以探索性資料分析技術發展心臟血管疾病臨床輔助預知模型

蔡詩怡
國立臺北護理健康大學資訊管理所研究生
640980145@ntunhs.edu.tw

謝楠楨
國立臺北護理健康大學資訊管理所教授
nchsieh@ntunhs.edu.tw

摘要

本研究是以心臟血管疾病為範圍，針對資料探勘中的分類問題做為研究主軸，希望以探索性資料分析技術，發展出一套合理與有效的臨床輔助預知模型(Prognosis model)。所用方法為：使用子集合屬性選擇(subset attribute selection)演算法做為特徵選取，以基本模型(base model)方法與集成方法(ensemble methods)的計算結果做比較。當中，基礎模型，選擇以人工智慧類神經網路(Artificial Neural Network)、支援向量機(Support Vector Machine)、決策樹(Decision Tree)等方法為主。接著以組合異質模型的方式，來建立心臟血管疾病術後的預測模型。研究結果為：發展出之模型在準確度，或者是型Ⅰ錯誤與型Ⅱ錯誤上，相較於傳統統計與貝氏資料分析方法有顯著之效能改善。結論為：本研究期望所建構出的預知模型，用於心臟血管疾病診斷來說，具有良好自動化之預測效能。其可應用於決策支援系統，用於提供醫生作為診斷預測術後是否有併發症的參考。

關鍵詞：心臟血管疾病、探索性資料分析、類神經網路、決策樹、支援向量機。

1. 緒論

1.1 研究背景

據世界衛生組織(World Health Organization ; WHO)的報告，心臟血管疾病(Cardiovascular disease ; CVD) 是全球主要死因。每年比起其他任何疾病，更多的人死於心臟血管疾病。心臟血管疾病代表與心臟和血管不適有關。心臟血管病包括：冠狀動脈心臟病、腦血管疾病、週邊動脈阻塞疾病、風濕性心臟病、先天性心臟病。心臟病發作和中風通常是急性事件，主要為血液流入心臟或腦部時，受到阻塞。目前全球排名前 200 大的處方藥中，約有四分之一是心臟血管疾病用藥。心臟血管疾病患者年齡似乎也有下降的趨勢。

1.2 研究動機

心臟血管病變所導致的心臟衰竭是有其順序的，只要能留意身體警訊，在警訊發生時，採取適當的預防措施，是可以有效阻斷心臟衰竭。許多心臟學流行病學家認為，多數心臟病多半是可以預防的。因此，假使能使用一個科學的方法，提供決策支援預測模型，幫助醫生在病人治療前提供簡單且可信度高之風險評估。那麼，對於病人與醫生來講，又多增加醫療上之另一層面的把關了。

1.3 研究目的

本研究是以心臟血管疾病為範圍，針對資料探勘中的分類問題做為研究主軸，希望以探索性資料分析技術，發展出一套合理又有效的臨床輔助預知模型(Prediction model)，故本研究的主要研究目的如下：(1) 比較傳統統計資料分析、貝氏資料分析、探索性資料分析三者，在臨床應用上的差異。(2)發展以人工智慧為基礎之探索性資料分析(Exploratory Data Analysis)參考架構。(3)依據所發展出之探索性資料分析架構，實務應用於發展一個心臟血管疾病臨床輔助預測模型，對於醫師判斷預測術後是否有病發症的情形給予參考使用。

2. 文獻探討

2.1 在醫療資訊領域資料分析方法

一、 探索性資料分析

探索性資料分析(Exploratory Data Analysis, EDA)，是由 John Tukey 的貝爾實驗室中發展出來的。探索性資料分析的觀念底下，資料分析者應該要進行「經驗導向」的資料處理過程。因此極端值被視為和在非經驗模型(機率模型)下會產生之數值同等重要。探索性資料分析其精神為，不忽略資料中任何的線索。探索性資料分析在臨床上的應用為：預測手術、用藥、診斷或是流程控制的效率、生物晶片分析、預防醫學分析、院內感染分析、臨床病徵分析、基因圖譜比對、基因定序、演化分析。傳統資料分析在臨床上的應用為：存活分析、實驗設計的可行性、藥品臨床試驗的有效性，臨床試驗受試者的估算、資料收集的方法、資料庫的建檔及品質管制。貝氏之資料分析在臨床上的應用為：醫療影像處理、臨床試驗。探索性資料分析、傳統資料分析、機率之資料分析，三者雖都為資料分析方法，但是其在分析步驟、使用順序上，是有些微的不同的。如表一所示，其描述探索性資料分析、傳統資料分析、貝氏之資料分析，三者分析步驟之比較。當中，在表一內提到了一些名詞，在此定義與解釋：1. 問題：發現問題，因而提出問題。2. 目標：了解所想達成的研究目標。3. 收集資料：收集和檢核資料。4. 分析：整理與分析所蒐集到的資料，將資料處理分析所得到的結果，客觀、嚴謹、忠實地呈現出來。5. 建構模型。6. 結論：根據研究問題作價值判斷與推論。

