

國立臺灣師範大學機電科技學系
碩士論文

指導教授：吳順德博士

基於 K-means 演算法、小波轉換及支持向量機之心電
訊號辨識系統

An Arrhythmia Recognition System Based on K-means Clustering、Wavelet
Transform and Support Vector Machine.



研究生：張家熏 撰

中華民國 100 年 6 月

致謝

在這二年的碩士班生活中，首先感謝指導教授吳順德博士，在這短短二年間給予我課業上的指導及生活上的關懷與勉勵，回想起這時間的點滴，猶如昨日的情景，有甘甜的回憶，也有萬般不捨，將邁向人生另一階段的旅程。

尤其是老師在精神上的想法讓我們學生獲得不止在課業上，而是以後人生觀價值的影響，堅持著對於品德操守的要求，這些都讓我獲益良多。也感謝實驗室的所有成員：學長士宜、家齊、正倫、岳穎與本屆同學福豪、良榮及助教弘傳和學弟求文、庭毅、易宗、冠宇、威諭、偉恩等多人提供許多協助，特別感謝岳穎學長及求文的幫忙，讓本人難以忘懷和大家一起努力奮鬥的日子及給予種種建議與指導，感謝所以幫助過我的和關心我的人，謝謝大家我會永遠記得並且感恩。

最後要給感謝我的家人，因為有我父母親用心栽培與姐姐和妹妹的鼓勵，能讓我一直努力下去，特別是在我生病時他們是我的精神支柱，這樣的力量讓我能利完成研究，並且順利完成碩士學業，希望藉由本論文的完成，與大家分享我的榮耀與喜悅，並願大家永遠身體健康和一顆開朗快樂的心。

中文摘要

本論文利用小波轉換(Wavelet transform)、K-means 分群法(K-means clustering)及支持向量機(Support vector machine)等方法，建立一個辨識各種心律不整的心電辨識系統。本論文所提的方法可以大致區分為三個階段；第一階段使用 K-means 分群法把屬於同一類別但相異性卻很大的心律不整訊號分成數個次類別，在每一個次類別，各樣本會有較高的相似性。第二階段則把各次類別裡的每一個心搏樣本利用小波轉換擷取時頻特徵向量。第三階段以每一個心搏樣本的時頻特徵以及形態特徵為訓練資料，並運用支持向量機來建立本辨識系統的模型。為了驗證本系統的有效性以及可靠性，本論文利用 MIT-BIH 心律不整資料庫進行了三個實驗。實驗的結果本論文所提的方法具有相當高的辨識率達 98.2%，最後與各相關辨識系統文獻比較差異。

關鍵字：心電訊號辨識系統、k-means 演算法、小波轉換、支持向量機。

Abstract

This paper described an arrhythmia classification system based on the technologies of wavelet transform, k means clustering and support vector machine for the purpose of heartbeat recognition. The method consists of three stages. At the first stage, the waveform of a single heartbeat in each main group is classified into subgroups using k-means clustering technology. At the second stage, the time-frequency features of each heartbeat were extracted by using wavelet transform. At the third stage, the model of the proposed classification system is obtained by using support vector machine (SVM). The training vector of SVM is the combinations of morphological features and time-frequency features extracted using wavelet transform. Three experiments were done to examine the performance and reliability of the proposed classification system. Experiments show that the efficiency and feasibility of this proposed classification system.

Key words: Arrhythmia classification system, k-mean clustering, Wavelet Transform, support vector machine.

目錄

第一章 緒論.....	8
1.1 前言	8
1.2 研究動機與目的	8
1.3 論文架構	10
第二章 心電圖概論及相關研究探討	11
2.1 心電圖介紹	11
2.2 心電圖十二導程	12
2.3 心電圖 PQRST 型態及判讀	16
2.4 MIT-BIH 心律不整資料庫	18
2.5 心電訊號相關的干擾	26
第三章 系統架構、心電圖辨識演算法及相關分析理論	29
3.1 心電圖辨識系統流程簡介	29
3.2 零相位延遲高通濾波器	32
3.3 Segmentation(訊號切割)	34
3.4 K-means 演算法分類器	37
3.5 小波轉換演算法	44
3.6 心電圖訊號之特徵值擷取	46
3.7 支持向量機簡介	50
3.8 最佳化特徵擷取	53
3.8 心電圖病症辨識	56
第四章 實驗結果	58
4.1 心電圖訊號辨識結果	58
4.2 本論文與相關文獻比較	61
參考文獻	63

圖 目錄

圖 2-1 細胞去極化與再極化過程.....	12
圖 2-2 心臟電傳導細胞.....	13
圖 2-3 十二導程各貼片位置.....	13
圖 2-4 誘導 I、誘導 II、誘導 III.....	14
圖 2-5 愛因托分三角形.....	14
圖 2-6 AVR、AVL、AVF.....	15
圖 2-7 胸前誘導.....	15
圖 2-8 心電圖區間、段、間期.....	16
圖 2-9 各種類型 QRS 綜合波.....	17
圖 2-10 標準 PQRS 心電波形圖	18
圖 2-11 MIT-BIH 心律不整資料庫 V2 導程心電圖.....	19
圖 2-12 MIT-BIH 心律不整資料庫 MLII 導程心電圖.....	19
圖 2-13 正常心搏波形.....	20
圖 2-14 左分支束阻塞波形.....	21
圖 2-15 右分支束阻塞波形.....	21
圖 2-16 變形的心房早期收縮波形.....	22
圖 2-17 心室性早期收縮波形.....	22
圖 2-18 心室融合心跳波形.....	23
圖 2-19 結節性過早收縮波形.....	23
圖 2-20 心房性過早收縮波形.....	24
圖 2-21 心室性脫逸收縮波形.....	24
圖 2-22 結節性脫逸收縮.....	25
圖 2-23 心律調節器心跳.....	25
圖 2-24 節律器所融合心跳.....	26
圖 2-25 心電訊號基準線飄移.....	27
圖 2-26 心電訊號肌肉波.....	27
圖 2-27 心電訊號市電頻率干擾.....	28
圖 3-1 心電圖辨識系統流程圖.....	30
圖 3-2 零相位延遲濾波器流程圖.....	33
圖 3-3 QRS 波偵測流程圖	34
圖 3-4 原始心電圖訊號.....	35
圖 3-5 低通濾波器處理後的心電圖訊號.....	35
圖 3-6 同一基準點心電圖訊號.....	35
圖 3-7 心電圖波形 192 個樣本數示意圖.....	36
圖 3-8 K-means 分群過程示意圖	39

圖 3- 9 適應性 K-means 演算法流程圖	40
圖 3- 10 適應性 K-means 演算法示意圖	41
圖 3- 11 心電圖訊號分類示意圖	43
圖 3- 12 小波分解示意圖	45
圖 3- 13 心電訊號小波分解示意圖	48
圖 3- 14 統計特徵及時域特徵類型	49
圖 3- 15 支持向量機二元分類圖	51
圖 3- 16 兩兩之間特徵向量區別值示意圖	54
圖 3- 17 最佳化特徵向量排序示意圖	55
圖 3- 18 心電圖病症辨識流程圖	56

—

表 目錄

表 2-1 心電圖病症代碼表	20
表 4-1 第一種實驗結果	59
表 4-2 第二種實驗結果	60
表 4-3 第三種實驗結果	60
表 4-4 本論文與相關文獻比較表	61

第一章 緒論

1.1 前言

隨著時代科技的進步，現代的人們壓力也隨之增加，國人主要疾病由心臟病取代了以往的傳染病和慢性病。依據行政院衛生署 2009 年的統計資料，國人因心臟病死亡人數為 15093 人、死亡率高達 65.4%，排名國人十大死因疾病的第二位[1]，而且死亡人數逐年遞增顯示心臟病嚴重性非常明顯。心臟病也是慢性病的其中一種，慢性病則需要長期治療和照顧，除了對病人本身是心理和生理上的壓力外，對於身旁的家人也是一個沉重的負擔，可見心臟病的嚴重性不容小覷。

心臟病患，發病後都往往都會造成行動上的不便，並且一再地復發，所以在事前的預防和監測是刻不容緩的，目前醫療的模式已由傳統的床邊照護慢慢演變為居家遠端監測，因此完整的心電圖辨識系統發展是非常的重要。

在量測心電訊號時往往會有干擾雜訊出現，若心電訊號中出現雜訊干擾，將使得心電圖辨識系統的辨識效果變差且會造成醫療人員誤判的可能，嚴重的會影響到患者所需的正確診療，無形中將浪費了醫療資源且延遲治療。有鑑於此，發展出優良的心電圖辨識系統是刻不容緩。

1.2 研究動機與目的

由於心電圖是目前診斷心臟疾病最基本也是最為方便的量測工具。心電圖信號能提供多方面的心臟功能訊息，且能判斷心臟相關疾病。這些由病患身上所量測到的訊號內容能詳細記錄病患生理狀況及病情，使得專業醫療工作人員能夠即

時有效診斷病患且給予正確的治療病症。不過這些由病患身上所量測得到的心電訊號是一堆龐大的數據，必須經過一系列有效的處理，才能得到對專業醫療人員有所幫助的數據。本論文開發一套有效率且高可靠性的心電圖辨識系統，透過電腦分析判斷心電圖訊號並且篩選正確的數據，提供給第一線的專業醫療人員重要的訊息協助，做出快速且正確的醫療診斷。

心電訊號是心臟在收縮時所產生的電壓活動紀錄。這些電位差變化與心臟健康狀況有著很緊密的關連。藉由異常的心電訊號，可推測出心臟疾病發生的原因，其診斷的方法為非侵入式技術。量測心電訊號時可能因量測的方法及過程而摻入許多不同且無法避免的雜訊，這些雜訊的摻雜會影響系統的判斷甚至大大降低辨識率，所以我們要去評估與改善心電訊號的雜訊對系統的影響。

一般心電訊號通常都會包含許多不同的雜訊，最常見的是：60Hz 市電干擾、0.3Hz 基準線飄移及 5~100Hz 肌肉波。本研究將評估雜訊干擾對心電圖辨識系統的影響。量測心電訊號過程中或多或少都會摻雜干擾雜訊，其對心電圖辨識系統影響程度為何，是否對於心電圖辨識系統做辨識時會降低心臟疾病辨識率，以及從患者身上量測到的心電訊號經過濾波處理後，能否提高本心電圖辨識系統的辨識率等，都是本論文所欲探討的課題。

1.3 論文架構

本論文開發一心電圖辨識系統，並探討雜訊對系統的影響，全文可分成五個章節：

第一章，簡介本論文研究目的及動機與問題描述。

第二章，敘述心電圖概論和相關研究探討。

第三章，描述本論文整個心電圖辨識系統架構以及介紹相關分析理論與心電圖辨識系統的各種演算法及其原理。

第四章，實驗的結果的評析。

第二章 心電圖概論及相關研究探討

在做心電圖訊號分析前，必須先了解什麼是心電訊號。本章節簡略敘述基本心電圖的概念，包含心電圖簡介、十二導程心電圖、心電圖 PQRST 型態及判讀、MIT-BIH 心律不整資料庫、波形檢測與病徵探討、心電訊號相關的干擾...等。

2.1 心電圖介紹

心電圖(Electrocardiogram;ECG)[10]是心臟收縮時所產生電壓的活動記錄。藉由心臟細胞一連串的去極化與再極化(Depolarization and Repolarization)現象[11]，促使心臟收縮讓心肌之間能互相傳遞能量，心電圖即是記錄這能量傳遞的過程。

每一個心臟細胞都含有離子溶液(鈉、鉀、鈣離子)，當細胞處於休息狀態下，細胞膜內會呈現負電荷而細胞膜外呈現正電荷，當離子移動或穿過細胞膜時，會形成一股電流，即為心電訊號。

當心臟被電激動刺激後(electrical impulse)，心臟細胞內部的離子會進而轉移，且呈正電壓，此現象稱為去極化(depolarization)。電刺激於心臟細胞之某一端，傳到另一端後細胞恢復回原來休息狀況稱之為再極化(repolarization)。再極化後細胞膜內的電荷回到原來的負電荷。再極化始於先去極化之一端。再極化完畢後呈現休息狀態，直到下一波之去極化來臨。

心臟細胞去極化與再極化的過程可由圖 2- 1 來說明。由 1 可觀察出，細胞去極化，且去極化波由單一細胞傳遞至下一個細胞。由 2 傳遞到所有細胞都已完成

去極化。由 3 完成去極化後而再極化作用。由 4 再極化作用完後可使得原本細胞復原成原本的電極性。

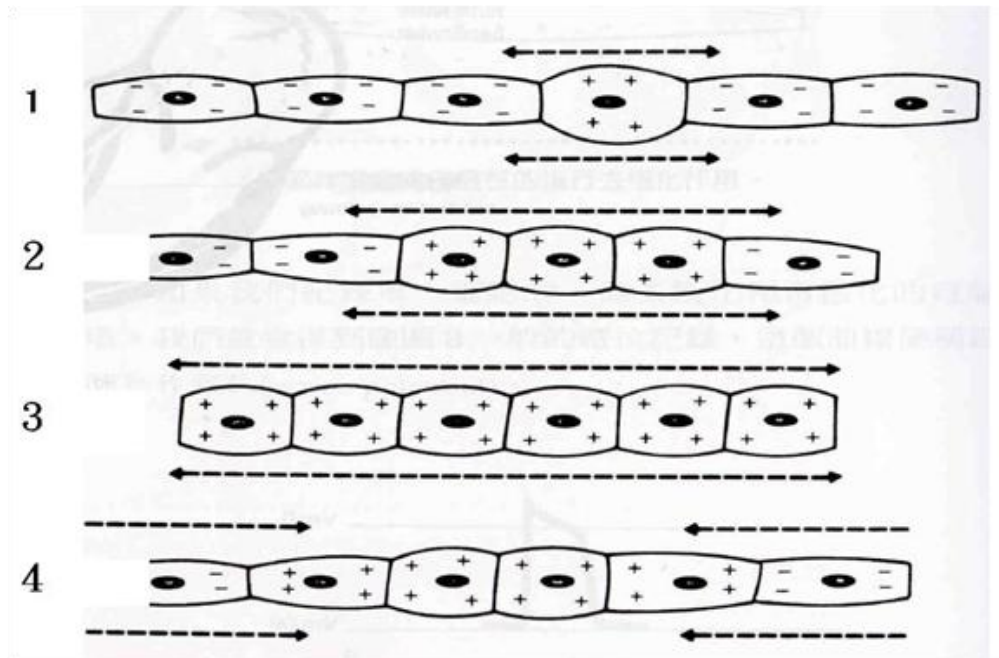


圖 2- 1 細胞去極化與再極化過程[10]

電傳導細胞屬於又薄又長的細胞，其作用就像電流的導線，能使電流有效率的傳到心臟的遠端區域，而這些細胞主要分布於心臟右心房上方，稱之為竇房結(Sinoartial Node)，如圖 2- 2，此細胞正常的活動狀況為每分鐘激活 60 到 100 次，但會隨著自主神經的活動和身體對心臟輸出量的需要而改變。

2.2 心電圖十二導程

由於心臟是一個三度空間的器官，因此在觀察心臟活動的方式也是要以三度空間的觀念來解析。把電極貼片放在身體之特定部位，利用電極正負不同組合，便可產生十二種不同角度的電壓紀錄，而每一個導程都是以獨特的角度來觀察心臟，觀察的角度越多，就能夠到更多的心臟資訊[12]。

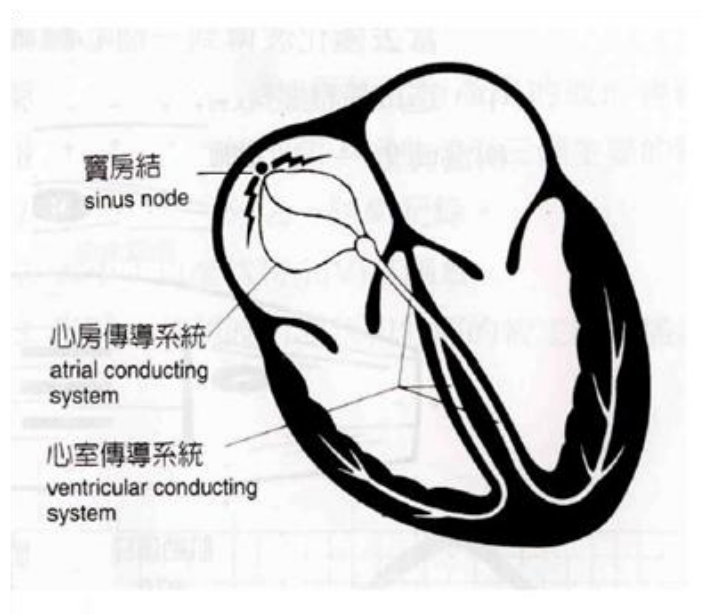


圖 2- 2 心臟電傳導細胞[12]

做十二導程心電圖時，在病人的手腕上放兩個電極，另兩個放在病人的腳踝上，這四個電極可以作出六個肢誘導，包括三個標準誘導(Standard Leads)、三個加壓誘導(Augmented Leads)。且另外六個電極並放在胸前，形成六個心前導程稱為胸前誘導(Chest Leads or Precordial Leads)圖 2- 3。

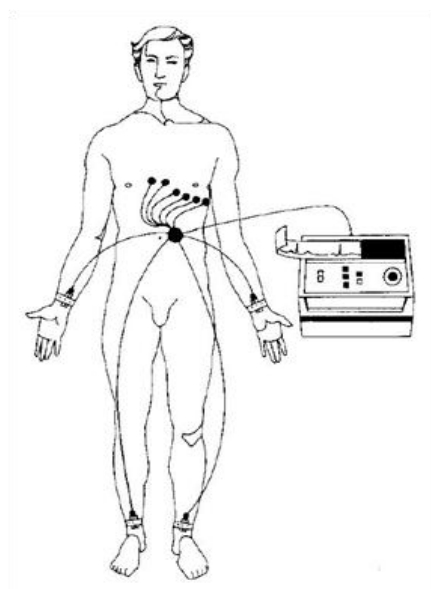


圖 2- 3 十二導程各貼片位置[10]

