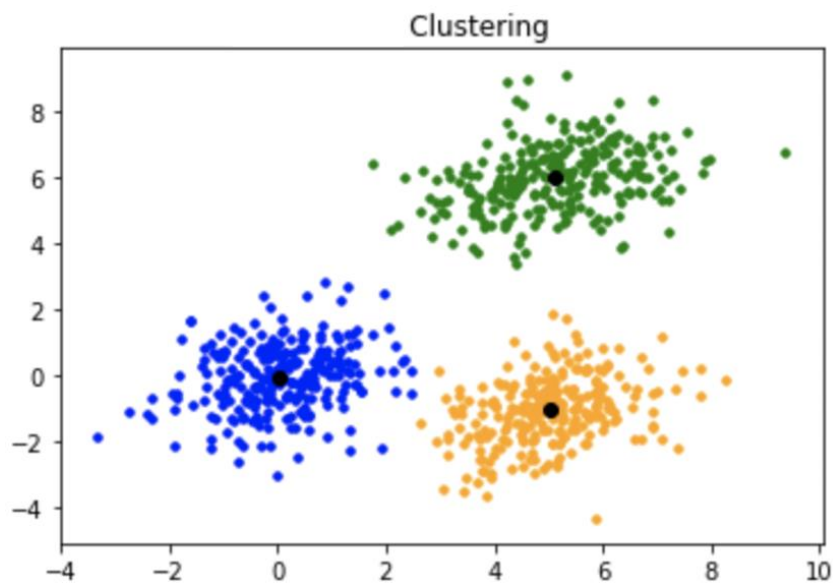


圖 16、二維特徵示意圖

依據分群中心點進行有評定資料的類別並進行歸納，以找到每筆資料所屬的類別，這樣當有新資料進入系統時，就不需要再重新訓練模型，直接依據中心點歸納到該類別所屬的分群，是十分節省效能與時間的做法。

3.9 模型測試

圖 17 為模型測試方法流程圖在模型測試階段輸入資料為連續回診兩次以上的病患尿流速圖 1186 張共 786 對。

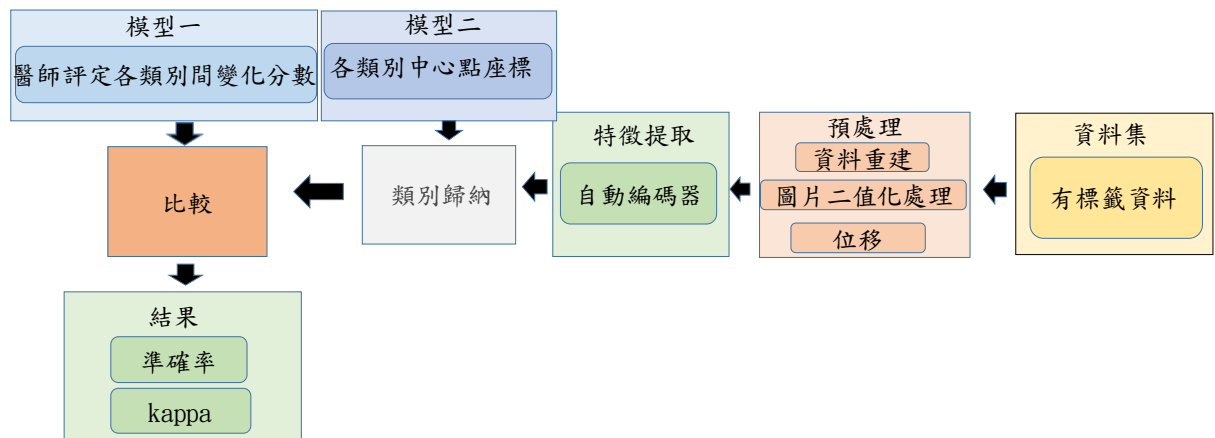


圖 17、模型測試方法流程圖

模型測試階段我們藉由 CNN Autoencoder 將測試資料的特徵提取出並依據特徵相似度對應到模型二:特定類別中心點座標，歸納到所屬的類別，再將結果對應到模型一:醫師評定各類別間變化分數，在比較的這個部分，測試資料經過圖 17 流程，系統給出的進退步結果將與有標籤資料的標籤(醫師測試資料評定)做比較，計算出準確率 (Accuracy)以及 Kappa 值。