表 1、三種資料分析步驟之比較[13]

資料分析方法	分析步驟
探索性資料分析	問題 → 目標 → 收集資料 → 分析 → 建模 → 結論
傳統資料分析	問題 → 目標 → 收集資料 → 建模 → 分析 → 結論
貝氏之資料分析	問題 → 收集資料 → 建模 → 先驗分佈 → 分析 → 結論

2.2 集成方法

Stacking 法

Stacking 是一種可以組合多種模型的方法，但是它不像 Boosting，Stacking 的使用可以混合多種學習演算法使用，如：Stacking 的分類可以結合 Naïve Bayes、decision tree、和 rule-based 分類。而每個分類都有自己的預分類評估機率。最後的分類是使用 meta 分類，像是多線性迴歸，這就是以概率的機礎學習模型。Stacking 是在 1992 被 Wolpert 學者提出[5]。而 Seewald 學者[2]製造了一個增強版的 Stacking，稱它為 StackingC，它改善了 Stacking 在多重分類預測問題的效能。Stacking 通常的效能比單一學習模型的效能相對來的佳[5,14,16]。而每個分類都有自己的預分類評估機率，Stacking 在總體上而言是相對較佳的方法[6]。Stacking 是一個提供學習子集評估和校正的方法，和單一模型使用，其被發現可以減少很多一般問題的泛化錯誤率。Stacking 也被發現，結合不同種類的學習演算法，在預測問題的準確度上，可以達到更好[5]。

2.3 機器學習預測性模型選擇

一、類神經網路

類神經網路(Artificial Neural Network)，其包括不同的層次，其彼此相互連結。隱藏層介於輸入層和輸出層之間。每一層節點由參數與權重連結，可以藉由此調整錯誤與偏差。而改變連結的權重的過程，被稱做訓練[15]。

二、支持向量機

支持向量機(Support Vector Machine)，其輸入樣本進行分類，只有在線性情況下。因此，在當輸入空間是非線性的情況下，該技術需要從非線性的輸入空間轉換到高維特徵空間時，其結構風險最小化，且提供更好的泛化能力、神經網絡使用經驗風險最小化。其研究領域包括分類的虹膜數據、數字數據、甲狀腺數據、血細胞的數據、平假名數據、數據挖掘、

客戶欺詐檢測、圖像分類用等[4]。使用支援向量機，可以將資料分為兩組，而無需過度擬合。支援向量機可以與大量資料集配合使用，可以知道哪些是含有大量預測變數欄位的資料集。

三、決策樹

決策樹(Decision Tree)，是一個快速又有效的方法，對於分類資料集來說，可以提供良好的決策支援能力[19]。在決策樹中的結點，其功能包括測試一個特定的屬性，通常此屬性為一個常數值。然而，樹的兩個屬性相互比較。葉節點給定適用於所有情況下，即可到達葉的分類。對一個未知的例子做分類時，它是沿著樹根，根據屬性的測試值，來做連續的結點[10]。

2.4 機械學習變數選擇技術

一、關聯性特徵選擇

關聯性特徵選擇 Correlation-based feature selection(CFS)，為自動化屬性選擇的屬性子集(attribute subset)評估方法，是用來評估有用的屬性子集、計算它的冗餘、和多個屬性之間的相互作用[8]。其主要為給定屬性優異的計算，考慮到相關的屬性與目標類別，以及資料集裡屬性與其他屬性的關係。屬性和目標類別有較強的相關性，和其他屬性排序較高的，則具有較弱的相關性。優點：其為快速和獨立的目標學習方法。缺點：其為不計算屬性間潛在的交互作用[11]。

二、Consistency

Consistency-based，其可以識別屬性子集的值，去做資料子集的劃分，包含大部份強大的單一分類。優點：其為獨立的目標學習方法，可以計算冗餘和屬性間的相互關係。缺點：相關的特徵選擇較低[9]。