標準誘導又稱雙極誘導(Bipolar Leads)，利用一個負極和一個正極，記錄兩電極的電位差。此記錄方式有三種方法。分別為誘導 I、誘導 II、誘導 III，如圖 2-4 所示。

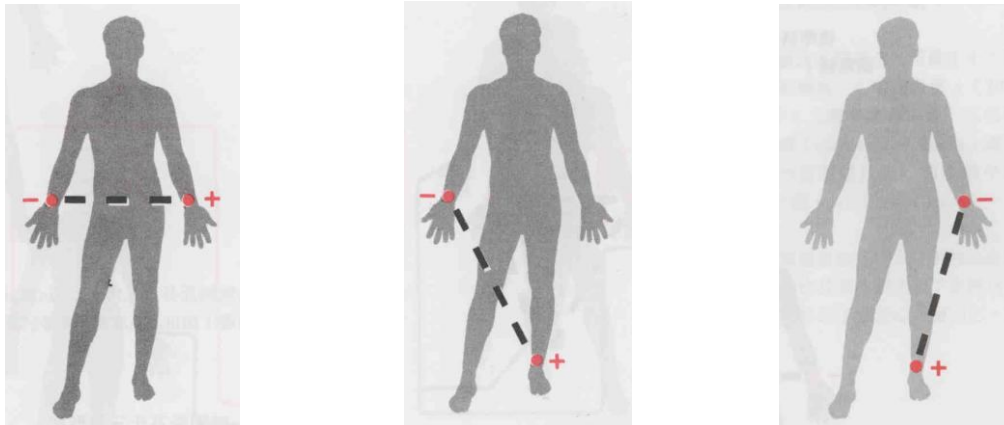


圖 2-4 誘導 I、誘導 II、誘導 III[12]

由上述的三種標準誘導可形成愛因托分三角形(Einthoven's Triangle)，其數學關係式為：誘導 I 加誘導 III 之間的高度或深度會等於誘導 II 的高度或深度，如圖 2-5 所示。

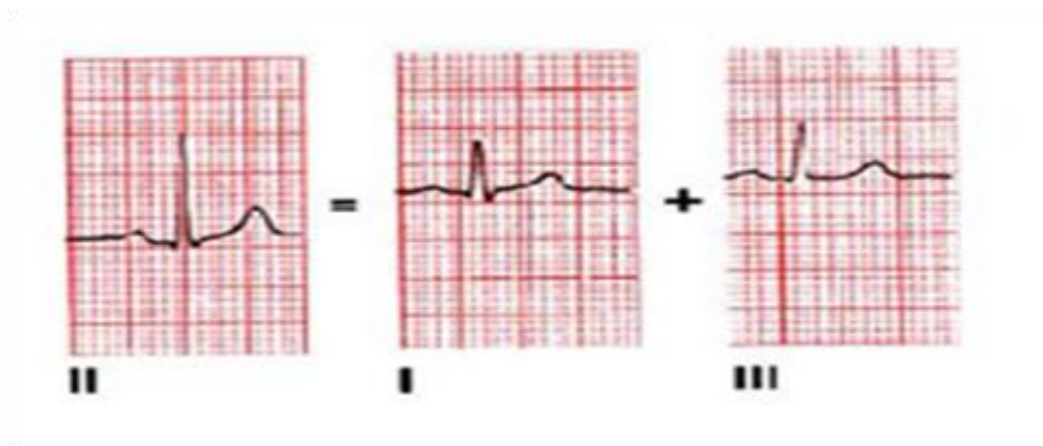


圖 2-5 愛因托分三角形[12]

加壓誘導亦稱單極誘導(Unipolar Leads)，也是利用一個負極和一個正極，記錄心電圖，只是由不同的組合方式形成。加壓誘導可分為 AVR 誘導、AVL 誘導、AVF 誘導(圖 2- 6)。

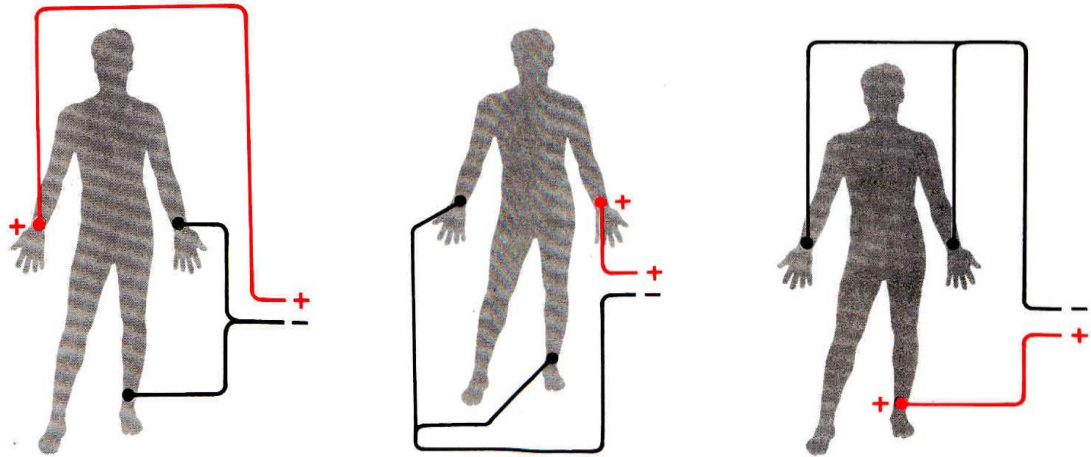


圖 2- 6 AVR、AVL、AVF[12]

胸前誘導(Chest Leads or Precordial Leads)是用六個都是屬於單極的胸前誘導貼片，每一個胸前電極都被設定為正極，而人體則被設定為共同接地極。胸前電極放置位置由第四肋間依續往左腋下貼片，胸前誘導可觀察到心臟水平面方向的電位圖，如圖 2- 7。

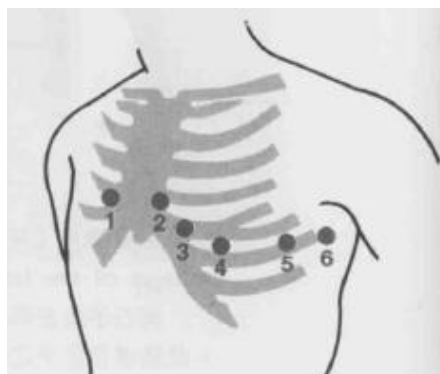


圖 2- 7 胸前誘導[12]

2.3 心電圖 PQRST 型態及判讀

標準週期心電圖包含 P 波、QRS 綜合波、T 波及 U 波(U 波不一定會出現，由心室再極化所引起的)等波型，而各個波與波之間的間段、間期都隱含一些有關心臟方面的資訊(如圖 2-8)，傳統的心臟疾病就是根據這些資訊來進行判斷的。

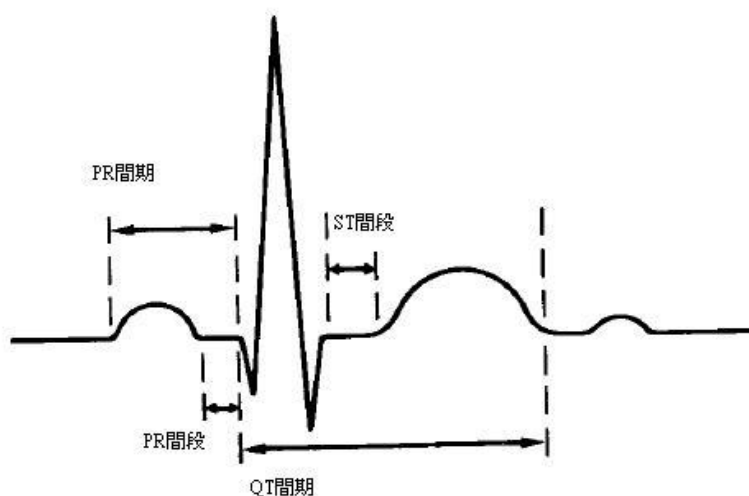


圖 2-8 心電圖區間、段、間期[14]

各種波、間段、間期所代表的涵義分別詳述如下：

P 波(The P Wave)：

從竇房節發出的去極化電波傳至左右心房所引起的左右心房之間的去極化，即表示心房去極化稱 P 波，通常呈現往上且略為圓形的正弦波。

QRS 波(The QRS Complex)：

去極化波傳導至心室間的心臟活動。由左至右的心室中隔去極化波由左側導程記錄到也就是所謂的 Q 波。心室中隔去極化波會在左心室比右心室大得多，故由竇房節所發出的電流平均會往左偏移，便可得到 R 波。在

水平面上，因心室去極化波是自右向左傳導，所以在右心室上可以記錄到一個深的 S 波(如圖 2-9 所示)。

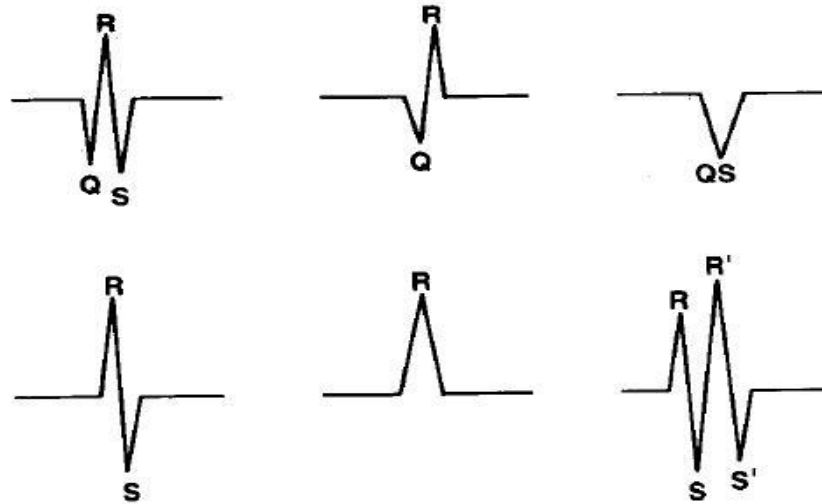


圖 2-9 各種類型 QRS 綜合波[14]

T 波(The T Wave)：

心室的再極化代表的是 T 波，T 波很容易受到外界的因素所影響。

間期與間段(Intervals and Segments)：

1. PR 間期(PR Interval)；是指 P 波前端到 QRS 綜合波前端之間的時間
代表左右心房去極化和去極化波傳導至房室節的時間。
2. PR 間段(PR Segment)；PR 間段代表 P 波與 QRS 綜合波間的時間。
3. ST 間段(ST Segment)；是指 QRS 綜合波終端與 T 波起端之間的距離
這間段是心肌梗塞極為敏感的指標。
4. QT 間期(QT Interval)；是指 QRS 綜合波前端與 T 波終端之間的時間
代表左右心室去極化和再極化的時間。

圖 2- 10 表示上述所示之各種間期與間段。

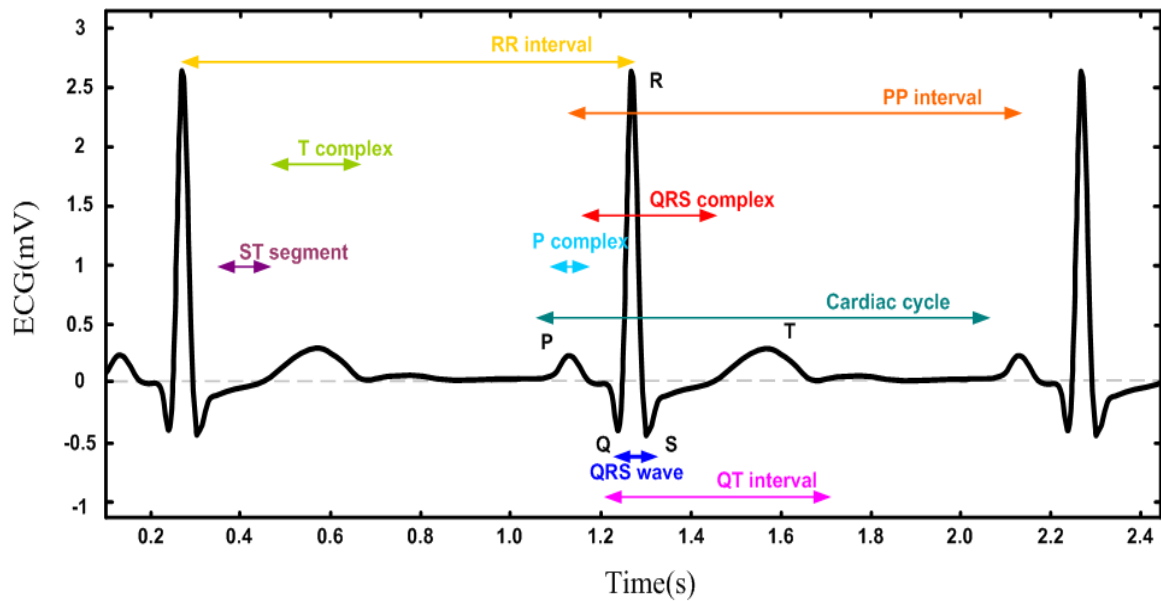


圖 2- 10 標準 PQRST 心電波形圖[12]

2.4 MIT-BIH 心律不整資料庫

本研究採用的心電訊號來自 MIT-BIH 心律不整資料庫，此資料庫由麻省理工學院(Massachusetts Institute of Technology)和 Beth Israel Hospital 研究團隊共同合作建立。其所發行的資料庫已經被世界上超過 500 個研究單位和機構所使用，學術論文也廣泛使用此資料庫作為演算法的測試標準[15][16][17]。

資料庫共有 48 組資料，每組資料長度為 30 分鐘，取樣頻率 360Hz，解析度為 11bits。每組資料中包含兩組不同的導程分別為 V2(圖 2- 11)和 MLII(圖 2- 12)的心電圖訊號。每一組資料內有一個標頭檔(Head File)、一個說明檔(Annotation File)、一個資料檔(Data File)。

- 標頭檔 (.hea)：內含資料名稱、資料長度、資料所屬的導程、取樣頻率等，以 ASCII CODE 方式儲式，且為 512 位元組。
- 說明檔 (.atr)：標示每一筆資料檔中心電圖資料的屬性，記錄 R 波與 R 波之間的距離和所屬的疾病。
- 資料檔 (.dat)：每筆資料內含兩個導程的資料，每個導程為 650000 筆，依二進制方式儲存。

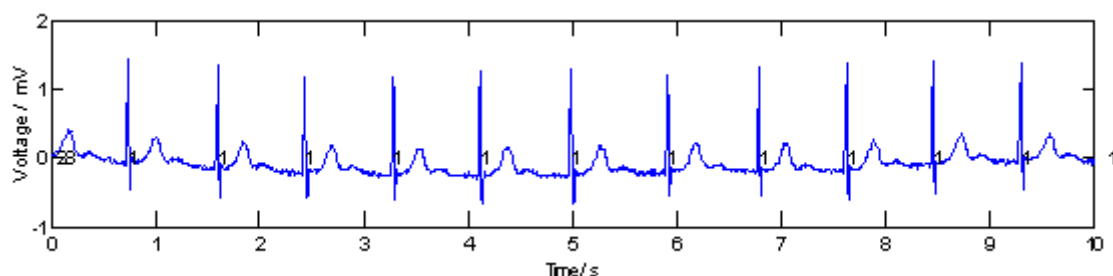


圖 2- 11 MIT-BIH 心律不整資料庫 V2 導程心電圖

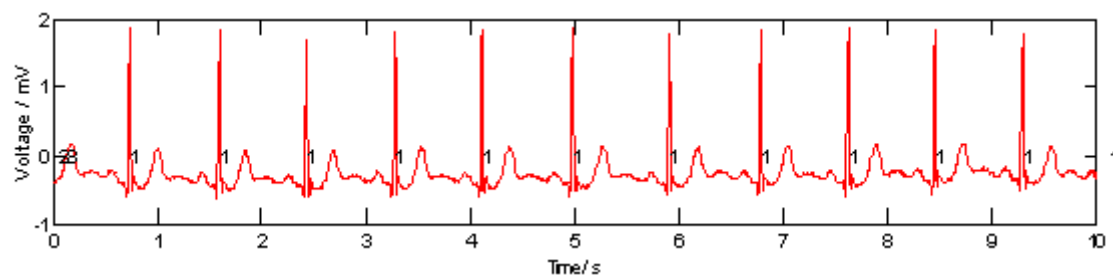


圖 2- 12 MIT-BIH 心律不整資料庫 MLII 導程心電圖

MIT-BIH 所收錄的心電資料可分為以下十二種不同的形態(表 2- 1):

MIT/BIH 標註符號	Type of arrhythmia	病症形態
N	Normal Beat	正常波形
L	Left bundle branch block beat	左分支束阻塞
R	Right bundle branch block beat	右分支束阻塞
a	Aberrated atrial premature beat	變形的心房早期收縮
V	Premature ventricular contraction	心室性早期收縮
F	Fusion of ventricular and normal beat	心室融合心跳
J	Nodal (junctional) premature beat	結節性過早收縮
A	Atrial premature beat	心房性過早收縮
E	Ventricular escape beat	心室性脫逸收縮
j	Nodal (junctional) escape beat	結節性脫逸收縮
/	Paced beat	心律調節器心跳
F	Fusion of paced and normal beat	節律器融合心跳