第四章 系統評估與比較

第四章架構編排如下，我們將第四章分成兩個部分 4.1 與 4.2，其中 4.1 針對我們系統中的無監督式模型進行效能比較，而 4.2 針對整個系統進行系統效能的評估並針對圖片比例與特徵維度等變因調整後實驗結果的討論。



4.1 無監督式模型效能評估

許多無監督學習相關的論文會使用外部資料集 MNIST 等等的開放資料驗證無監督模型的準確率[23] [24]，本研究為了驗證層次聚類模型各種鏈接方式的分群效能，我們使用了容易取得的開放式資料集 MNIST 手寫辨識資料集以及與尿流速圖相似都是曲線資料的 ECG 心電圖資料集[22]來驗證我們的模型，如圖 18 為外部資料驗證層次聚類鏈接方式方法流程。

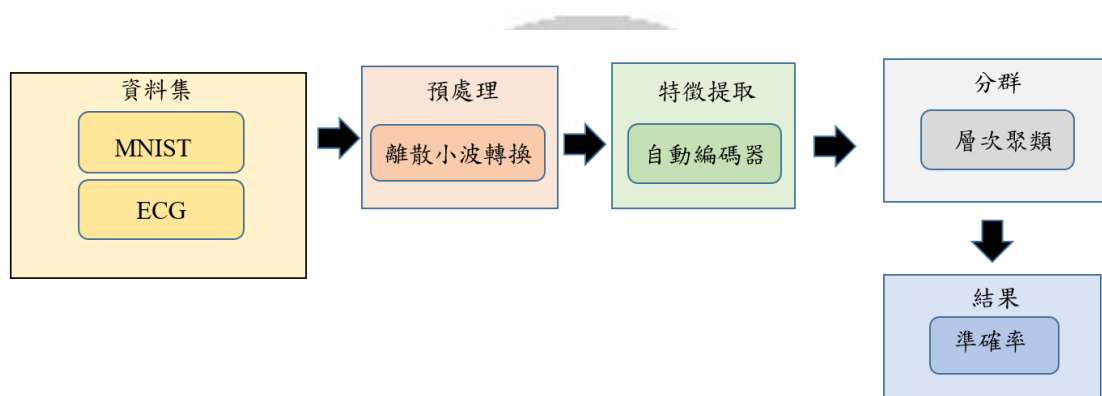


圖 18、外部資料驗證層次具類鏈接方式方法流程

4.1.1 外部資料介紹

圖 17 為外部資料方法流程，而 ECG 的部分我們採用的資料集是 MIT-BIH Arrhythmia Database（心律失常資料庫）它是由 Massachusetts Institute of Technology - 麻省理工學院和 Beth Israel Hospital – 美國麻薩諸塞州波士頓 Beth Israel 醫院一起聯合建立的。資料集中每條資料時長為 30 分鐘，提供了對應的 label 標籤檔案，標註的資訊可以從 label 標籤檔案中找到對應的資訊。

MNIST 是一個手寫數字的圖像資料集常用於機器學習訓練和測試的教學。資料集包含 60000 個訓練圖片和 10000 個測試圖片，每個圖片大小是 $28 * 28$ 像素。

4.1.2 外部資料預處理

ECG 的資料集有雜訊的比例蠻高的，因此應用 DWT (Discrete Wavelet Transform) 進行預處理。

4.1.3 統計相似度準確率計算

由於無監督式學習的分群結果的 label 是隨機的，例如 MNIST 的分類數字 1 的群，無監督式分群的標記不一定也會是有可能是 0 到 9 其中一個標記，因此無法像一般監督式學習一樣直接對應 label 去計算準確率，無監督式學習必需要運用比較統計相似度的方式計算準確率，並使用公式(4)進行準確率計算。

$$accuracy(y, \hat{y}) = \max_{perm \in P} \frac{1}{n} \sum_{i=0}^{n-1} 1(perm(\hat{y}_i) = y_i) \quad (2)$$