三、Wrapper

Wrapper，其可以使用目標學習演算法去評估屬性子集的值。使用搜尋演算去測試許多組合的屬性，並且找到最佳解決方案。優點：可以計算冗餘和屬性間的相互關係。一般來講，其結果會比其它技術來的好，因為其第二解決方案評估是使用目標學習演算法。缺點：特定的學習演算法被使用在評估子集合的值(已經被重新運行每個學習法)。比其它方法慢，不能應用在高維度的資料庫和較慢的學習演算法[17]。

2.5 分類器效能可靠評估方法

一、以 ROC 曲線評估

ROC (Receiver Operative Characteristic curve)曲線下面積較大的是指該分類器的性能優於其他[3,18]。

二、十摺交叉驗證法

十摺交叉驗法 (10-fold cross-validation)，是常用的精度測試方法。將數據集分成十份，輪流將其中 9 份做訓練 1 份做測試，10 次的結果的均值作為對算法精度的估計，一般還需要進行多次 10 倍交叉驗證求均值，但如以 10 次 10 倍交叉驗證，其效果更精確一點[7]。使用此方法，在每一次個案內，會有數量較大的資料物件被訓練，推測可增加在分類上的準確率，且隨機抽樣的問題及出現重複訓練相同的資料物件多次的狀況，可較為避免，也因為訓練次數隨著每一次資料集合的物件數而改變，所以當資料物件數增加時，相對的資料分割訓練次數也會因此而增加，如此間接的會提高計算上的成本。且當每一資料集內有 n 筆資料物件時，會使整個學習程序須執行 n 次，時間成本也因此相對的提高了。因此對於用於大量的資料庫較不適合。此方法較適合使用在資料數較小的資料集合[1]。

三、混淆矩陣

混淆矩陣(confusion matrix)，其常被用來評估人工智能領域之模型評估，如表二，每一列在例子中代表其真實的分類。混淆矩陣其優點在於它可以很清楚的看出被受分類混淆的模型。其評估的準則包括：平均準確度、型Ⅰ錯誤(Type-Ⅰerror)、型Ⅱ錯誤(Type-Ⅱ error)。型Ⅰ錯誤和型Ⅱ

錯誤不適用在所有模型，型Ⅰ錯誤被稱為”假陽性”。型Ⅱ錯誤，被稱為”假陰性”。在大多數情況下，型Ⅱ錯誤不像型Ⅰ錯誤，可以用來預測術後併發症。對於預測手術後併發症來講，判斷認為好的模型必需符合良好的平均準確度、降低型Ⅰ錯誤，且型Ⅰ錯誤和型Ⅱ錯誤需一致最好[12]。

表二、混淆矩陣用於評估模型

		實際情況 (Actual Condition)	
		目前有併發症 (Complication present)	目前無併發症 (Complication sabsent)
測試結果 (Test result)	陽性結果 (Positive result)	真陽性 (True Positive；TP)	假陽性 (False Positive；FP)
	陰性結果 (Negative result)	假陰性 (False Negative；FN)	真陰性 (True Negative；TN)