表 2- 1 心電圖病症代碼表

MIB-BIH 資料庫內各病症波形示意圖，整理如下：

Normal beat 正常波形 (N)(圖 2- 13)：

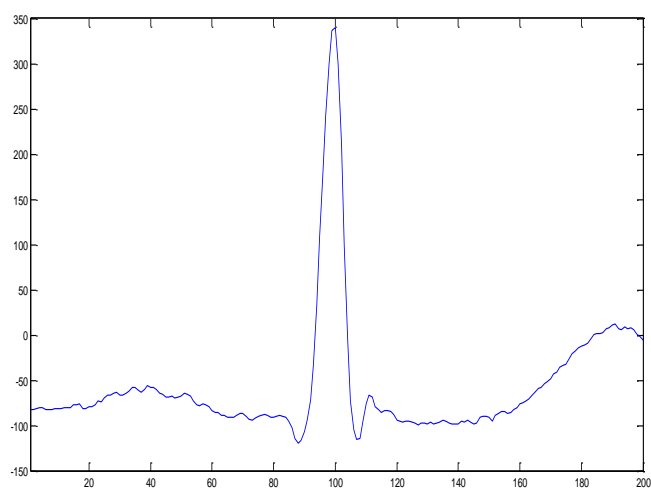


圖 2- 13 正常心搏波形

Left bundle branch block beat 左分支束阻塞 (L) (圖 2- 14): 圖 2- 14 左分支束阻塞
波形

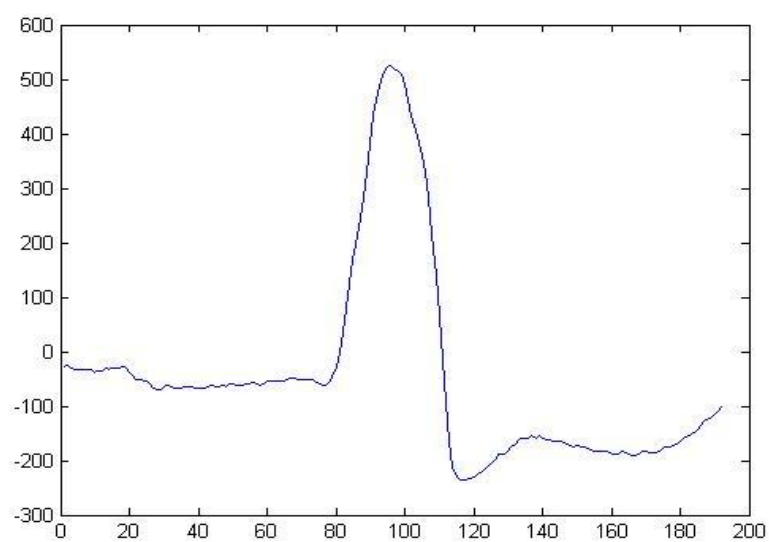


圖 2- 14 左分支束阻塞波形

Right bundle branch block beat 右分支束阻塞 (R) (圖 2- 15):

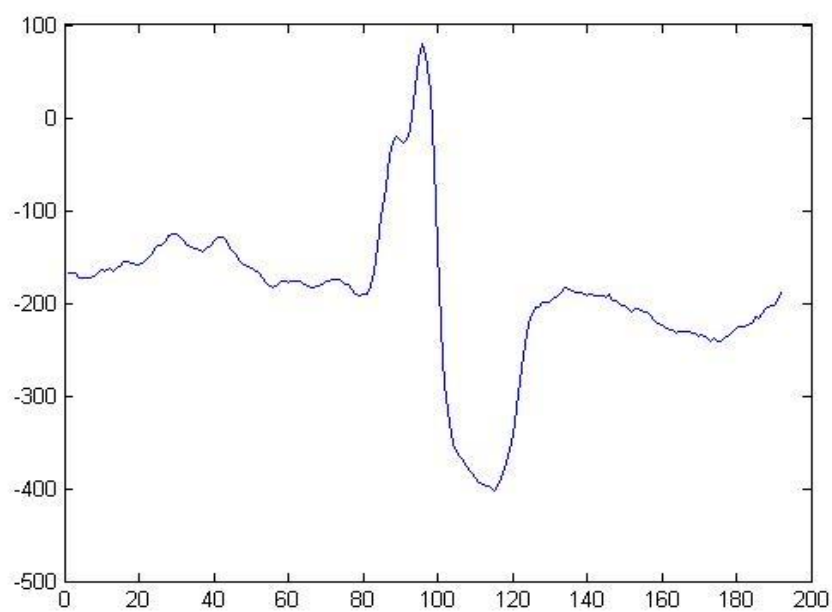


圖 2- 15 右分支束阻塞波形

Aberrated atrial premature beat 變形的心房早期收縮 (a) (圖 2- 16) :

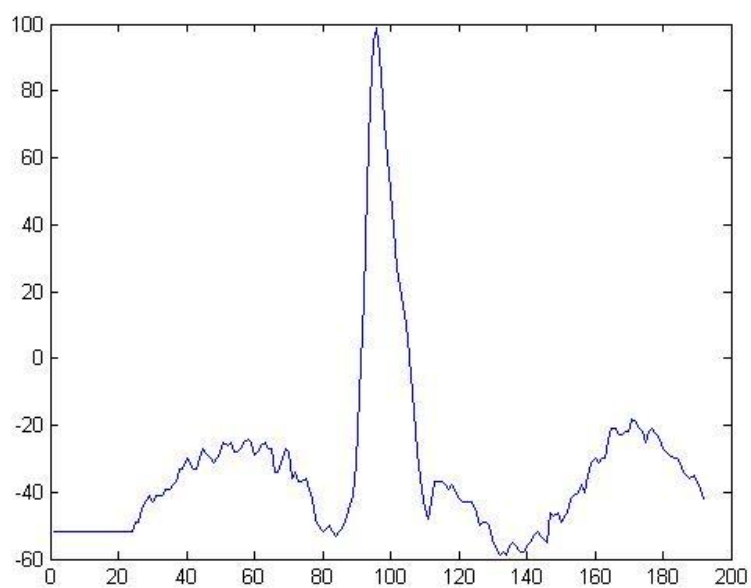


圖 2- 16 變形的心房早期收縮波形

Premature ventricular contraction 心室性早期收縮 (V) (圖 2- 17) :

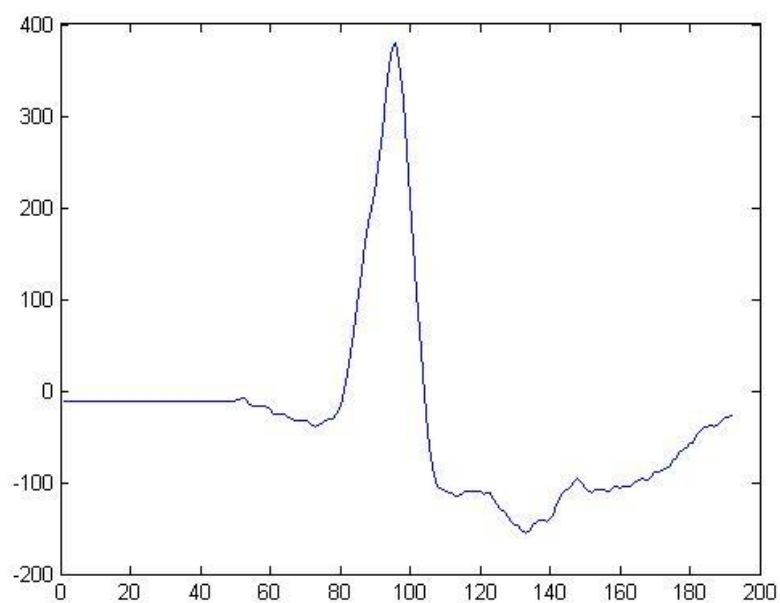


圖 2- 17 心室性早期收縮波形

Fusion of ventricular and normal beat 心室融合心跳 (F)(圖 2- 18) :

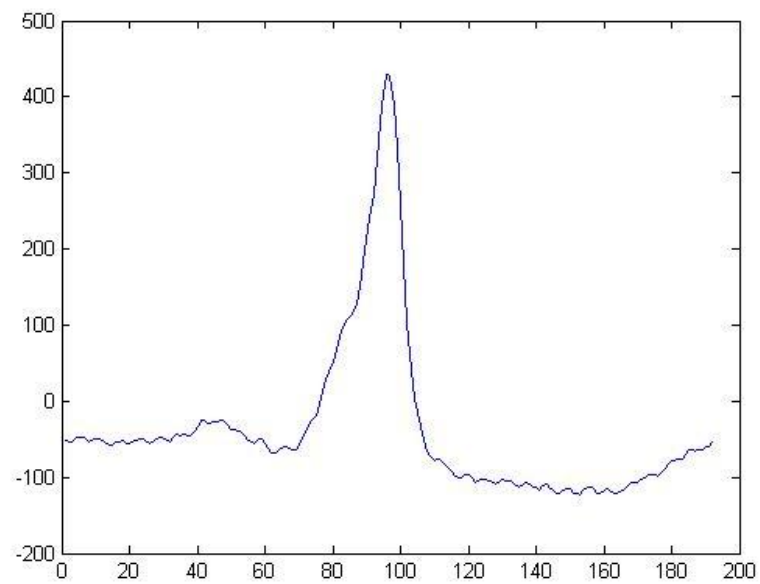


圖 2- 18 心室融合心跳波形

Nodal (junctional) premature beat 結節性過早收縮 (J) (圖 2- 19) :

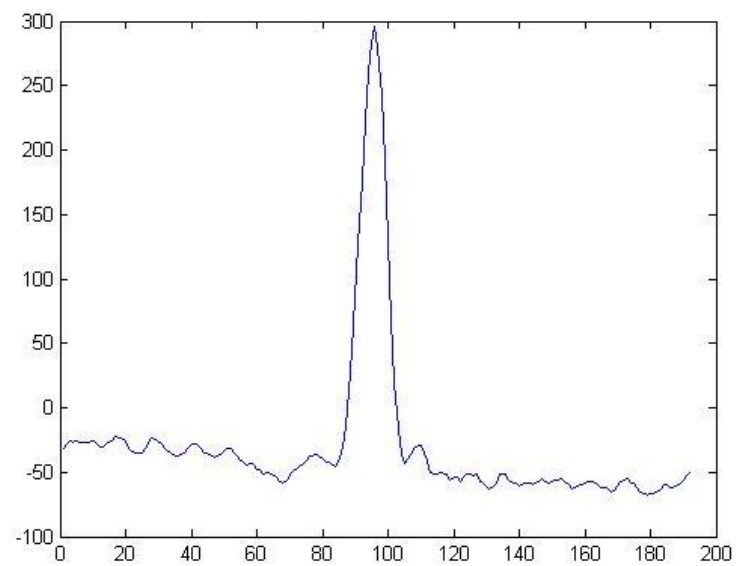


圖 2- 19 結節性過早收縮波形

Atrial premature beat 心房性過早收縮 (A) (圖 2- 20) :

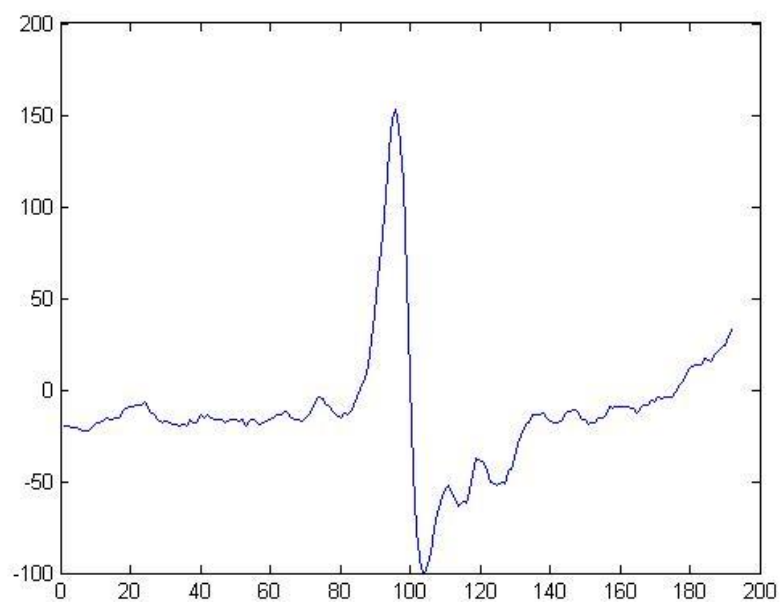


圖 2- 20 心房性過早收縮波形

Ventricular escape beat 心室性脫逸收縮 (E) (圖 2- 21) :

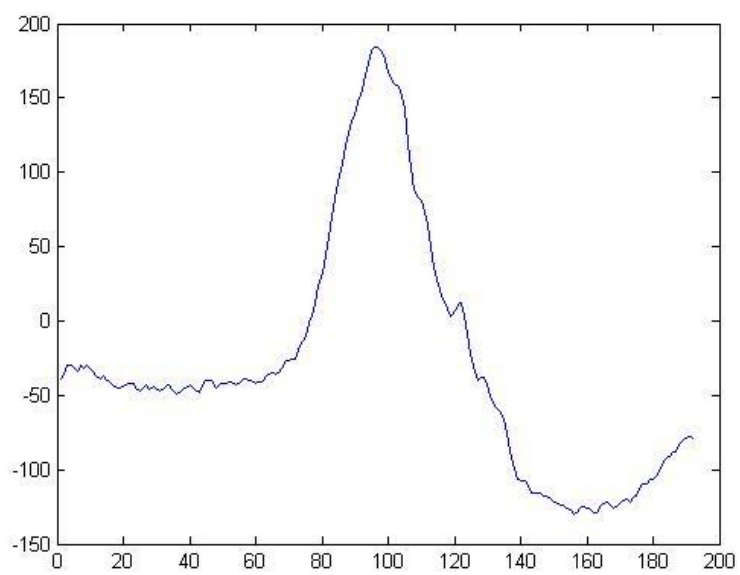


圖 2- 21 心室性脫逸收縮波形

Nodal (junctional) escape beat 結節性脫逸收縮 (j) (圖 2- 22) :

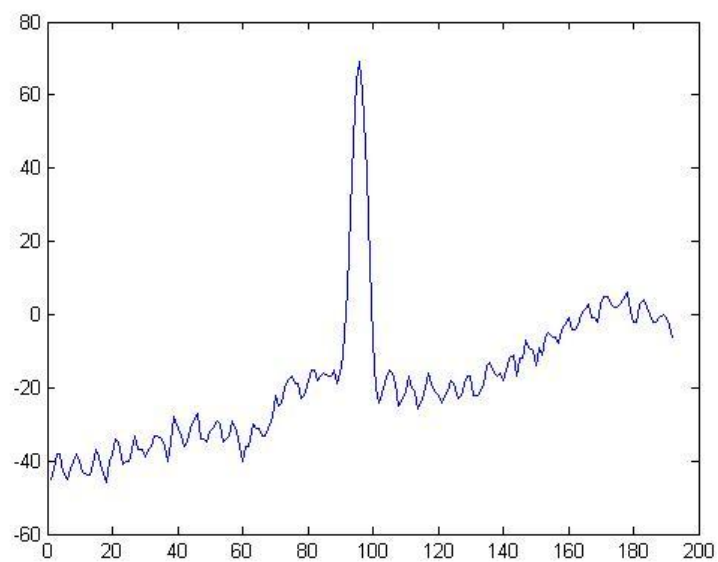


圖 2- 22 結節性脫逸收縮

Paced beat 心律調節器心跳 (/) (圖 2- 23) :

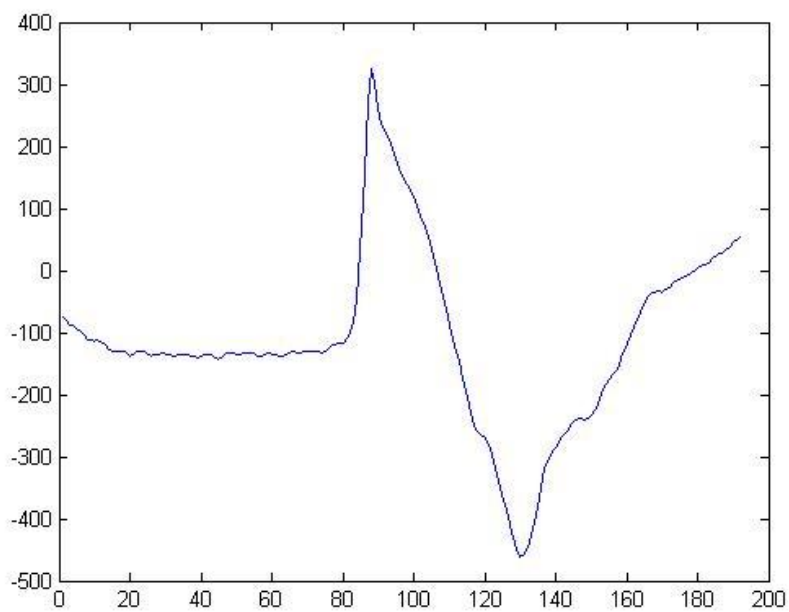


圖 2- 23 心律調節器心跳

Fusion of paced and normal beat 節律器所融合心跳 (F) (圖 2- 24) :