4.1.4 kappa 值與準確率計算

在醫學的領域中，當針對類別型資料(categorical data)，實務上最常使用的是 **kappa** 用來計算預測標註與實際標註的一致性。Kappa 係數計算上簡單且不會大量耗費資源，也容易理解如表 2 為 Kappa 值算法的簡例。最早是由 Galton (1892) 提出，計算方式如下：

$$kappa = \frac{p_0 - p_e}{1 - p_e} \quad (3), \text{ 其計算結果範圍為 } [-1 \sim 1]。$$

假設每一類的真實樣本個數分別為 a_1, a_2, \dots, a_n ，而預測出來的每一類的樣本個數分別為 b_1, b_2, \dots, b_n ，以圖 17 為例。

$$P_0 = \frac{a_{11} + b_{22} + c_{33}}{n} \quad (4) \text{ 每一類正確分類的樣本數量之和除以總樣本數}$$

$$P_e = (a_1 \times b_1 + a_2 \times b_2 + \dots + a_n \times b_n) / (n \times n) \quad (5) \text{ 為期望一致率}$$

	變好	不變	變差
變好	10	5	5
不變	2	35	2
變差	8	5	15

a1

a2

a3

b1

b2

b3

$$P_0 = (10 + 35 + 15) / 87 = 0.689$$

$$a_1 = 10 + 2 + 8 = 20$$

$$a_2 = 5 + 35 + 5 = 45$$

$$a_3 = 5 + 2 + 15 = 22$$

$$b_1 = 10 + 5 + 5 = 20$$

$$b_2 = 2 + 35 + 2 = 39$$

$$b_3 = 8 + 5 + 15 = 28$$

$$P_e = (a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3) / (87 \times 87) = 0.455$$

$$Kappa = (P_0 - P_e) / (1 - P_e) = 0.4293578$$

表 2、kappa 算法範例

由於參考過的醫學相關論文[21][22]都有使用 kappa 值代表研究結果與醫師判斷結果的一致性，因此我們也決定加入 kappa 值做為實驗結果的參考。

4.1.5 準確率 (Accuracy)

準確率可以定義為正確分類實例的百分比 $(tp + tn) / (tp + tn + fp + fn)$ 。其中 tp 、 fn 、 fp 和 tn 分別表示真陽性、假陰性、假陽性和真陰性的數量。準確率計算方式如下

$$accuracy = \frac{tp+tn}{tp+fp+tn+fn} \quad (6)$$

4.1.6 層次聚類鏈接方式效能評估

為了尋找層次聚類最適合的鏈結方式，我們對於每種常見的鏈結方式[12]做了實驗，層次聚類有 4 種常用的鏈接方式分別是

AC-Single Linkage 計算方式如下

$$d(C_i, C_j) = \min_{a \in C_i, b \in C_j} d(a, b) \quad (7)$$

AC-Complete Linkage 計算方式如下

$$d(C_i, C_j) = \max_{a \in C_i, b \in C_j} d(a, b) \quad (8)$$

AC-Average linkage 計算方式如下

$$d(C_i, C_j) = \sum_{a \in C_i, b \in C_j} \frac{d(a, b)}{|C_i||C_j|} \quad (9)$$

其中， a 是 C_i 中的元素， b 是 C_j 中的原素。

AC-Ward linkage 計算方式如下

$$d(C_i, C_j) = \sum_{a \in C_i \cup C_j} ||a - \mu_{C_i \cup C_j}||^2 \quad (10)$$

其中， a 是 $C_i \cup C_j$ 中的元素。

下表為 MNIST 與 ECG 資料集在層次聚類模型上的結果。以下表

3、4 為 ECG 與 MNIST 的不同鏈結方法的實驗結果。

表 3、比較 MNIST 資料集在不同鏈接方式下的準確率

實驗方法(MNIST)	準確率(%)
AC-ward-linkage	90.6
AC-average-linkage	84.0
AC-complete-linkage	86.7
AC-single-linkage	80.9

表 4、比較 ECG 資料集在不同鏈接方式下的準確率

實驗方法(ECG)	準確率(%)
AC-ward-linkage	81.6
AC-average-linkage	75.4
AC-complete-linkage	77.1
AC-single-linkage	65.7

4.1.7 DBSCAN 與層次聚類效能比較

在第二章有討論到不需要事先知道 K 的演算法有層次聚類與 DBSCAN 兩種，這邊去比較這兩種模型在我的資料集上表現的效能。

表 5、尿流速圖層次聚類與 DBSCAN 的準確率比較

模型	準確率(%)
層次聚類	90.1
DBSCAN	66.8

由表 5 可以觀察出 DBSCAN 表現結果較差，原因為 DBSCAN 是基於密度的分群方式，若是資料密度差異過大，會影響分群結果，而我們的資料密度差異蠻大的，在顏色較淺的地方，資料密度分散；而在顏色較深的地方，資料密度愈集中。圖 19 為二維資料密度示意圖，由於我們的資料維度高，無視覺化上面有難度，因此用二維的示意圖做說明。