平均準確度、型Ⅰ錯誤、型Ⅱ錯誤公式如下：

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

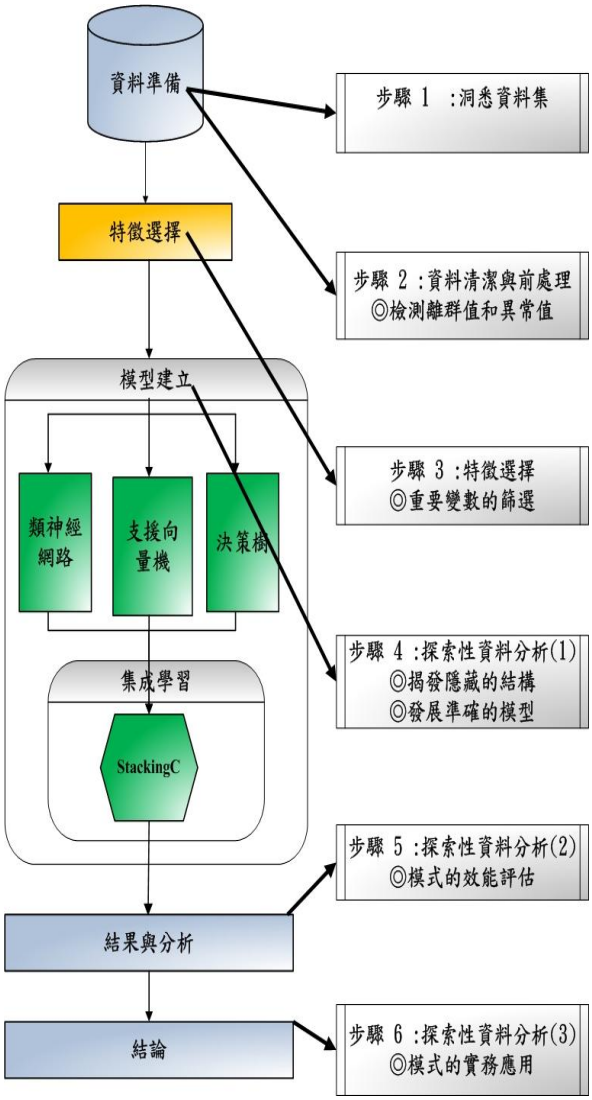
$$\text{Type-I error} = \frac{FN}{TP+FN} \quad (2)$$

$$\text{Type-II error} = \frac{FP}{FP+TN} \quad (3)$$

3.研究方法

3.1 探索性資料分析方法論

探索性資料分析方法論為資料探勘的一種流程。如圖一所示，本研究資料探勘中之探索性資料分析順序為：搜集到的資料、而後做資料清潔與前處理、再來給予特徵技術選擇(本研究使用 CFS、Consistency、Wrapper 之特徵選取)、接著使用探索性資料分析、接著就可以依據資料推論到知識(規則的萃取)，其可以多維數據顯示(ex：OLAP)、電子表格、數據可視化、報表等…產生。



圖一、探索性資料分析方法之步驟

3.2 探索性資料分析方法之步驟描述

此節依據探索性資料分析方法，在其執行的步驟內，給予描述，如下：

步驟 1：洞悉資料集。

本研究所蒐集之資料由台北市某醫學中心所提供，自 2007 年 10 月至 2009 年 9 月之心臟血管疾病資料庫，研究之範圍為心血管疾病患者之個案資料，而這些資料中之個案案例共有 122 筆、有 156 個欄位，這 156 個欄位 當中包含了 13 種併發症，如表三所示。在個案案例 122 筆當中，刪除一筆資料不完整的案例資料，剩餘 121 筆。而在 156 個屬性欄位中，刪除資料 116 個屬性欄位後，剩餘 40 個屬性欄位。

表三、個案併發症清單

中文名稱	英文名稱
出血	Bleeding
心房顫動	Atrial Fibrillation
心室早期收縮或心室心博過速	VPCs or VT
永久性心律器置放術	Permanent Pacemaker (placement)
肺炎	Pneumonia
呼吸衰竭	Resp Fail
急性呼吸衰竭	Acute Respiratory Failure
肝衰竭	Hepatic Failure
腦中風	Cerebral Vascular Accident
缺血性下肢	Ischemic leg
缺血性腸道疾病	Ischemic Bowel Disease
敗血症	Sepsis
其他併發症	Others

步驟 2：資料清潔與前處理

◎檢測離群值和異常值：

在此主要目的是為了確認資料的正確性以及完整性，使得資料探勘能夠順利進行。因此，利用人工處理與資料探勘分類(classification)的分式，分為兩階段的方式，來檢測離群值和異常值。藉由篩選、清理、增加、改變資料，對屬性做資料調整。且對於不完整的資料，與單一變數相同值過多，則皆直接刪除該欄位，原由為，本研究蒐集的資料為醫學臨床資料庫，如自行填入缺失值、或單一變數相同值過多屬性欄位留下，將導致資料失真。而使用分類的方法主要是想把已分類的資料來研究罹患心臟血管疾病的患者資料特徵屬性，日後，可再根據這些特徵對其他未經分類或是新的資料做預測。

步驟 3：特徵選擇

◎重要變數的篩選：

屬性選擇為一個識別的流程，其可從資料集中，去移除單一變數相同值過多和冗餘的資訊。本研究利用資料探勘軟體工具 Weka，在資料前處理中，將資料使用(CFS、Consistency、Wrapper 之特徵選取)的方式，來分別進行重要的變數的挑選。當中，在 CFS 法內，將最佳屬性搜尋方式設定為 Best First。此三種特徵選取法，挑選出特徵值如表四所示：