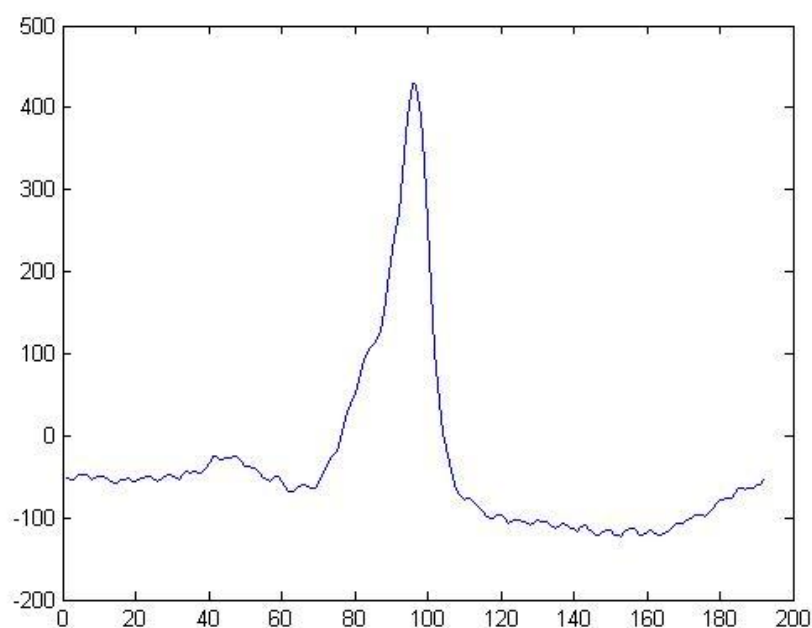


圖 2- 24 節律器所融合心跳

2.5 心電訊號相關的干擾

心電訊號量測時的干擾主要有基準線飄移、肌肉波、市電頻率干擾等。茲詳述如下：

基準線飄移(Baseline Wander)：

此種人為干擾的頻率約為 0.3Hz，主要為患者量測時過於緊張導致呼吸急促或患者皮膚與電極貼片接觸不良所造成，如圖 2- 25 所示。

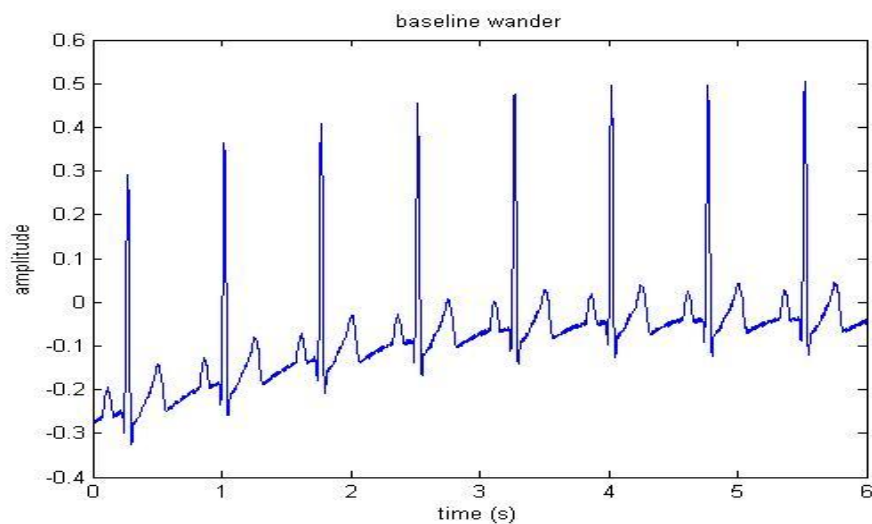


圖 2- 25 心電訊號基準線飄移[32]

肌肉波(Eletromyographic)：

主要分布的頻率為 6~100Hz 之間，因患者本身過於緊張刺激和量測時移動導致肌肉顫動，如圖 2- 26 所示。

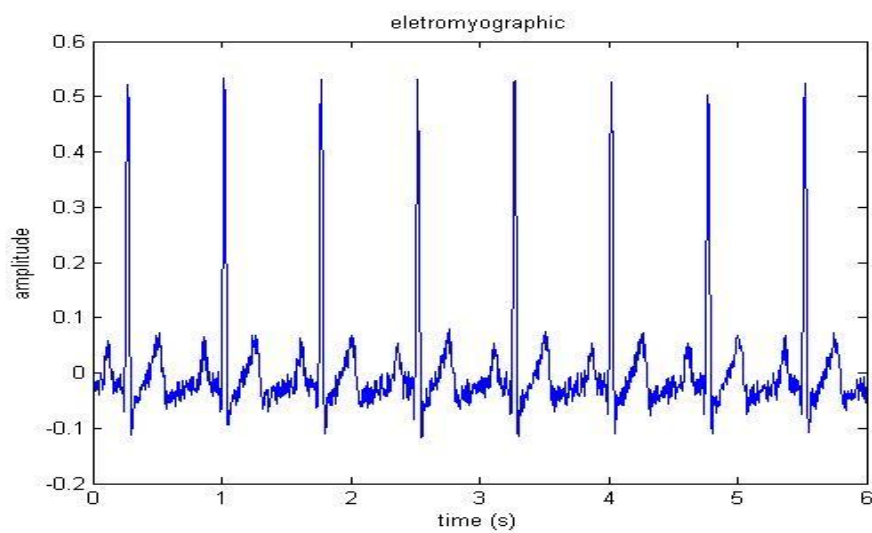


圖 2- 26 心電訊號肌肉波[32]

市電頻率干擾(Power Line Interference)：

分布頻率為 60Hz，主要是在量測儀器受到 60Hz 交流電源的干擾所造成，如圖 2-27 所示。

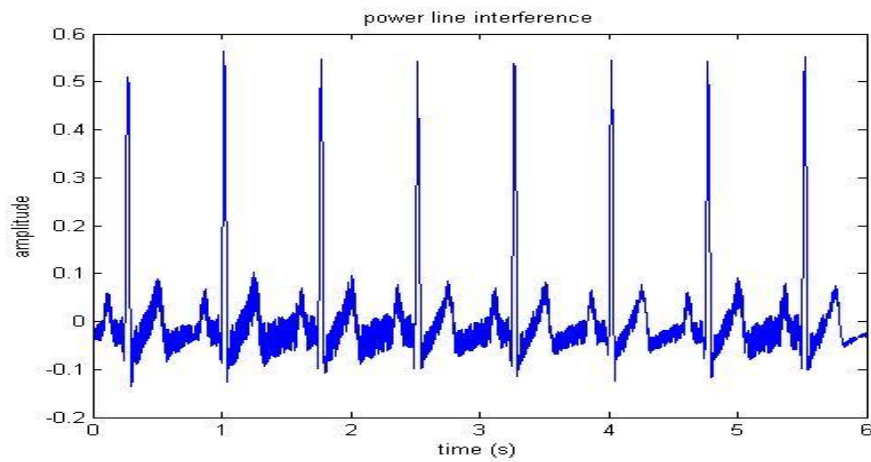


圖 2-27 心電訊號市電頻率干擾[32]

第三章 系統架構、心電圖辨識演算法及相關分析理論

在本章中，我們將介紹心電圖辨識系統的整體架構、去除基準線漂移的濾波器、訊號切割、K-means 分群演算法、訊號的特徵擷取、適應性特徵選取、支持向量機分類器等流程的原理與應用。

3.1 心電圖辨識系統流程簡介

本論文所提出的心電圖辨識系統，主要分為五大區塊，如圖 3-1 所示：

- (一)、 將量測的連續心電訊號，輸入至本心電圖辨識系統，經本系統的方法把連續心電訊號切割為實驗用的樣本，每一個樣本中包含著一個完整心電訊號波形。
- (二)、 同一種病症心電圖波形可能差異甚大，故將相同病症的樣本資料做粗略分群，這動作能大大提升分類器辨識率。
- (三)、 將心電訊號樣本分為訓練辨識模型用與測試用兩種，並擷取樣本中隱含的特徵，本系統將會依據這些特徵進行辨識。
- (四)、 把訓練辨識模型用的樣本特徵放置支持向量機分類器中訓練。
- (五)、 將測試用樣本的特徵丟入辨識模型中，並記錄辨識結果。

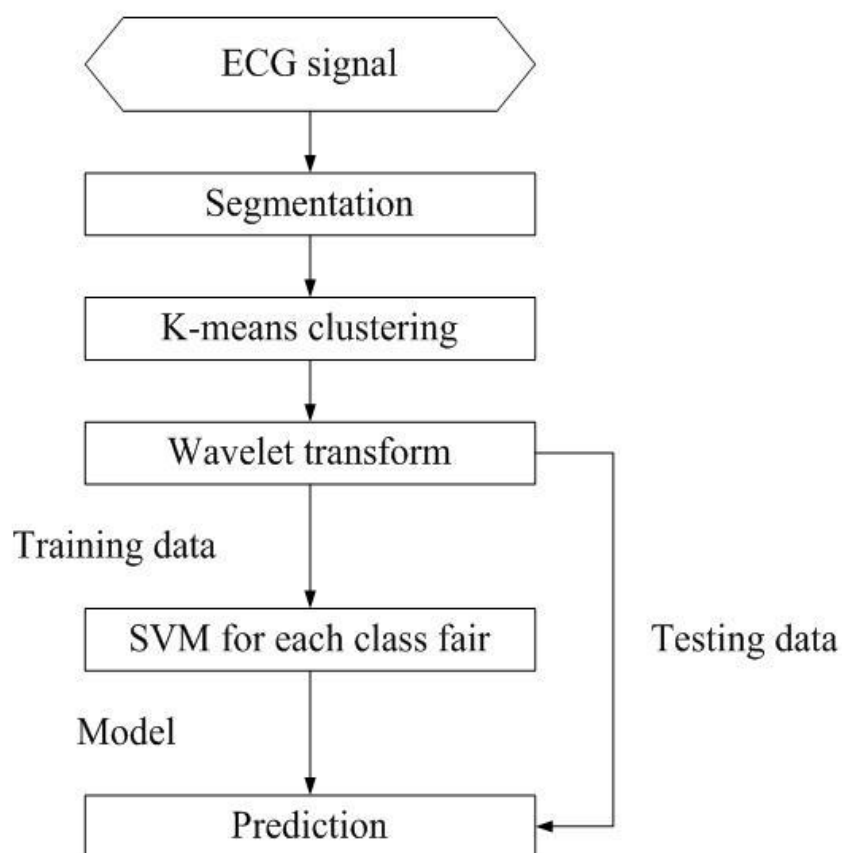


圖 3- 1 心電圖辨識系統流程圖

以下，我們將針對每個流程做較完整的說明。

流程一：因為 R 波是一個心博週期中變化最為明顯的部位，因此首先偵測 R 波的位置，再以 R 波的位置向前後搜尋其他波形，接著切割出一個完整的心跳波形作為實驗的樣本。

流程二：對同一類病症的實驗樣本做初步的分群，樣本經過 K-means 分群演算法分群後，同一類心臟疾病中波形相似的樣本將被分在一起，故同一類心臟疾病會被分為許多次類別，此初步分類能提高系統的辨識率。

流程三：抽取每個樣本內隱含的特徵，本系統的特徵擷取分為頻域特徵、統計特徵及時域特徵三類，茲分述如下：

- 頻域特徵：使用小波轉換做六階的頻帶分解，分別取出高、低頻小波係數。在一般心電圖訊號分解過程中，心電圖訊號在不連續點或是轉折點處經過小波轉換，能夠在多重解析度中呈現明顯且不一樣的特徵。在低頻的部份，能突顯波形中的趨勢部份，而高頻部份能突顯變化較劇烈的特徵。
- 統計特徵：由上述小波轉換所得小波係數和心電圖的波形，分別計算出每一頻帶絕對值的平均值、絕對值的最大值、平均能量、及頻帶之間平均值的比值等。
- 時域特徵：時域特徵蒐集了傳統以心電波形來判斷心臟疾病的相關準則，包含 R-R 間段的距離、R-R 間段與後五個心搏週期的比值、R-R 間段與後一個心搏週期的比值、R-R 間段與前一個心搏週期的比值、P 波與 R 波的振幅比值、R-S 波之間的落差值、R-R 間段與 R-S 間段的比值、R-R 間段與 P-R 間段的比值、R 波與 S 波之間距離等。這些特徵各會對應到不同種類的心電疾病，例如：R-R 間段與前後數個心搏週期的比值即是用來判斷心律是否不整的準則之一[20]。

流程四：把擷取出的特徵資料取其每一個次類別樣本總數的一半，放入支持向量機的分類器中訓練。在訓練的過程中，我們比較兩類之間的 hyper-plane 是否為最佳化，並歸納出兩兩次類別之間特徵值的差異，以計算出本身的支持向量。之後建立出兩兩次類別的分類模型。

流程五：將其餘的資料，運用流程四所訓練出來的支持向量機來辨識。當我們將資料輸入模型中辨識時，一個模型會認定該筆資料為某一次類別，當資料通過所有模型辨識後，系統會根據每一個類別的得票數來決定該樣本最後的歸屬類別。

3.2 零相位延遲高通濾波器

如第二章所述，心電訊號經常會有基準線漂移的干擾，基準線干擾會影響小波轉換係數(也就是我們所採用的時頻特徵)，進而影響到辨識系統整體的辨識率，因此，我們需要設計一高通濾波器來濾除此基準線飄移干擾。

一般的數位濾波器常會有相位延遲的現象，此現象會造成心電圖波形失真，進而影響到時頻特徵的擷取。為了避免這個現象發生，我們採用了零相位延遲的設計理念。

零相位延遲的實現方式是將有限長度的輸入訊號，先通過數位濾波器 $G(Z)$ 的處理，其輸出再經過時間逆轉處理，然後再將此時間逆轉處理後的結果，輸入同一個數位濾波器 $G(Z)$ ，最後將輸出的結果再做一次時間逆轉處理，如此就可以得到零相位的數位濾波器(如圖 3- 2 所示) [32]。

假設 $X(e^{j\omega})$ 為輸入訊號 $x[n]$ 的傅立葉轉換， $Y(e^{j\omega})$ 為輸入訊號 $y[n]$ 的傅立葉轉換，則

$$\begin{aligned} Y(e^{j\omega}) &= W_2^*(e^{j\omega}) = G^*(e^{j\omega})V^*(e^{j\omega}) = G^*(e^{j\omega})W_1(e^{j\omega}) \\ &= G^*(e^{j\omega})G(e^{j\omega})X(e^{j\omega}) = \|G(e^{j\omega})\|^2 X(e^{j\omega}) \end{aligned} \quad (3.1)$$

其中 $G(e^{j\omega})$ 為數位濾波器 $G(Z)$ 的頻譜響應。

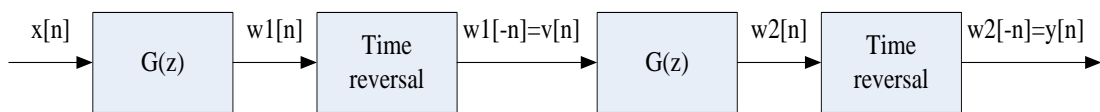


圖 3- 2 零相位延遲濾波器流程圖[32]

本論文用來濾除基準線飄移的零相位延遲高通濾波器為 Butterworth 高通濾波器，其轉移函數如下：

$$Y(z) = \frac{0.8187 - 2.24561z^{-1} + 2.4561z^{-2} - 0.8187z^{-3}}{1 - 2.6012z^{-1} + 2.278z^{-2} - 6.702z^{-3}} X(z) \quad (3.2)$$

實驗證明，本濾波器可以有效濾除 0.3Hz 的基準線飄移干擾。

3.3 Segmentation(訊號切割)

本論文所採用導程為 MLII，因 MLII 是電壓變化最為明顯的導程，可清楚觀察心臟變化。本辨識系統是辨識單一心跳的病徵，因此將連續的心電訊號切割成單一心跳的實驗樣本，偵測心電圖訊號 PQRST 波的相關研究已經有很多的方法被提出，本論文採用文獻[21]所提的 PQRST 波偵測方法，其演算法流程圖，如圖 3-3 所示。

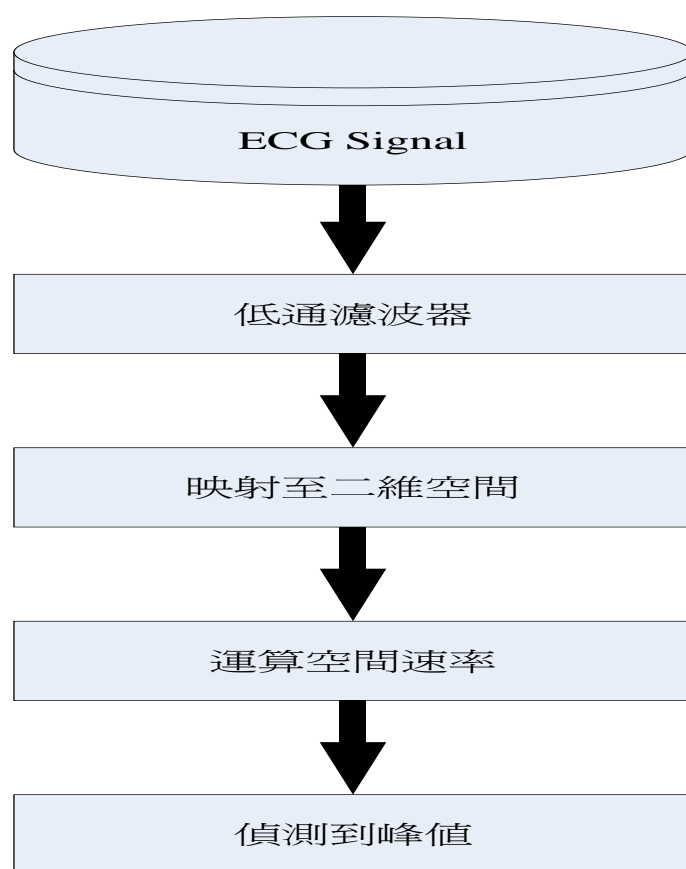


圖 3-3 QRS 波偵測流程圖

QRS 波偵測演算法中的低通濾波器的功能是去除心電圖訊號中的高頻成份，如圖 3- 4 所示。接著，我們再對通過低通濾波器的心電圖訊號作 5 單位的时间延遲，再映射到二維空間判斷其與原始訊號的相關性圖 3- 5，這樣可以使心電圖訊號中的每一個特徵點都很容易的辨識出來。然後再計算出心電圖訊號的空間速率，空間速率可以判斷出訊號對於空間以及速率的影響，如圖 3- 6。由此圖可明顯看出，波形較為突出的地方是峰值所在的位置，此偵測區域最大值的方法便能判斷出峰值位置。峰值位置與原始波中的 R 波位置幾乎一樣，因此該峰值的位置便是 R 波位置。