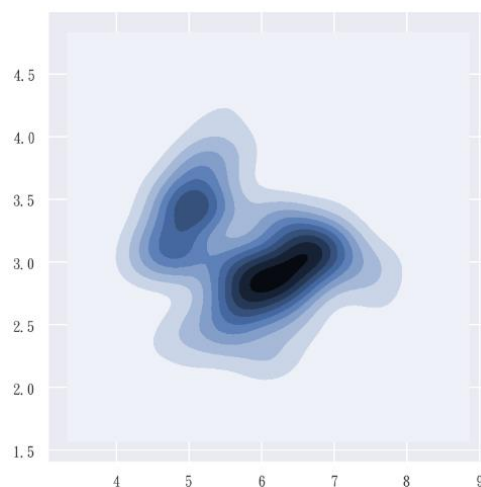


圖 19、二維資料密度分布示意圖

4.2 系統評估

在本節中將說明對於整個系統的準確率評估，依照過往實驗經驗，在眾多可調整變因中例如 kernel size、filter、圖片比例、圖片解析度、特徵維度...等等。我們發現，圖片比例、圖片解析度以及特徵維度對於準確率影響較大，因此以下 4.2.1 與 4.2.2 將分別針對圖片比例與解析度和特徵維度去做比較。



4.2.1 圖片比例與解析度對於自動編碼器特徵提取的影響

在使用自動編碼器提取特徵時，一開始因為寫程式方便而改變圖片比例，將本來是 1500*500 像素的圖片壓縮為 100*100 像素，因為這個步驟中導致圖片的比例失真，也連帶的影響到自動編碼器提取出來的特徵，導致準確率很低，經過多次討論決定將圖片比例保留，準確率就有大幅度的提升，下表 6 為是否改變圖片比例的比較表。由表 6 可知在解析度較低的情況下，圖片比例維持的情況下準確率還是較高的。

表 6、是否改變照圖片比例

	Input shape	Accuracy(%)
圖片比例維持	60*22	80.3
改變圖片比例	100*100	68.9

為了驗證我們調整圖片比例之後為何會嚴重的影響準確率，其最重要的影響變因為何，我們將保留圖片比例的各個不同的解析度與不改變圖片比例的各種解析度進行比較，下圖 20 為比較圖。

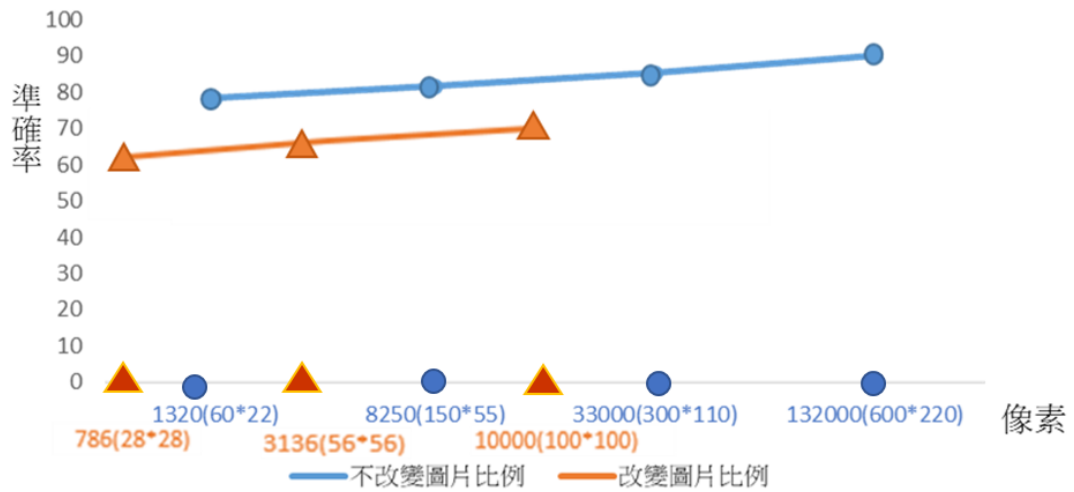


圖 20、圖片比例改變與否與各種不同的解析度的比較

由上圖 20 可得知，主要影響解析度的變因為圖片比例是否改變，而解析度雖然也有影響到準確率，但並非主要的變因。

4.2.2 最適特徵維度

本系統是藉由自動編碼器提取特徵，因此提取出的特徵相當重要，經由實驗與[23][24]，我們可以知道特徵維度對於準確率有一定程度的影響，因此我們應用二分逼近法找出對於資料最適合的特徵維度。