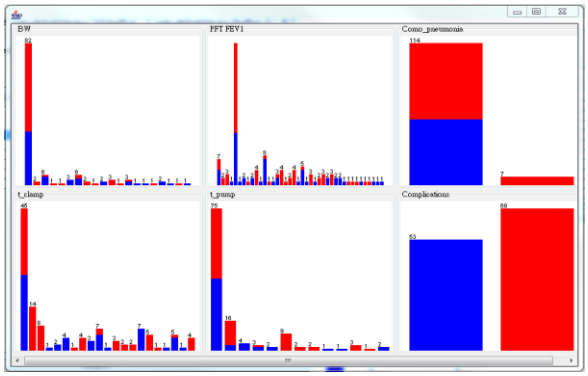
表四、使用屬性子集合評估演算法之結果

	屬性	CFS	Consistency	Wrapper
01.	BW			x
02.	PFT_FEV1			x
03.	Como_pneumonia		x	x
04.	t_clamp			
05.	t_pump		x	x
06.	PH COPD	x		x
07.	Cre	x		x
08.	PH PAOD	x	x	
09.	GOT	x	x	
10.	echo_clamp	x	x	x
11.	echo_IVS	x	x	
12.	Complications			

步驟 4：探索性資料分析(1)

◎揭發隱藏的結構：

利用 Weka 軟體之 Explorer 選單，開啟資料，其為資料可視化，可因此了解資料的分佈狀態、資料內所蘊含的資訊，進而去發現資料內部隱藏的結構(關係)、也可藉此檢測資料的品質程度。圖二為資料預處理後，特徵屬性之分佈狀態。



圖二、視覺化探視資料分佈狀態

◎發展準確的模型：

在此步驟內，依據資料的特性，而去發展適合的演算法模型，其應用資料探勘演算法進行模型的建構。分別選定類神經網路、支援向量機、決策樹的單獨模型，與 stackingC 方法。選擇以類神經網路之原由為：其準確率為目前公認所接受的。選擇以支援向量機之原由為：其可以解決非

線性的問題。選擇決策樹的原由為：其可以支援不穩定的結構。

步驟 5：探索性資料分析(2)

◎模型的效能評估：

在此步驟內，使用 Weka 軟體的 Weka Explorer 介面。在選定之模型當中，類神經網路、支援向量機、決策樹，在 weka 軟體內使用的演算法分別為：MultilayerPerceptron、SMO、J48。而後把所選定的模型，給予軟體運算，計算出平均準確度、混淆矩陣。再把混淆矩陣套入公式(2)與(3)，得出型Ⅰ錯誤、型Ⅱ錯誤之數值。最後，使用平均準確度、型Ⅰ錯誤、型Ⅱ錯誤的值來評定模型效能，與做為選擇模型之參考使用。如表五所示：其列出了類神經網路、支援向量機、決策樹、StackingC 四種機器學習預測模型分別在特徵選取(CFS、Consistency、Wrapper)三者上之評估(準確度、型Ⅰ錯誤、型Ⅱ錯誤)結果。準確度愈高，則愈好。而型Ⅰ錯誤、型Ⅱ錯誤判別高低的好壞主要取決於想預測之疾病類型與在治療檢驗上，是否有極大的風險，來調整取決的高低。結果發現，使用 Consistency 屬性選擇法與 SVM 預測模型，其準確度(79.3388%)、型Ⅰ錯誤(0.2083%)、型Ⅱ錯誤(0.2055%)上，相對於其它屬性屬擇與預測模型組合上，來得佳。