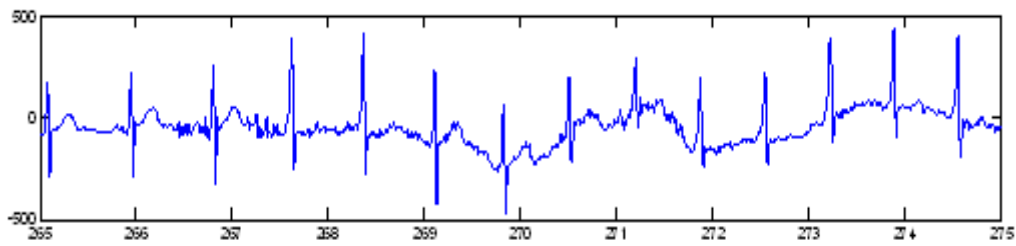


圖 3- 4 原始心電圖訊號

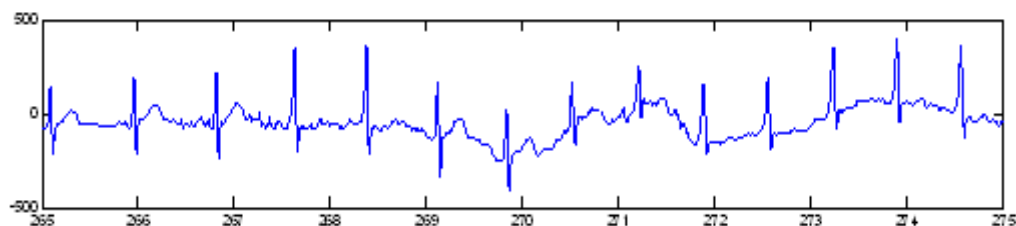


圖 3- 5 低通濾波器處理後的心電圖訊號

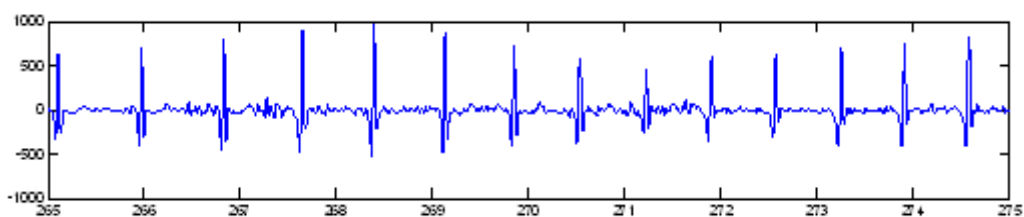


圖 3- 6 同一基準點心電圖訊號

一般心搏重要的部份都集中於 PQRST 綜合波中，而每一段正常的心搏週期 P 波至 T 波所經過的時間為 0.5 秒到 0.88 之間，而擷取一個心搏週期的時間至少要 0.5 秒，本論文所提出的心電圖辨識系統擷取一段心搏為 192 個樣本數，所需的時間大約 0.52 秒左右，而這樣所擷取的長度可覆蓋整個 MIT-BIH 心律不整資料庫心電圖整個 PQRST 綜合波範圍，如圖 3-7。

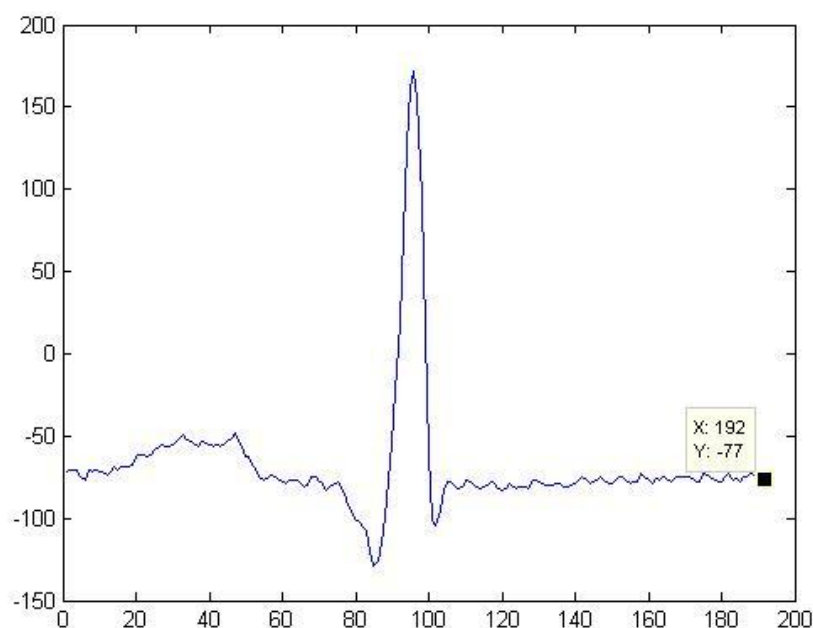


圖 3-7 心電圖波形 192 個樣本數示意圖

3.4 K-means 演算法分類器

K-means 分群演算法(K-means clustering)[22]是屬於最為典型的分割式分群演算法(Partition clustering)[23]，而主要應用是要在大量高維度的資料點，找出最具代表性的資料點。這些所謂的資料點可以稱為群聚中心(cluster centers)，並根據這些群聚中心來進行資料分類，用少數的代表點來代表特定的類別資料，可以大為降低系統的計算量和資料量。

此分群演算法可以將資料依據分群的目標分成許多群聚，而被劃分為同一群資料會有某些特性極為相似，相反的被分為不同群的資料就會有某些特性會有明顯的不同。此方法須先預定群聚的數目，經由反覆疊代運算，逐次降低每一個目標函數的誤差值，直到目標函數不再產生變化，就能產生分群的效果。此演算法已成功應用於不同的領域，從市場分析、電腦視覺、統計學、農業、天文學等。

在使用分割式分群演算時，是希望使每一個群聚中心，每一個資料點，與群聚中心的距離都能有最小的平方誤差(square error)。如果我們預設資料內含有 h 個群聚中心，其中第 k 個群聚可以用集合 G_k 來表示，假設 G_k 含有 n_k 筆資料

$\{x_{1k}, x_{2k}, x_{3k} \dots x_{n_k}\}$ ，此群聚中心為 y_k ，此群聚的平方誤差 e_k 則可定義為：

$$e_k = \sum_i |x_{ik} - y_k|^2 \quad (3.3)$$

其中 x_{ik} 屬於群集 G_k 。

則 h 個群聚數的總和平方誤差 E 即是每個群聚平方的總和，稱為分群的「誤差函數」(error function)：

$$E = \sum_{k=1}^h e_k^2 \quad (3.4)$$

k-means clustering 就是一種可以找出 h 以及相關群聚中心使得 E 值為最小的演算法。

k-means 演算法的流程如下：

- 步驟一：設定分群群聚數目為 h ，並選取 h 個樣本作為 h 個群聚的群聚中心。
- 步驟二：輸入全部的樣本，計算每一筆樣本到各個預設群集中心之間的距離，然後再比較該筆的樣本距離哪一個群集中心最為接近，這筆資料記錄就會被指派到最為接近的那個群集中心，此時就會產生了初始群集的成員集合。
- 步驟三：再根據群內的每一個樣本重新計算出該群集的質量中心，利用新的質量中心來做為該群新的群集中心。指定完新的群中心之後，再一次比較每一筆樣本與新的群集中心之間的距離，然後根據距離，再度重新分配每一個資料所屬的群集。
- 步驟四：持續反覆步驟三，一直執行到群集成員不再變動為止，如圖 3-8。

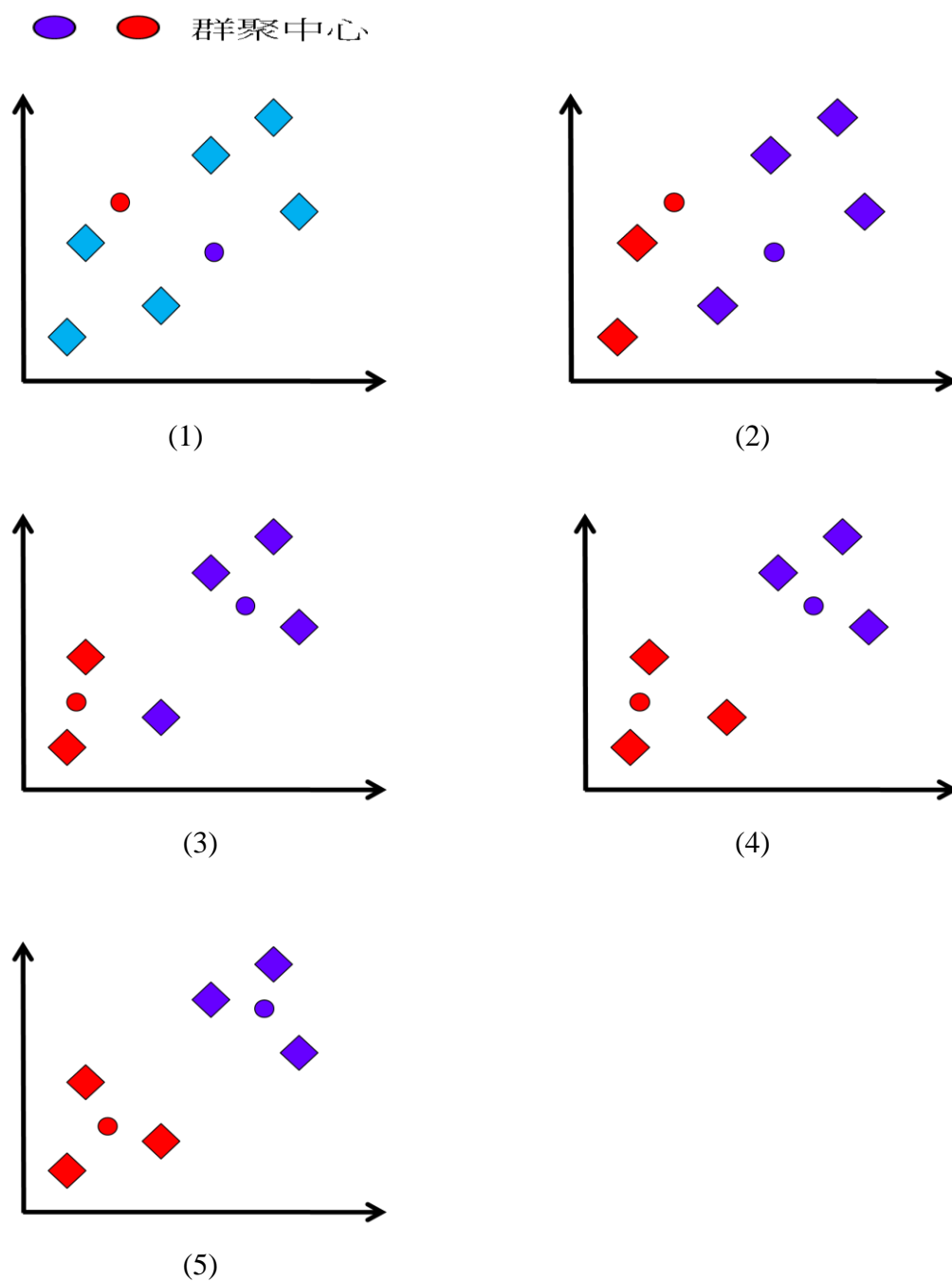


圖 3- 8 K-means 分群過程示意圖

本論文改良了上述的方法，我們的方法不需事先預設群聚個數，而是在所有樣本中隨機挑選一個當作初始的群聚中心，然後再將資料逐筆加入該群聚中心，然後重新計算群聚中心，若該筆資料使群聚中心移動的太遠，則將該筆資料設為

新的群聚中心，因此群聚的數目會依輸入資料的特性不同，而做適應性的調整，我們稱此方法為適應性 K-means 演算法。

適應性 K-means 演算法流程如圖 3-9 所示，圖 3-10 為本演算法示意圖：

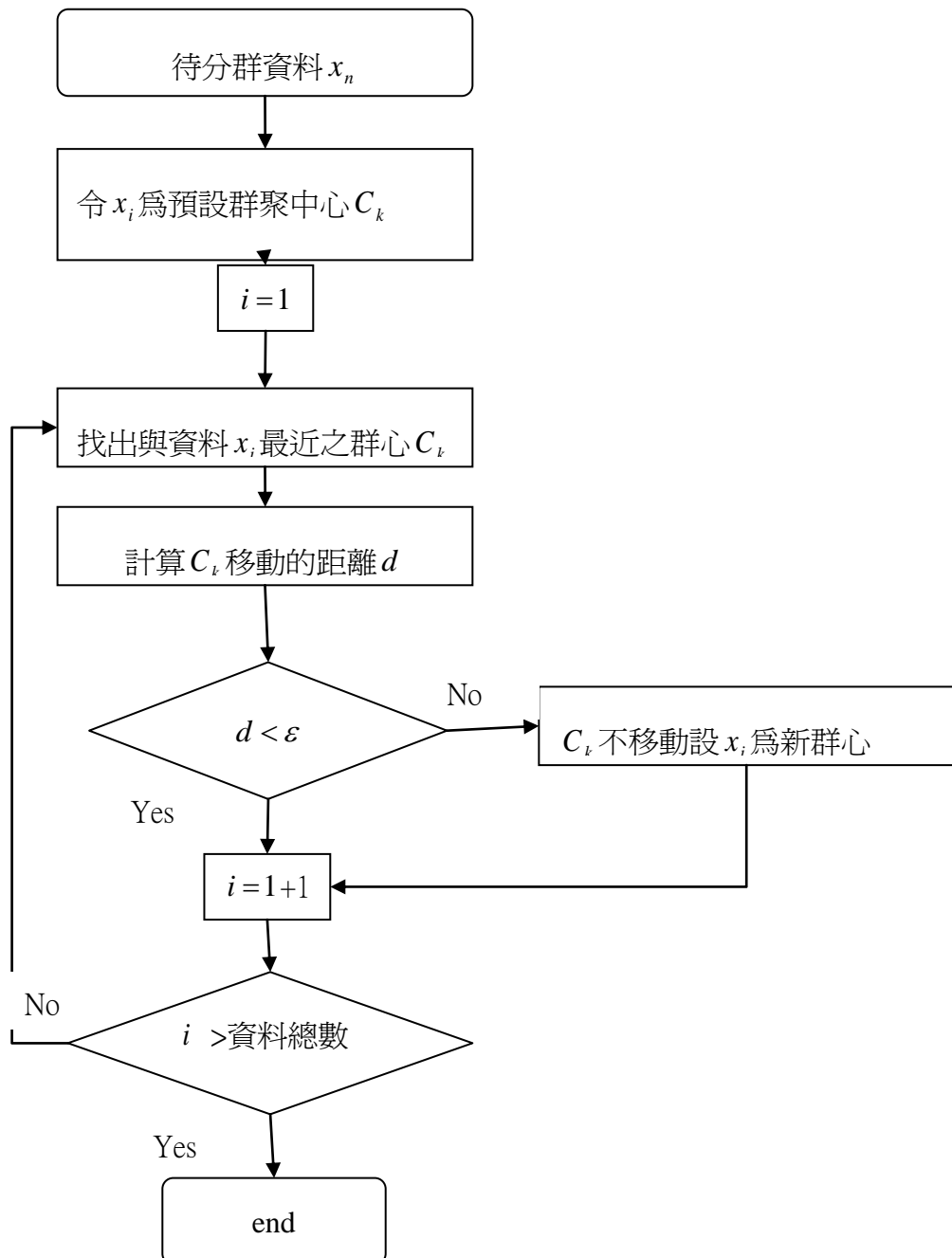


圖 3-9 適應性 K-means 演算法流程圖

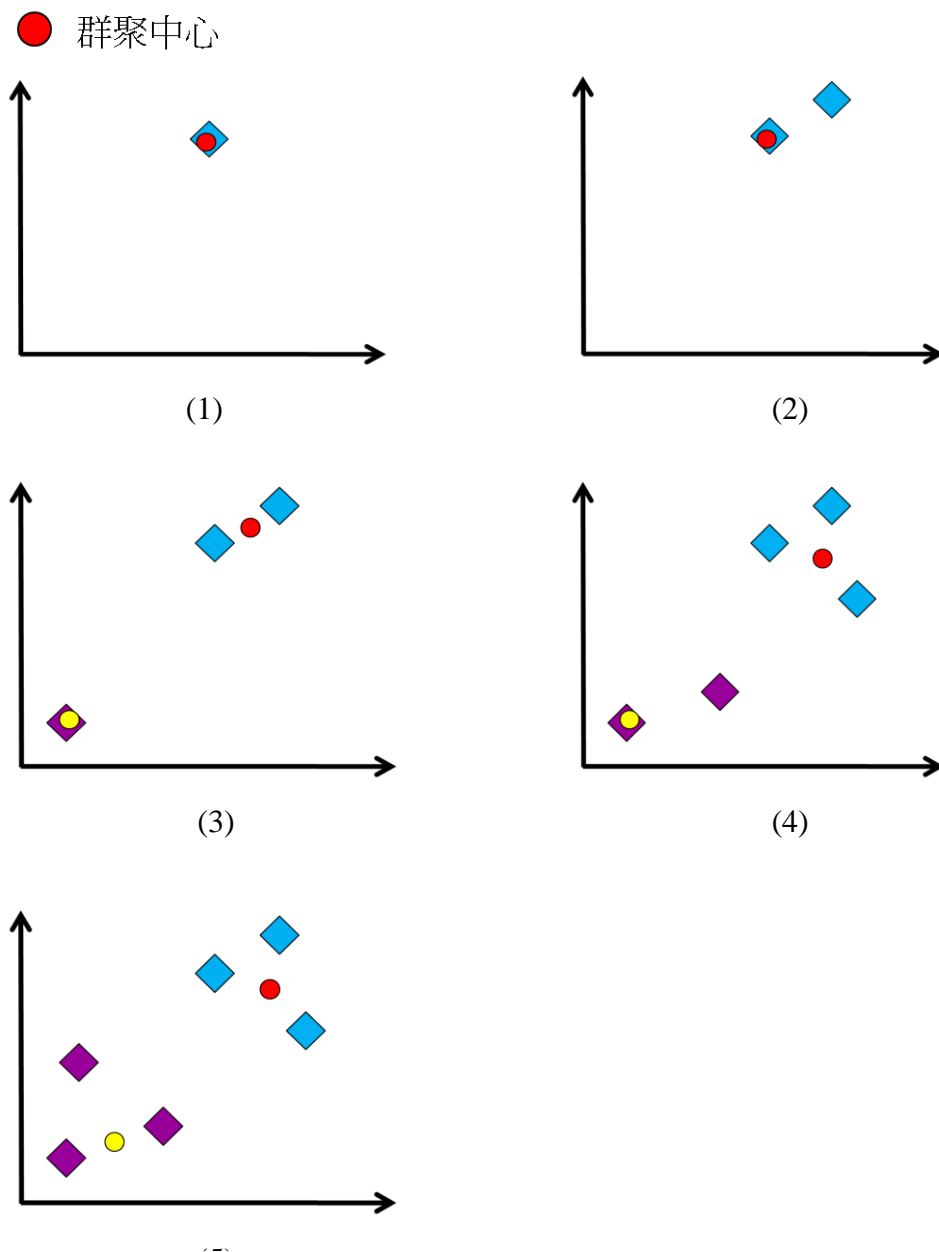


圖 3- 10 適應性 K-means 演算法示意圖

接下來我們將以 MIT-BIH 心律不整資料庫為例來測試本論文所提適應性 K-means 演算法的效能。MIT-BIH 心律不整資料庫中的資料共有 12 種病症，本研究採用其中的九種病症，另三種病症，因資料樣本數太少，樣本間的變異性也太大，故本論文將不考慮這三種病症。