表 7、比較本研究在不同特徵維度的準確率

Code size	Accuracy(%)	Kappa
64	80.2	0.677
128	83.5	0.730
256	86.7	0.785
320	87.3	0.794
430	88.9	0.819
485	89.8	0.835
513	90.04	0.838
540	90.1	0.840
580	89.7	0.834
630	88.9	0.817
737	87.5	0.802
844	85.8	0.76.9

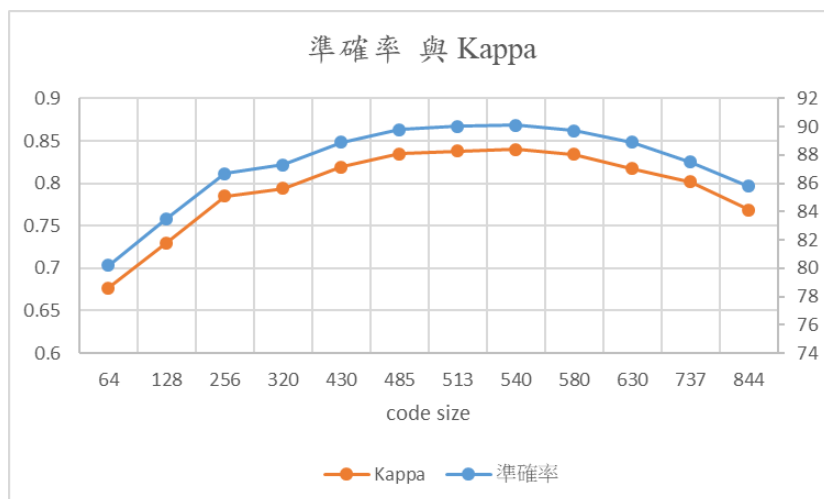


圖 21、準確率與 Kappa 在各個特徵維度的實驗結果

我們比較了不同數量的自動編碼器提取的特徵維度[25][26]。可以從表 6、圖 21 的實驗結果觀察出 540 的維度是最適合的特徵大小維度太少或太大都不太適合。