我們觀察到即使同一類的病症也會有不同的波形，如病症 Left bundle branch block beat 就至少有三種波形用肉眼就能判斷發現其不同(如圖 3- 11 (A))，因此在訓練時不可以把病症視為單一種類，否則樣本太過分散會影響最後的辨識率。

病症 Left bundle branch block beat 的樣本，經過本論文所提出之適應性 K-means 演算法進行粗略的分群後，便可分出四種次類別以上如圖 3- 11 (1)、(2)、(3)、(4)。

接下來，我們將這個適應性 K-means 演算法應用到兩種不同資料的分群上：

- 第一種資料：心律不整資料庫中的心電訊號樣本，不經過 3-2 節所提的零相位誤差濾波器進行基準線飄移干擾的濾除，因此樣本中含有基準線飄移干擾。我們使用本論文所提出的 K-means 演算法進行本資料進行分類，可將原本 9 種心臟病症分類出 72 種次類別。
- 第二種資料：將心律不整資料庫的心電訊號經過 3-2 節所提的零相位誤差濾波器進行基準線飄移干擾的濾除，再由本論文所提之適應性 K-means 演算法分類，結果將 9 種心臟病症分類成 47 種次類別。各病症所分出次類別詳細資料如表 3- 1 所示

這個實驗說明了，移除基準線飄移之後，再利用適應性 K-means 分群演算法進行分類，將可以得到數量比較少的次類別。由於我們所採用的支援向量機分類法必須建立兩兩類別之間的模型，因此次類別越少，我們所需的建立的支持向量

機模型也會跟著變少，這將會節省建立分類器模型所需的時間，並提升本辨識系統的效率。

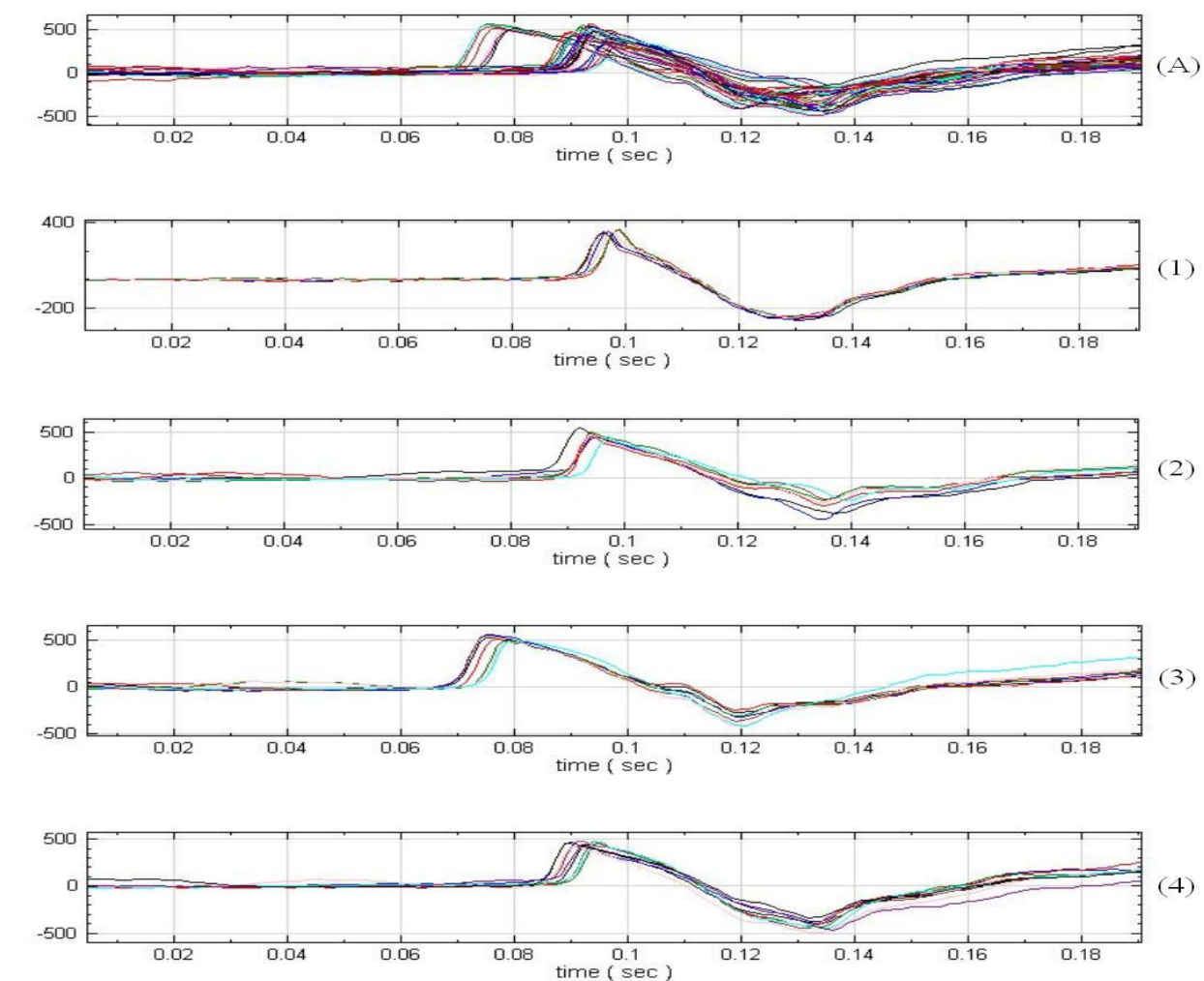


圖 3- 11 心電圖訊號分類示意圖

MIT/BIH 資料庫 病症名稱	次類別數量	
	含有雜訊	濾波後
N	33	16
L	6	6
R	4	4
V	9	10
A	3	3
E	1	1
j	4	1
/	8	4
F	2	2
Total	70	47

表 3- 1 病症分類統計圖

3.5 小波轉換演算法

小波分析起源於 1910 年 Haar 所提出的小波規範正交基底[24]，經過數十年沉寂，在 1984 年由法國物理學家 Morlet 重新發現，並將其概念引入訊號分析中，而後，Grossmann 對訊號作一個可對其函數的伸縮、平移展開研究與定義，並正式將此類函數命名為小波(Wavelet)。1986 年 Meyer 建構出具有衰減性的平滑小波函數，而這個函數的二進制伸縮與平移再構成正交基底。往後一年，Mallat 透過多重解析度分析來產生小波基底，建構了小波函數的構造及方法，讓訊號透過小波函數能加以分解與重建。1988 年由 Daubechies[27][28]提出具有正交性以及緊支撐這兩種特性。到此，小波分析理論才有了初步的建立。經過二十年來的發展，小波轉換已被應用到諸如影像壓縮、訊號壓縮、雜訊濾除等相關領域並取得相當大的成功。

小波轉換的數學式表示如下：

$$W_{\psi} f(b, a) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{b,a}(t)} dt \quad (3.5)$$

其中 $\psi(t)$ 為小波函數：

$$\psi_{b,a} = a^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad a > 0 \quad (3.6)$$

其中 a 為壓縮或拉伸的比例， b 為整段訊號的平移。

小波轉換是利用多重解析度(Multi-resolution)對訊號作分解及運算，訊號中的不連續點及轉折點的地方，經過小波轉換處理後，則會呈現明顯不同的特徵。多重解析度的觀念是先對訊號做小波一階段濾波可得 A_1 和 D_1 ，再對訊號 A_1 作小波濾波得 A_2 和 D_2 ，其餘依此類推。如圖 3-12 所示。

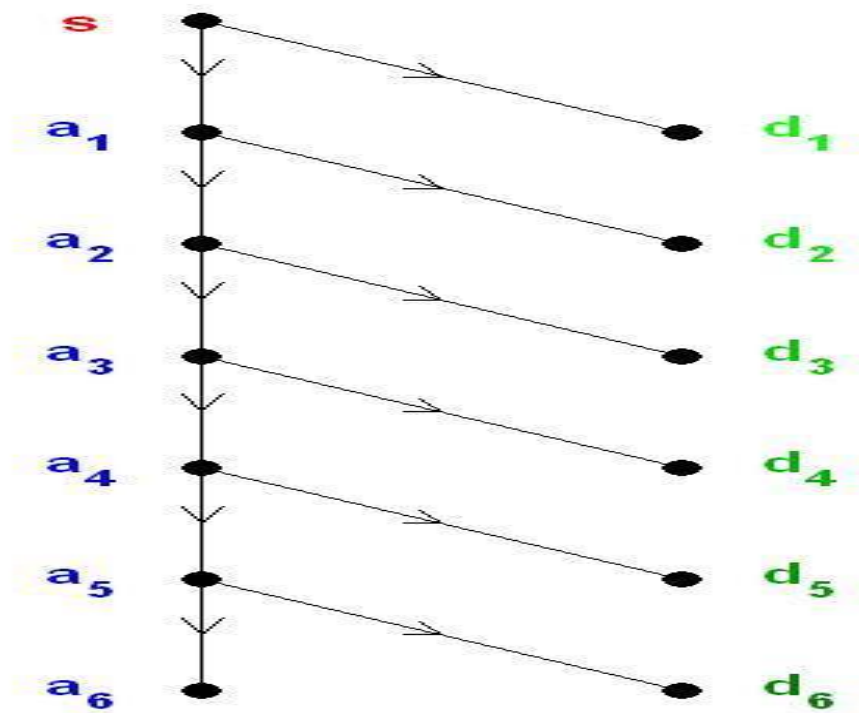


圖 3-12 小波分解示意圖

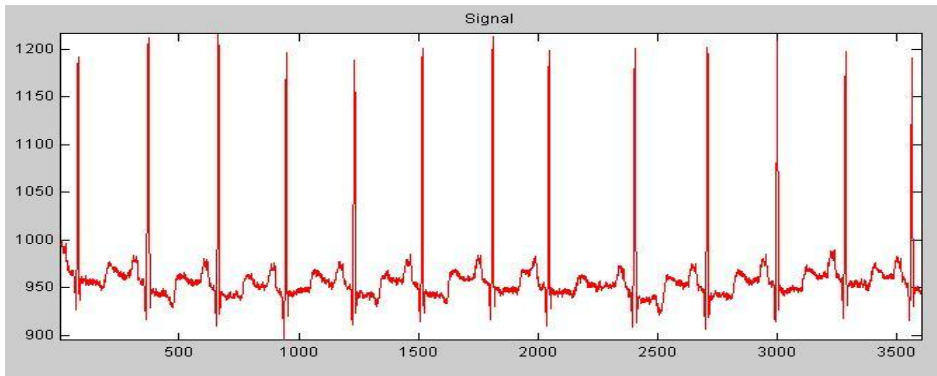
本研究採用 Daubechies 的小波轉換演算法，這演算法的特性是，它為有限支撐，同時也具有良好的時頻局部定位特性。

3.6 心電圖訊號之特徵值擷取

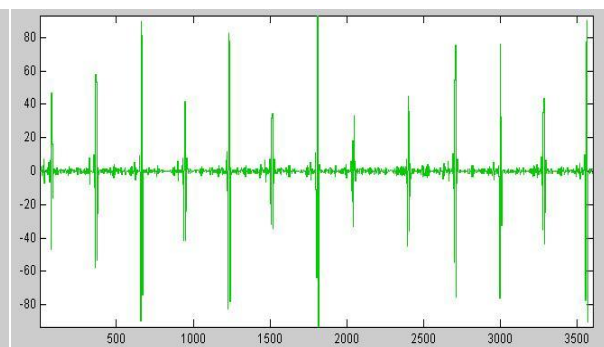
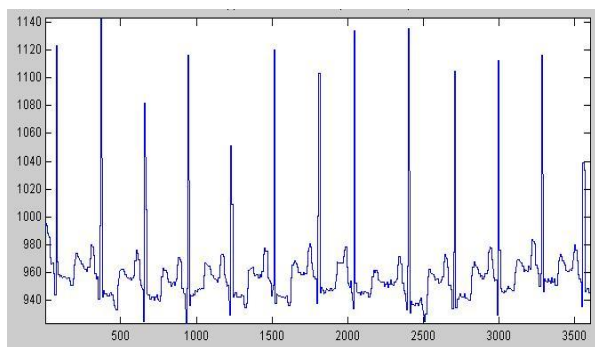
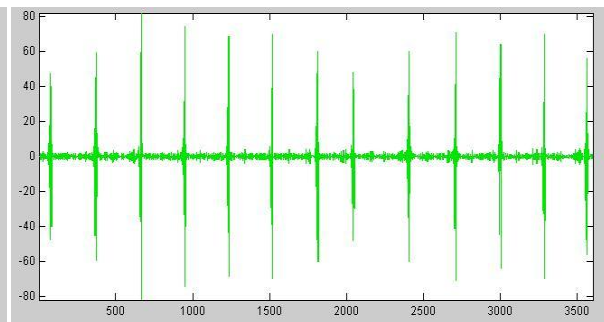
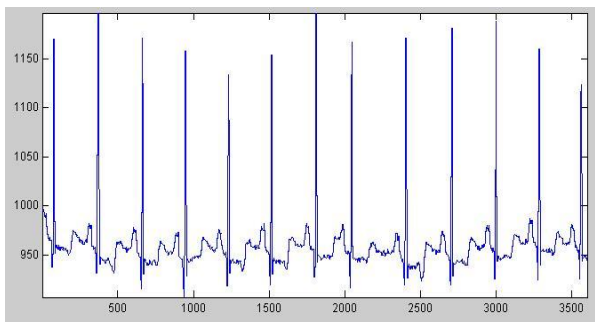
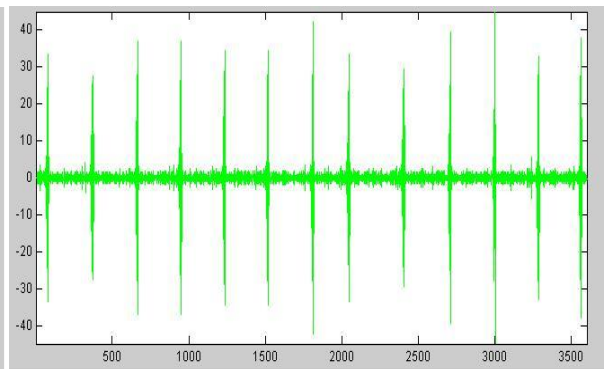
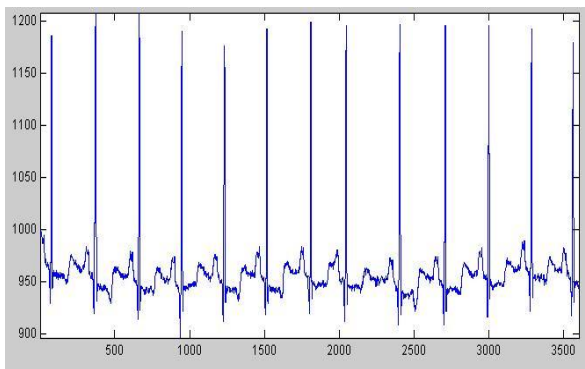
心電圖心臟疾病資料所做的特徵擷取，主要的特徵可分為頻域特徵、統計特徵及時域特徵三類。

- 頻域特徵：使用小波轉換做六階的頻帶分解，分別取出高、低頻小波係數。在心電圖分解過程中，心電訊號在不連續點或是轉折點處經過小波轉換後，能夠在多重解析度中呈現明顯且不一樣的特徵。在低頻的部份，能突顯波形中的趨勢部份，而高頻部份能突顯變化較劇烈的特徵。

本論文辨識系統中採用六階頻帶的小波轉換。六階頻帶小波轉換是利用小波的連續性多重解析度(Multi-Resolution)特性，藉由兩個數位濾波器產生兩組分解後的小波訊號，其中 $g[\cdot]$ 是高通性質的小波濾波器，第二層 $h[\cdot]$ 是低通的映射。為了降低其運算量，本論文使用的母小波採用 Lipschitz 係數， $g[\cdot]$ 的係數是 $[0, -2, 2, 0]$ 和 $h[\cdot]$ 的係數是 $[0.125, 0.375, 0.375, 0.125]$ ，而採用這些相關的小波係數是其它研究相關心電圖識別系統的專家學者也是使用此方法，並驗證此方法的有效性 [29][30]。心電圖訊號經由小波作多重分解(Multi-Level Decomposition)，分別取出低頻部份與高頻部份如圖 3- 13 (A)。而第一階分出的低頻係數相關的近似係數(Approximations)，還會包含了低頻的資訊極少部分有高頻雜訊，而高頻的細節係數(Details coefficients)，包含心電圖訊號中的高頻部份變化資訊，其中 k 代表小波轉換的尺度係數(Scales)。因此，完整六階高通係數共有 189 個係數，記為 $D[189]$ 圖 3- 13 (7)、(8)、(9)、(10)、(11)、(12)，相同地，六階低通係數共有 189 個係數，記為 $A[189]$ 圖 3- 13 (1)、(2)、(3)、(4)、(5)、(6)。



(A)	
(1)	(7)
(2)	(8)
(3)	(9)
(4)	(10)
(5)	(11)
(6)	(12)