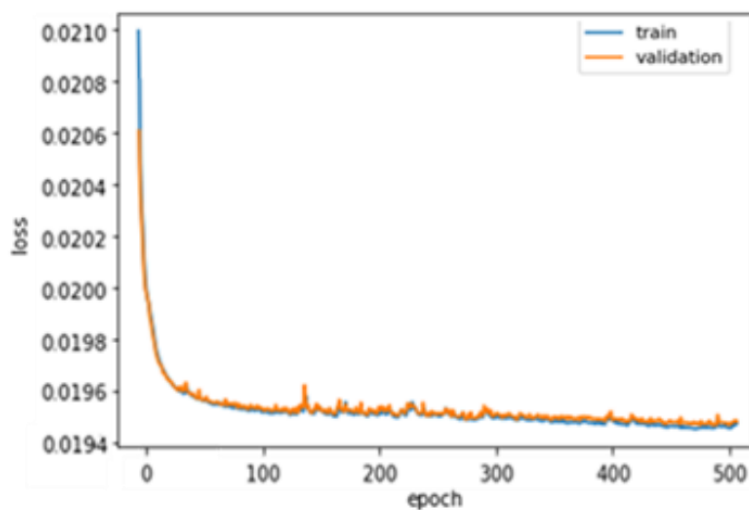


圖 22、540 特徵維度的收斂狀況

32 維過度壓縮特徵導致資訊遺失過多，影響分類結果。而過大的特徵因為過擬和的問題而導致結果不佳。

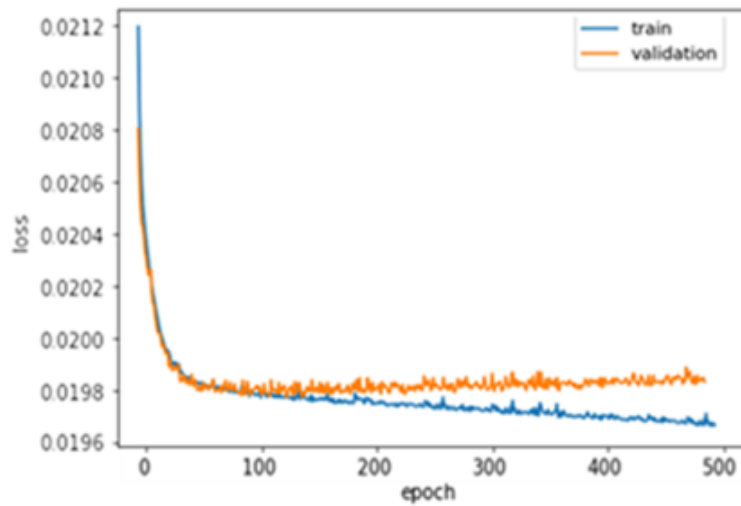


圖 23、844 特徵維度的收斂狀況

圖 22 為特徵維度 540 之收斂狀況，圖 23 為特徵維度 844 之收斂狀況，可以看出並沒有過擬合的情形，而維度 844 的收斂情況可以看出雖然不嚴重，但還是有過擬合的情況。經由實驗證明過多的特徵會產生過擬和的情況[27]。

第五章 結論

在我們的解決方案中，尿流速圖是一種很常見的尿流速測量結果，常用來評估排尿功能障礙。但我們尿流速圖資料龐大醫師很難解釋和分類，因此我們的解決方法為先用自動編碼器提取出每張圖的特徵再用層次聚類將相似的尿流速圖分在一類，此方法不僅大量的減少醫師的負擔，也讓結果更加的客觀，再請醫師去判斷每個尿流速圖的類別與對應令每一類別的進退步。通過與泌尿科醫師評分的比較，我們的結果達到了 90% 的準確率。

未來工作與改良方向，由於目前的資料來源皆為病人在醫院進行測量所得，但是病患並不是在熟悉的環境下排尿，這可能導致測量結果與病患日常生活的排尿狀況有一些差異，若是可以取得病人在家放鬆的情況下的排尿數據，分析的結果將會更貼近病人生活，更能夠協助醫師診斷。我們相信之後若是可以取得 BOO (Bladder Outlet Obstruction) 的資料讓機器學習判斷病人是否有 BOO 的情形將會是一個很棒的醫療輔助，但是經過與醫師的討論後是否 BOO 不是單獨依靠尿流速圖就可以判斷的，還要再加上殘餘尿量...等等，但是目前資料取得不易，若是未來資料蒐集的難度降低了，將能夠對醫師提供更多的醫療輔助。

參考文獻

- [1] R.Tewari et al., “Diet, Obesity, and Prostate Health: Are We Missing the Link?, ” J. Androl., vol. 33, no. 5, pp. 763–776, Sep.2012.
- [2] J.JB, J.KM, B.-B.NE, andM.P, “Uroflowmetry in asymptomatic elderly males, ” Br. J. Urol., vol. 58, no. 4, pp. 390–395, 1986.
- [3] S.W et al., “Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies, ” Neurourol. Urodyn., vol. 21, no. 3, pp. 261–274, 2002.
- [4] V.FA, Z.PE, B.GR, andN.PP, “Modelized analysis of pressure-flow studies of patients with lower urinary tract symptoms due to benign prostatic enlargement, ” Neurourol. Urodyn., vol. 22, no. 1, pp. 45–53, 2003.
- [5] Valentini, Françoise A., "A mathematical micturition model to restore simple flow recordings in healthy and symptomatic individuals and enhance uroflow interpretation." ,Neurourology and Urodynamics: Official Journal of the International Continence Society vol.19,no.2, pp.153-176,2000.
- [6] F.I, F.J, L.YS, C.EK, and H.SW, “Can a quantitative means be used to predict flow patterns: Agreement between visual inspection vs. flow index derived flow patterns, ” J. Pediatr. Urol., vol. 12, no. 4, pp. 218.e1-218.e8, Aug.2016.
- [7] F.I, S.-D. Y.S, C.SJ, N.B, and F.JA, “A quantitative approach to the interpretation of uroflowmetry in children, ” Neurourol. Urodyn., vol. 35, no. 7, pp. 836–846, Sep.2016.
- [8] J. M. B.Netto, A.Hittelman, S.Lambert, K.Murphy, T.Collette-Gardere, and I.Franco, “Interpretation of uroflow curves: A global survey measuring inter and intra rater reliability,” Neurourol. Urodyn., vol. 39, no. 2, pp. 826–832, Feb.2020.