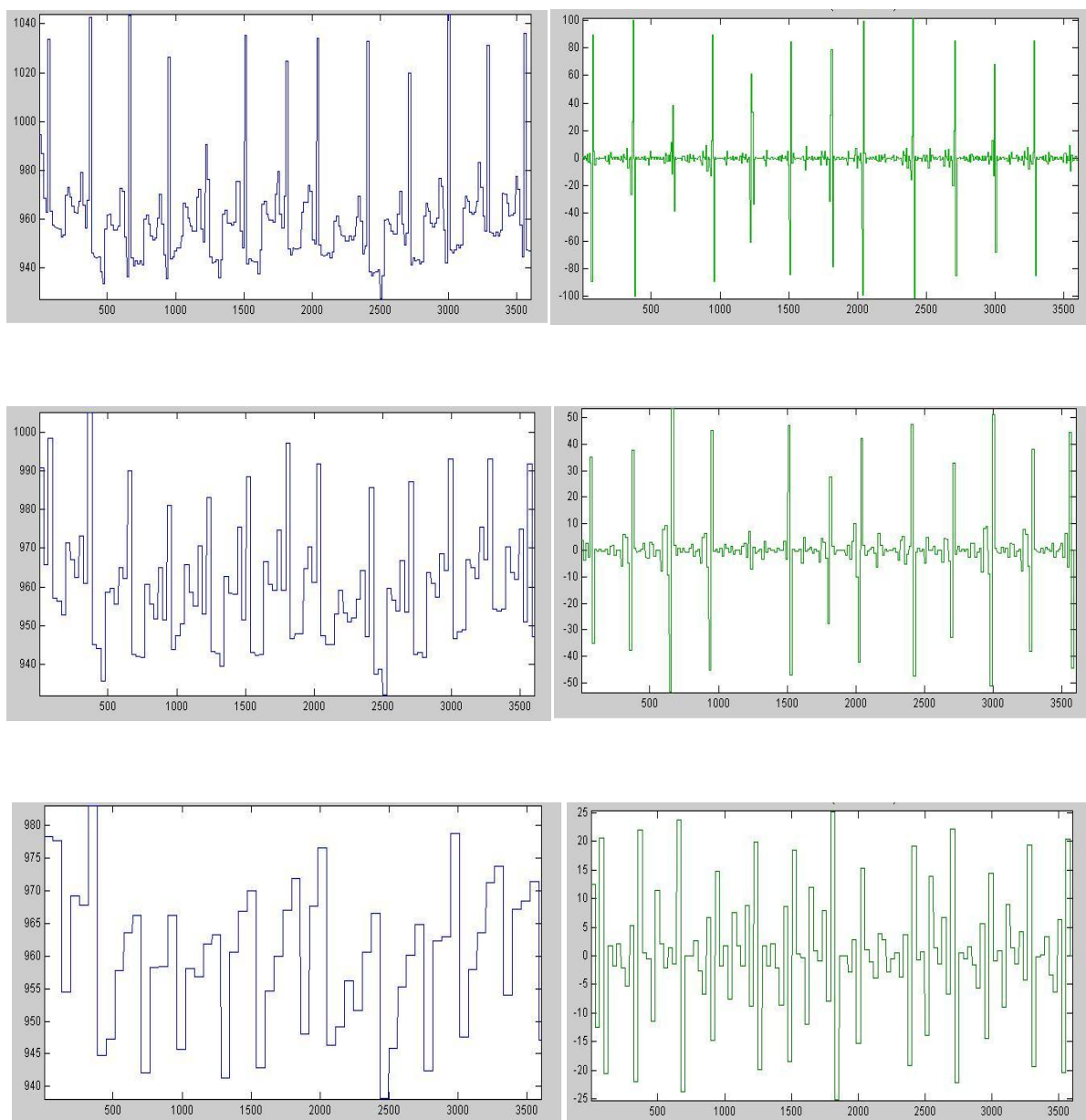


圖 3- 13 心電訊號小波分解示意圖

- 統計特徵：根據上述小波分解所得的小波係數和心電圖的波形，計算出統計的特徵，包含每一頻帶絕對值的平均值、標準差、絕對值的最大值、平均能量、及頻帶之間平均值的比值。
- 時域特徵：心電圖訊號中波形和波形之間的關係，包括心電訊號中週期性的關連性、R 波與 R 波的時間間隔、週期是否一致性、前五筆的週期平均、前後距離的比值、R 波與 S 波的振幅等，以上的特徵是傳統上判斷心臟的一些

準則(圖 3- 14)。

特徵類型	類型數目
小波係數絕對值的平均值	12
小波係數絕對值的最大值	12
小波係數平均的能量	12
小波係數的標準差	12
小波係數能量與絕對值的平均	11
R-R 間段的距離	1
R-R 間段與後五個心搏週期的比值	1
R-R 間段與後一個心搏週期的比值	1
R-R 間段與前一個心搏週期的比值	1
P 波與 R 波的振幅比值	1
R-S 波之間的落差值	1
R-R 間段與 R-S 間段的比值	1
R-R 間段與 P-R 間段的比值	1
R 波與 S 波之間距離	1
總數	68

圖 3- 14 統計特徵及時域特徵類型

綜合上述頻域特徵、統計特徵及時域特徵所擷取出的特徵值總共為 446 個特徵，其中 378 個為小波轉換後的小波係數(低頻係數 189 個+高頻係數 189 個)，統計特徵 59，另外是 9 個時域特徵為心跳的生理特徵，綜合以上所述總共所擷取特徵總數為 446(小波係數 378+統計特徵 59+時域特徵 9)。最後將所有特徵正規化，完成特徵擷取流程。

3.7 支持向量機簡介

支持向量機(Support Vector Machine ; SVM)是可以用來做為分類及迴歸的方法。是以統計理論為基礎的機器學習理論。簡單的說 SVM 是預先給予一群分類好的資料，藉由支持向量機加以訓練 (Training) 而獲得一組模型 (Model)。因此，支持向量機是一種監督式的學習方法也是屬於一般化線性分類器。

SVM 主要的觀念可分為兩類：第一主要是針對線性系統進行分析，對於非線性系統的部份，是利用通過使用非線性映射運算將低維空間線性不可分的樣本轉化為高維特徵空間使其為線性可分，從而使得高維特徵空間採用線性算法對樣本的非線性特徵進行線性分析。第二基於結構風險最小化理論上在特徵空間中，建立最佳化的分割超平面，讓機器學習機使整個系統得到最佳化，並且能使得樣本空間的期望風險以某個概率滿足一定上界。而支持向量機最終的目的是基於結構風險最小化原理，建構一個目標函數將兩類模型盡可能的區分開來。

支持向量機可經由最簡單的二元分類開始做起，再推廣到多類以上分類。先以二分法來加以討論：在低維度空間中，一群資料需以非線性函數做為分類的依據，而這非線性函數並無法輕易的求得，但可以藉由一核函數(Kernel functions)將資料映射到高維度空間，這個過程能使整個資料點分散且滿足線性函數可分類的特性。將這些資料點映射到高維度空間中，即可找出一最佳分離超平面(hyperplane)將這群資料區分出兩群(ie: A 群組、B 群組)，如圖 3- 15 所示，A 群組的資料均位於超平面的一側，B 群組的資料也均位於超平面的另一側。此超平面稱之為最佳分離超平面(Optimal separation hyperplane)，超平面的作用是盡可能的把不同類型的資料樣本點正確地分開，同時分開的資料樣本點與最佳分離面的垂直距離為最大，讓不同類型的資料樣本點能有最大的分離間隔。

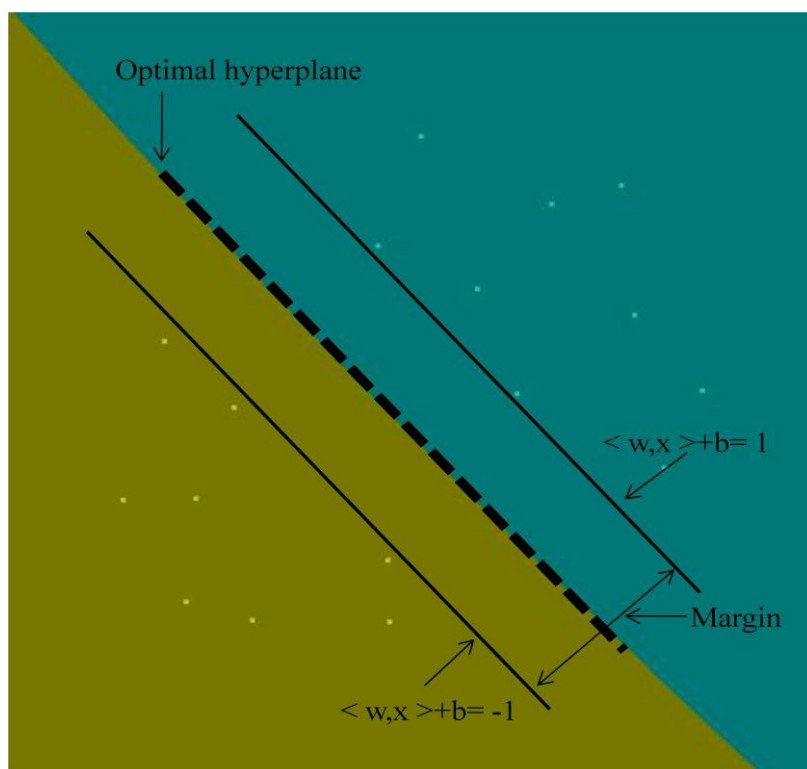


圖 3- 15 支持向量機二元分類圖

假設每個樣本點具有 n 個維度向量則：

$$\{x_i, y_i\}, i=1, \dots, n, \quad x_i \in R^d, y_i \in \{-1, +1\}$$

$\{x_i\}$ 稱為特徵向量， $\{y_i\}$ 稱為標籤(Label)

其中權重向量 w (Weight vector) 及一個偏移值 b (Bias)。支持向量的最佳分離超平面可表示成 $\vec{w} \cdot \vec{x}_i + b = 0$ ，所以有的分類資料函數向量均可表示為 $y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 (i=1, 2, \dots, n)$ 為了使分類面間距離 $(2/\|\vec{w}\|)$ 為最大，必須讓 $(\|\vec{w}\|/2)$ 為最小，則根據拉格朗日法(Lagrangian)變換導入 Lagrangian multiplier α_i 、Lagrangian

$$\text{函數： } L_d(w, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 + \sum_{i=1}^n \alpha_i [1 - y_i (\langle \vec{w}, x_i \rangle + b)], \alpha_i \geq 0, i=1, 2, \dots, n$$

及 KKT(Karush-Kuhn-Tucker)條件分別對 w 及 b 偏微分令梯度為零便可得到對偶函數 $W(\alpha)$ ：

$$\begin{aligned}
W(\alpha) &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle, (i=1, 2, \dots, n)
\end{aligned} \tag{3.7}$$

支持向量(Support Vector)是在約束條件(constraint) $\sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0$ 下，滿足所有的 $\alpha_i > 0$ 的輸入向量。

但在實際上並不是所有的問題都是線性可分離的關係。而要解決此類型的問題，則必須導入軟化間隔(Soft Margin)與鬆弛變量(Slack) ξ 來修正最佳分離超平面為 $y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i, (i=1, 2, \dots, n)$

當鬆弛變量 ξ 充分大時，原本無法以超平面劃分的樣本點將滿足上式的限制而被正確分類，但 ξ 要是被任意的放大，則支持向量機最大間隔原則將變的毫無意義。

因此原始的最佳化分離超平面問題的目標函數須導入平均鬆弛變量項 $\sum_{i=1}^n \xi_i$ ，使得上式得以修正為：

$$J(\bar{w}, \xi_i) = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n \xi_i (i=1, 2, \dots, n), \tag{3.8}$$

上式中 C 是一個權衡常數(Tradeoff Constant)是用來平衡目標函數前後項之間隔最大化與訓練誤差最小化之間的衝突。

從低維度空間映射至高維度空間必須透過核函數(Kernel functions)得以求取最佳化分離超平面的最佳化參數值，核函數有多種不同的選擇，本論文採用了徑向基核函數(Radial Basis Function, RBF)來當作 SVM 的和函數：

$$K(\bar{x}_i, \bar{x}_j) = \exp\left(-\gamma |x_i - x_j|^2\right) \tag{3.9}$$

3.8 最佳化特徵擷取

由 3.7 節中可得知支持向量機的優點是將兩個資料類別的特徵中較低維特徵映射至高維度的特徵空間，而後線性區分出兩個類別。所以在每一筆資料中特徵值的排列是相當重要的，這可影響到支持向量機分類的好壞。而本論文用採用最佳化特徵擷取的方式，把小波轉換、統計分析及生理特徵，所擷取的特徵給予重新排列。

所謂的適應性特徵選取是找出各特徵之間的差異，如果計算出來的差異值越大則表示兩類之間差別也越大，我們先定義：

$$d_{n,m}^k = (\eta_n^k - \eta_m^k)^2, \text{ for } k = 1, 2, \dots, 446 \text{ where } 1 \leq n, m \leq 47, n \neq m \quad (3.10)$$

其中 $d_{n,m}^k$ 代表為平均值相減的平方， η 代表為平均值，下標的 n 、 m 是表示所要分類的兩個類別， k 是所屬的特徵維度。而運算完 $d_{n,m}^k$ 後，我們須再求得特徵向量區別值 ε 定義為：

$$\varepsilon_{n,m}^k = \frac{d_{n,m}^k}{\left((v_n^k)^2 + (v_m^k)^2\right)}, \text{ for } k = 1, 2, \dots, 446 \text{ where } 1 \leq n, m \leq 47, n \neq m \quad (3.11)$$

上式 $\varepsilon_{n,m}^k$ 為特徵向量區別值、 v 為變異數 (variations)。此特徵向量區別值代表若同一類別之間的變異數越小，則表示兩類之間越相同。計算出的特徵向量區別值越大，則表示越可以區分出這兩類，此特徵越為重要，如下圖 3-16。

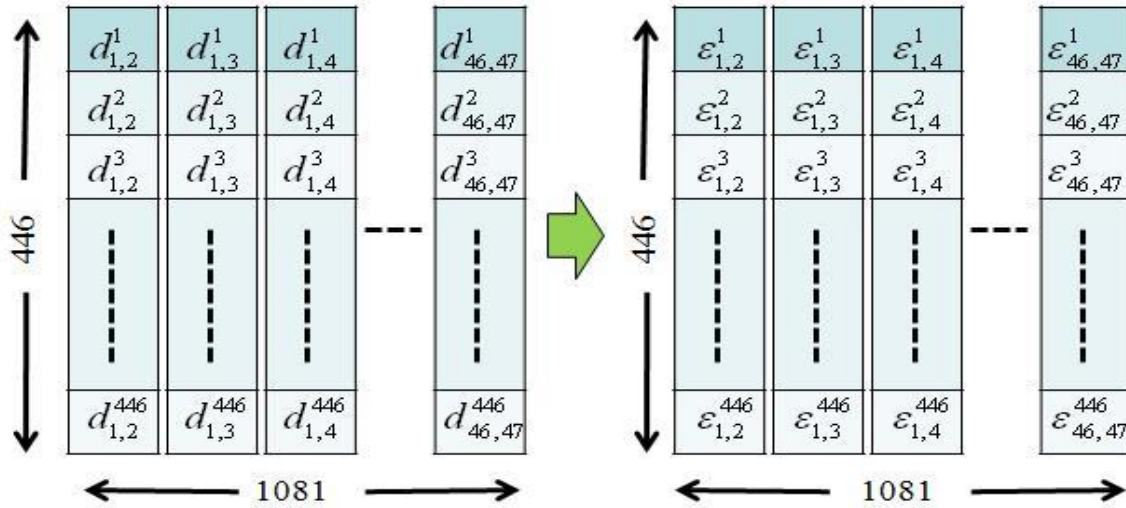


圖 3- 16 兩兩之間特徵向量區別值示意圖

本論文所研究的心律不整共有 9 類的不同病症，從表 3-1 我們得知，若心電訊號經過零相位延遲濾波器濾除基準線飄移，再經過適應性 K-means 演算法分群後，將可得到 47 個不同的次類別，因此會得到 $(47 \times 46) / 2 = 1081$ 個不同兩類之間特徵向量區別值 ε 向量，而每個 ε 向量的維度是 446。

計算完特徵向量區別值後，便可知道對於不同的兩類別，哪一個特徵向量維度的特徵值對於分辨這兩類是別比較有幫助的，區別值越大對於此分類越為有利。因此我們以特徵向量區別值 ε 來進行特徵向量的排序，排序動作如圖 3- 17，其中 P[68] 為統計特徵及時域特徵、D[189] 為小波高頻特徵及的小波低頻特徵 A[189]，經過此排序我們便可得知哪幾個特徵向量的維度比較重要，可以產生最佳辨識兩類別的特徵向量維度。如此可以使分類效果比一般所有特徵一起訓練出分類模型要來的好，並且可以減少讓使用的特徵，減低運算量，並提升系統的辨識率。