- [9] S. P. R.Baas, "A machine learning approach to the automatic classification of female uroflowmetry measurements, " 2016.
- [10] I. S.Dhillon, Y.Guan, and B.Kulis, "Kernel k-means, Spectral Clustering and Normalized Cuts, " 2004.
- [11] NAZI, Azade, "Generalized Clustering by Learning to Optimize Expected Normalized Cuts.", arXiv preprint arXiv:1910.07623, 2019.
- [12] U.vonLuxburg, "A tutorial on spectral clustering, " Stat. Comput. 2007 174, vol. 17, no. 4, pp. 395–416, Aug.2007.
- [13] X.He, D.Cai, Y.Shao, H.Bao, and J.Han, "Laplacian regularized Gaussian mixture model for data clustering, " IEEE Trans. Knowl. Data Eng., vol. 23, no. 9, pp. 1406–1418, 2011.
- [14] V.Vijaya, S.Sharma, and N.Batra, "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering, " Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 568–573, Feb.2019.
- [15] K.Khan, S. U.Rehman, K.Aziz, S.Fong, S.Sarasvady, and A.Vishwa, "DBSCAN: Past, present and future, " 5th Int. Conf. Appl. Digit. Inf. Web Technol. ICADIWT 2014, pp. 232–238, 2014.
- [16] S. M.Mousavi, W.Zhu, W.Ellsworth, and G.Beroza, "Unsupervised Clustering of Seismic Signals Using Deep Convolutional Autoencoders, " IEEE Geosci. Remote Sens. Lett., vol. 16, no. 11, pp. 1693–1697, Nov.2019.
- [17] C.D, "Siamese Neural Networks: An Overview, " Methods Mol. Biol., vol. 2190, pp. 73–94, 2021.

- [18] L.Song, D.Gong, Z.Li, C.Liu, and W.Liu, "Occlusion Robust Face Recognition Based on Mask Learning With Pairwise Differential Siamese Network." pp. 773–782, 2019.
- [19] A. M.Martinez and A. C.Kak, "PCA versus LDA, " IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 228–233,2001.
- [20] W. S.Noble, "What is a support vector machine?, " Nat. Biotechnol. 2006 2412, vol. 24, no. 12, pp. 1565–1567, Dec.2006.
- [21] M.Pal, "Random forest classifier for remote sensing classification, ", vol. 26, no. 1, pp. 217–222, Jan.2007.
- [22] M.Kachuee, S.Fazeli, and M.Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation, " Proc. - 2018 IEEE Int. Conf. Healthc. Informatics, ICHI 2018, pp. 443–444, Apr.2018.
- [23] J.Yang, D.Parikh, D.Batra, and V.Tech, "Joint Unsupervised Learning of Deep Representations and Image Clusters, " Accessed: Oct.04, 2021.
- [24] P.Ferré, F.Mamalet, and S. J.Thorpe, "Unsupervised Feature Learning With Winner-Takes-All Based STDP, " Front. Comput. Neurosci., vol. 0, p. 24, Apr.2018.
- [25] Y.Yang, Q. M. J.Wu, and Y.Wang, "Autoencoder with Invertible Functions for Dimension Reduction and Image Reconstruction, " IEEE Trans. Syst. Man, Cybern. Syst., vol. 48, no. 7, pp. 1065–1079, Jul.2018.
- [26] N.Renström, P.Bangalore, and E.Highcock, "System-wide anomaly detection in wind turbines using deep autoencoders, " Renew. Energy, vol. 157, pp. 647–659, Sep.2020.
- [27] M. H.Mohd Noor, M. A.Ahmadon, and M. K.Osman, "Activity Recognition using Deep Denoising Autoencoder, " Proc. - 9th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2019, pp. 188–192,

Nov.2019.