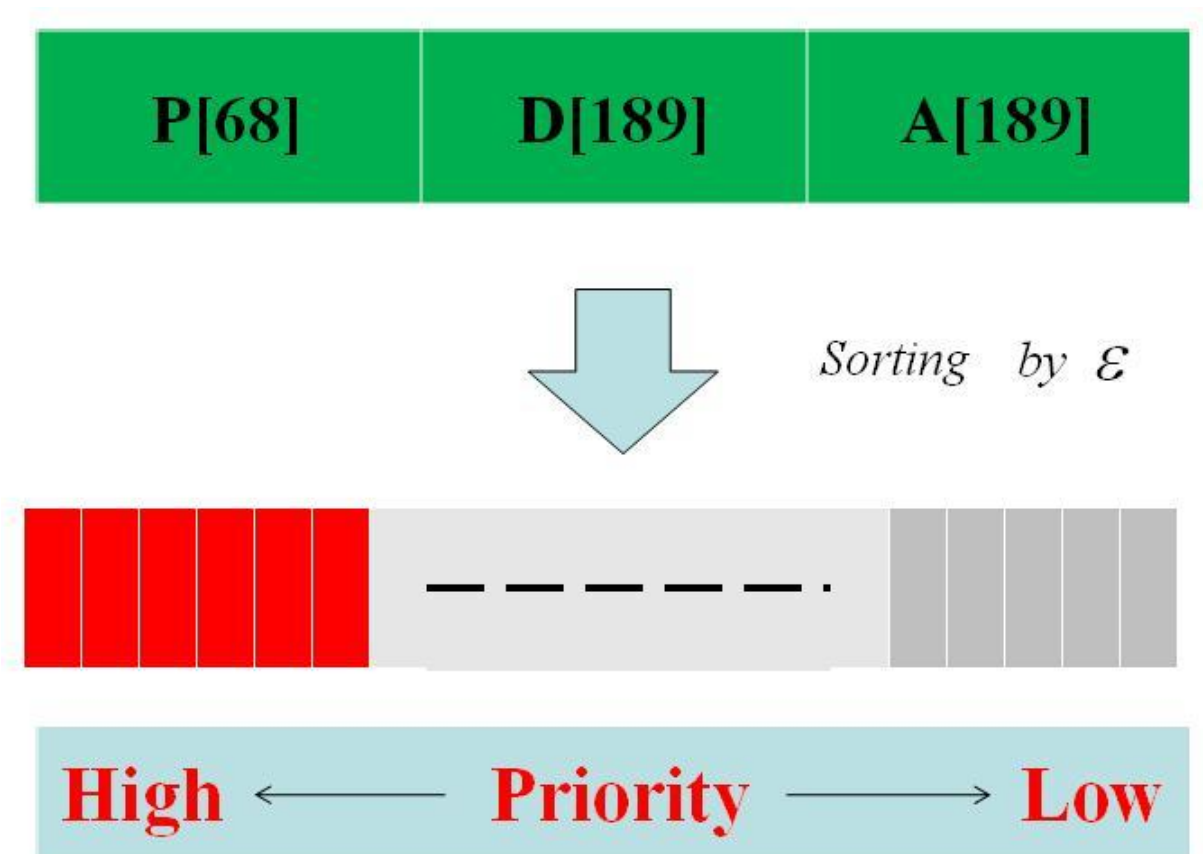


圖 3- 17 最佳化特徵向量排序示意圖

3.8 心電圖病症辨識

最後本研究將 MIT-BIH 心律不整資料庫所用的心電圖資料，把資料中的一半放入支持向量機訓練出分類器的模型，另一半則測試此模型的辨識率。一開始會將樣本的特徵向量輸入辨識系統，每一個樣本的特徵向量會依據先前最佳化特徵擷取選出的特徵做為樣本的特徵。如當輸入的樣本要分辨第一類與第二類時，就會採用第一類與第二類用的最佳分離超平面辨識，此未知的特徵向量經辨識模型的支持向量運算後，即會指定此特徵向量的類別，若此未知特徵向量被歸為第一類，則第一類得到一票，當特徵向量通過所有的 SVM 超平面計算後，即可得知此特徵向量屬於各個類別的票數，得票數最多的那一個類別，即是此未知特徵向量得所屬類別，如圖 3- 18。

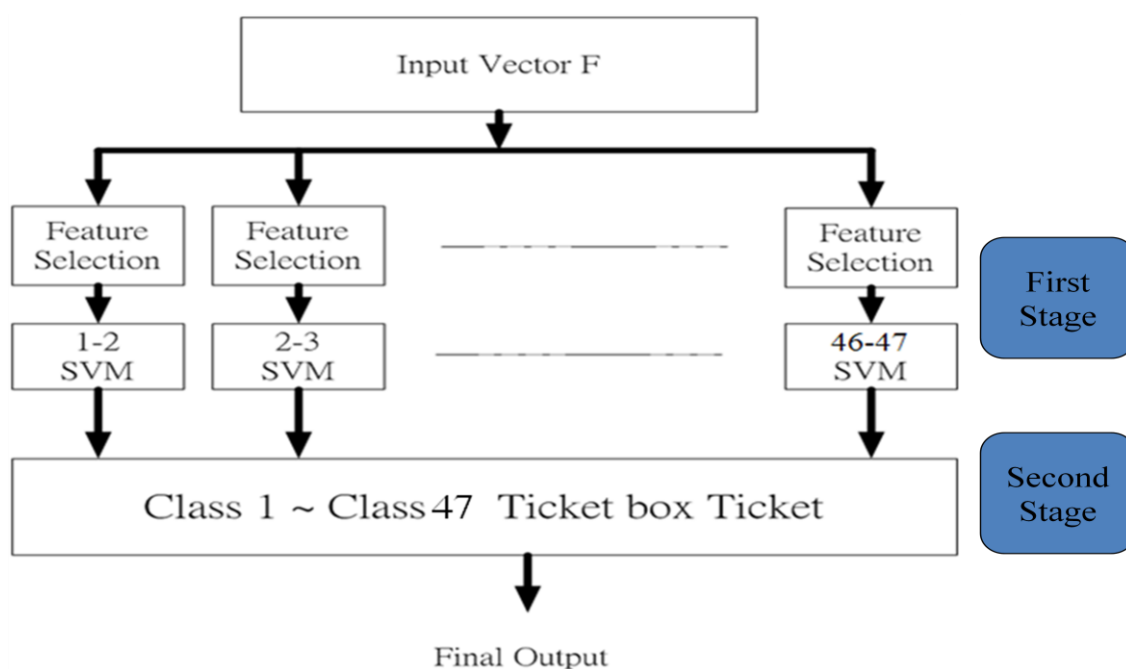


圖 3- 18 心電圖病症辨識流程圖

SVM 超平面數目由分類的類別數目所決定，以本論文的兩個實驗為例，當心電訊號濾除基準線飄移後，9 種病症將被適應性 K-means 演算法分為 47 個次類

別，則系統內共有 $C_2^{47} = 1081$ 個最佳分離超平面；當心電訊號含有基準線飄移干擾時，9 種病症將被適應性 K-means 演算法分為 70 個次類別，則系統內共有 $C_2^{70} = 2415$ 個最佳分離超平面，因此濾除心電訊號的基準線飄移干擾，將可減少 SVM 分類面的數目，對於減少辨識系統模型建立的時間，有相當大的幫助。

第四章 實驗結果

4.1 心電圖訊號辨識結果

本實驗使用的資料是 MIT-BIH 心律不整資料庫所提供的原始心電訊號，這些訊號本身帶有外在干擾雜訊，在不做前處理的情況下，套用本實驗流程，我們預期帶有雜訊的訊號使用在本研究的心電訊號辨識系統上應該會有較低的辨識率。

實驗一：

訓練樣本與測試樣本均未濾除基準線飄移干擾。

實驗二：

訓練樣本與測試樣本均以 3.2 節所介紹的零相位延遲濾波器濾除了基準線飄移干擾。

實驗三：

訓練樣本濾除了基準線飄移干擾，而測試樣本則未濾除基準線飄移干擾。本實驗的目的是為了評估基準線飄移對本系統辨識率的影響率。

實驗一所得的辨識率為 97.43%。此實驗的數據說明了就算是帶有雜訊的心電訊號使用在我們這套辨識系統上，建立出來的模型對於含有雜訊的心電訊號辨識效果仍然是非常的優良，但在 K-means 分群時，所分出的次類別高達 70 種，故在訓練特徵模型時，會比較耗時，如表 4- 1。

心律不整病症型態	心搏樣本總數	訓練樣本總數	辨識樣本總數	錯誤樣本總數	辨識率 (%)
N	72370	36227	36143	592	98.36
L	7875	3944	3931	196	95.01
R	7067	3538	3529	112	96.82
V	6459	3242	3217	105	96.73
A	2388	1197	1191	252	78.84
E	100	51	49	2	95.91
j	220	114	106	17	83.96
/	3508	1763	1745	5	99.71
F	223	111	112	6	94.64
Total	100210	50187	50023	1285	97.43

表 4- 1 第一種實驗結果

實驗二所得到的辨識率為 98. 21%，比第一種實驗的結果更高，而且在 K-means 分群時，只分出 47 種。明顯比第一種實驗的 70 種還要好，建立所有 hyper-plane 所需的時間僅為實驗二的 40%左右，如表 4- 2。這個實驗的結果說明，以零相位延遲濾波器濾除心電訊號的基準線干擾，不只可以提高辨識率，也可以降低系統建模的時間。

實驗三所得到的辨識率為 67.96%。很明顯可以得知，由於訓練模型的資料在 K-means 分群時，並沒有考慮基準線飄移的成分，因此在面對有雜訊的訊號時，無法有效的辨識，如表 4- 3 所示。

心律不整病症型態	心搏樣本總數	訓練樣本總數	辨識樣本總數	錯誤樣本總數	辨識率 (%)
N	72370	36185	36185	530	98.53
L	7875	3937	3938	46	98.83
R	7067	3533	3534	68	98.07
V	6459	2930	3529	58	98.35
A	2388	1144	1244	176	85.85
E	100	51	49	2	95.91
j	220	109	111	10	90.9
/	3508	1753	1755	7	99.60
F	223	111	112	3	97.32
Total	100210	49753	50457	900	98.20

表 4-2 第二種實驗結果

Type of Arrhythmia	Beat Cycles	Learning Beats	Testing Beats	Misrecognized Beats	Accurate Rate(%)
N	108555	36185	72370	22802	68.49
L	8812	937	7875	3358	57.35
R	10600	3533	7067	3678	47.95
V	9389	2930	6459	428	93.37
A	3532	1144	2388	1141	40.91
E	151	51	100	46	54
J	329	109	220	180	18.18
/	5261	1753	3508	49	98.6
F	334	111	223	148	33.63
Total	149963	49753	100210	32100	67.96

表 4-3 第三種實驗結果

4.2 本論文與相關文獻比較

表 4-4 是本論文與其它心電圖辨識系統的相關研究成果比較：

作者	T.H. Linh[33]	M. Engin[34]	S. Osowski[35]	C.P. Shen[24]	Our work
資料來源	MIT-BIH	MIT-BIH	MIT-BIH	MIT-BIH	MIT-BIH
擷取特徵	Hermit polynomials	WT	Higher order statistics/Hermit series	WT	WT
辨識系統	Neuro-Fuzzy network	Neuro-Fuzzy network	Support vector machine	Support vector machine	Support vector machine
資料量	7279	1200	12785	100449	100210
辨識病症數	6	4	13	12	9
辨識率	96%	98%	95.91%	95.97%	98.2%

表 4-4 本論文與相關文獻比較表

由上表可以得知目前心電圖辨識系統在辨識率都高達九成以上，且能觀察到擷取特徵方面以 Hermit function、高階統計及小波轉換為主流。本實驗所使用的資料量達十萬筆以上，雖然辨識的疾病種類比最高的 13 種少了 4 種 (Supraventricular premature beat、Aberrated atrial premature beat、Fusion of ventricular and normal beat、Nodal premature beat)，而以上 4 種病症不採用的原因是因為每一種病症資料數都相當的少，所以才不列入本實驗辨識樣本內。

與其資料量都達十萬筆以上的其它相關文獻比較，本實驗少了三種病症但其總資料數只少於 239 筆，但總體辨識率仍達 98.2%，與其他文獻相較，毫不遜色。

最後，我們要特別聲明的是，本論文是承續文獻[24][31]的工作，本論文與文獻[24][31]的主要不同之處如下：

1. 我們改良了訊號切割(Segmentation)的演算法。

2. 我們改善了時域特徵的計算方法。
3. 我們提出了適應性 K-means 演算法取代傳統的 K-means 演算法大幅降低次類別的數目，進而減少建立 SVM 分類面模型的時間。
4. 我們評估了基準線飄移干擾對此辨識系統的影響性，並提出零相位高通濾波器來濾除基準線飄移干擾。

參考文獻

- [1] 中華民國行政院衛生署國民健局 “98 年度死因完整統計表”.
- [2] D. Gorinevsky and S. Boyd, “Optimization-Based Design and Implementation of Multidimensional Zero-Phase IIR Filters,” IEE Transactions on Circuits and Systems. Vol. 53, NO.2, February 2006.
- [3] H. Elci, R. W. Longman, M. Q. Phan, J. Juang and R. Ugoletti, “Simple Learning Control Made Practical by Zero-Phase Filtering: Applications to Robotics,” IEEE Transactions on Circuits and Systems. Vol. 49, NO. 6, June 2002.
- [4] A. V. Gribok, M. J. Buller, W. Hoyt and J. Reifman, “A Real-Time Algorithm for Predicting Core Temperature in Humans,” IEEE Transactions on Information Technology In Biomedicine. Vol. 14, NO. 4, July 2010.
- [5] D. Romero, M. Ringborn, P. Laguna, S. Member, O. Pahlm and E. Pueyo, “Depolarization Changes During Acute Myocardial Ischemia by Evaluation of QRS Slopes: Standard Lead and Vectorial Approach,” IEEE Transactions on Biomedical Engineering, Vol. 58, NO. 1, January 2011.
- [6] S. Kadambe, R. Murray and G. Faye, “Wavelet Transform-Based QRS Complex Detector,” Transactions on Biomedical Engineering, Vol. 46, NO. 7, July 1999.
- [7] P. Ravier, F. Leclerc, C. Dumez and G. Lamarque, “Redefining Performance Evaluation Tools for Real-Time QRS Complex Classification Systems,” IEEE Trans , Vol. 54, NO. 9, September 2007.
- [8] M. Ferras, C. Leung, C. Barras and J. Gauvain, “Comparison of Speaker Adaptation Methods as Feature Extraction for SVM-Based Speaker Recognition,” IEEE Trans Vol. 18, NO. 6, August 2010.
- [9] C. Huai, K. Aik and H. Li, “An SVM Kernel With GMM-Super Vector Based on the Bhattacharyya Distance for Speaker Recognition,” IEEE Signal Processing Letters, Vol. 16, NO. 1, January 2009.
- [10] 呂嘉陞譯, “心電圖學必備,” 合記圖書出版社, 原著: S. Thaler “The Only EKG Book You’ll Ever Need” ,2001.
- [11] 黃天守, 陳清輝譯, “基本心電圖判讀,” 眾文圖書公司, 原著: D. Davis, How to Quickly and Accurately Master ECG Interpretation, 1997.
- [12] 邱國元, 「十二導程心電圖機研製」, 國立交通大學, 碩士論文, 民國 96 年 9 月.
- [13] 王宛, “臨床心電圖學,” 人民衛生出版社, 2005 年 4 月.
- [14] 方柏仁, “實用心電圖學,” 藝軒圖書出版社, 2008 年 5 月.
- [15] <http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>
- [16] G. B. Moody and R. G. Mark, “The Impact of the MIT-BIH Arrhythmia

- Database,” IEEE Engineering in Medicine and Biology Magazine, Vol.20, No.3, pp.45-50, May-June, 2001.
- [17] R. Mark and G. Moody, “MIT-BIH Arrhythmia Database Directory,” Cambridge, MIT, 1988.
- [18] C. W. Li, C. G. Zeng and C. F. Tai, “Detection of ECG Characteristic Points Using Wavelet Transform,” IEEE Trans. Biomedical Engineering, Vol.42, No.1, pp.21-28, January 1995.
- [19] H. A. Dinh, D. K. Kumar, N. D. Pah and P. Burton, “Wavelet for QRS Detection,” in Proc. 23rd Annual EMBS International Conference, pp.1883-1887, 2001.
- [20] 呂嘉陞,林建宇,邱顯邦,“臨床實證指引,”合記出版社,2006年6月。
- [21] J. Lee, K. Jeong, J. Yoon and M. Lee, “A Simple Real-time QRS Detection Algorithm,” in Proc. 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol.4, pp.1396-1398, Nov. 1996.
- [22] W. C. Kao, C. K. Yu, W. H. Chen, C. P. Shen and Y. W. Hung, “Automatic Electrocardiogram Recognition by Wavelet Transform and Support Vector Machine,” 2005 CACS Automatic Control Conference, Taiwan, Nov., 2005.
- [23] J. Z. Xiao and Li Xiao, “A Research of the Partition Clustering Algorithm,” International Symposium on Intelligence Information Processing and Trusted Computing, 2010.
- [24] 沈家平,「心電圖訊號分析演算法與硬體架構設計」,國立台灣師範大學,碩士論文,民96。
- [25] N. Acir, “Classification of ECG Beats by Using a Fast Last Square Support Vector Machines with a Dynamic Programming Feature Selection Algorithm,” Neural Comput, Vol.14, pp. 299–309. 2005.
- [26] M. Afrif, M. U. Akram and F. A. Afsar, “Arrhythmia Beat Classification Using Pruned Fuzzy K-nearest Neighbor Classifier Soft Computing and Pattern Recognition,” SOCPAR’09, pp.37-42, 2009.
- [27] I. Daubechies, “The Wavelet Transform, Time Frequency Localization and Signal Analysis,” IEEE Trans. Information Theory, Vol. 36, No. 5, pp.961-1005, Sep. 1990.
- [28] S. G. Mallat, “A Theory For Multiresolution Signal Decomposition: The Wavelet Representation,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 11, No. 7, pp. 674–693, July. 1989.
- [29] T. Stamkopoulos, K. Diamantaras, N. Maglaveras and M. Strintzis, “ECG analysis using nonlinear PCA neural networks for ischemia detection,” IEEE Trans. Signal Processing, vol. 46. no. 11, pp. 3058-3067, Nov. 1998.
- [30] N. Sivannarayana and D. C. Reddy, “Biorthogonal Wavelet Transforms For ECG Parameters Estimation,” Medical Engineering & Physics, Vol. 21, No. 3, pp.

167-174, April. 1999.

- [31] 楊岳穎，「以適應性特徵擷取及改進支持向量檢測心電圖心律不整」，國立台灣師範大學，碩士論文，民 99。
- [32] 徐士宜，「基於經驗模態分解法與零相位延遲濾波器之心電圖雜訊濾除法則」，國立台灣師範大學，碩士論文，民 99。
- [33] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, “Image Coding Using Wavelet Transform,” IEEE Trans. Image Processing, Vol. 1, pp. 205-220, Apr. 1992.
- [34] I. Güler and E. Übeyli, “Application of Adaptive Neuro-fuzzy Inference System for Detection of Electrocardiographic Change in Patients with Partial Epilepsy Using Feature Extraction,” Expert Systems with Application, Vol. 27, pp. 323-330, 2004.
- [35] N. Sivannarayana and D. C. Reddy, “Biorthogonal Wavelet Transforms for ECG Parameters Estimation,” Med. Eng. Phy. Vol.21, pp.167-174, 1999.