表五、屬性子集合與模型效能評估結果：

屬性子集合 評估 / 模型	CFS		Consistency		Wrapper	
ANN	準確度(%):	80.9917	準確度(%):	84.2975	準確度(%):	83.4711
	型Ⅰ錯誤(%):	0.2321	型Ⅰ錯誤(%):	0.1600	型Ⅰ錯誤(%):	0.2462
	型Ⅱ錯誤(%):	0.1538	型Ⅱ錯誤(%):	0.1549	型Ⅱ錯誤(%):	0.0714
SVM	準確度(%):	81.8182	準確度(%):	79.3388	準確度(%):	78.5124
	型Ⅰ錯誤(%):	0.2459	型Ⅰ錯誤(%):	0.2083	型Ⅰ錯誤(%):	0.2923
	型Ⅱ錯誤(%):	0.1167	型Ⅱ錯誤(%):	0.2055	型Ⅱ錯誤(%):	0.1250
DT	準確度(%):	73.5537	準確度(%):	73.5537	準確度(%):	82.6446
	型Ⅰ錯誤(%):	0.3433	型Ⅰ錯誤(%):	0.3433	型Ⅰ錯誤(%):	0.2500
	型Ⅱ錯誤(%):	0.1667	型Ⅱ錯誤(%):	0.1667	型Ⅱ錯誤(%):	0.0877
StackingC	準確度(%):	81.8182	準確度(%):	85.1240	準確度(%):	82.6446
	型Ⅰ錯誤(%):	0.2373	型Ⅰ錯誤(%):	0.1930	型Ⅰ錯誤(%):	0.2500
	型Ⅱ錯誤(%):	0.1290	型Ⅱ錯誤(%):	0.1094	型Ⅱ錯誤(%):	0.0877

步驟 6：探索性資料分析(3)

◎模型的實務應用：

將高效能高正確率的模型，實務應用在心臟血管疾病醫學資料預測中。

3.3 預測模型建立

建立模型的過程中，選擇以類神經網路、支援向量機、決策樹、及 StackingC 做為其模型的選定。圖一為研究架構圖。整體上主要是在於描述本研究搜集到的資料，如何在整個探索性資料分析方法論之六個步驟中實現。圖一，左側描述如下：一開始取得資料來源後，先做資料前處理，接著，在特徵選擇的部份，找到重要的變數，而後建立模型，使用類神經網路、支援向量機、StackingC 等方法。接著再來做結果與分析的部份。最後，得到研究結論。

3.4 資料探勘工具

本研究所使用的資料探勘工具為 Weka(Waikato Environment for Knowledge Analysis, WEKA)，它是由紐西蘭的懷卡托(Waikato)大學開發出來的，此系統是以 JAVA 撰寫的，它幾乎可以使用在任何平台，如 Linux、Windows、Macintosh 作業系統，甚至可以用在個人數位助理(Personal Digital Assistant，PDA)上使用。其提供了很多不同種類的學習演算的方法、廣泛的提供資料探勘實驗完整的流程，包括輸入數據、評估模型、視覺化的輸出資料。除了各式各樣的學習演算法，它還提供廣泛的預處理工具。其多樣化的功能，可供使用者可以比較不同的演算法，對於想解決的問題選擇最適當的演算法[10]。

4. 結論

此研究方法，不論是在準確度，或者是型Ⅰ錯誤與型Ⅱ錯誤上，皆大幅度的提升模型效能。研究所建構出的預知模型，用於心臟血管疾病診斷，有良好自動化之預測效能。並應用於決策支援系統，對於醫師判斷預測術後是否有併發症的情形給予參考使用。

參考文獻

[1] 韓歆儀 *應用兩階段分類法提升 SVM 法之分類準確率*. 2004.

[2] A. K. Seewald, "How to Make Stacking Better and Faster While Also Taking Care of an Unknown Weakness," *Proc. of 19th International Conference on Machine Learning(ICML)*, University of New South WALES, Sydney, Australia, July 2002, pp. 554-561.

[3] A. Osareh, and M. Mirmehdi, and B. Thomas, and R. Markham , "Comparative Exudate Classification Using Support Vector Machines and Neural Networks," *Medical Image Computing and Computer-Assisted Intervention*, Vol. 2489, 2002, pp. 413-420.

[4] C. Emre, and A. Ahmet, "A Biomedical Decision Support System Using LS-SVM Classifier with an Efficient and New Parameter Regularization Procedure for Diagnosis of Heart Valve Diseases," *journal of Medical Systems*, June 2010, pp. 1-8.

[5] D. H. Wolpert, "Stacked generalization, " *Neural Networks*, Vol. 5, 1992, pp. 241-259.

[6] D. Zhu, "A hybrid approach for efficient ensembles, " *Decision Support Systems*, Vol. 48, February 2010, pp. 480-487.

[7] E. Kretschmann, and R. Apweiler, "Automatic Rule Generation for Protein Annotation with the C4.5 Data-Mining Algorithm Applied on Peptides in Ensembl, " *Conference on Bioinformatics, German* , October 2001, pp. 53-57.

[8] G. D. Fiol, and P.J Haug, "Classification models for the prediction of clinicians' information needs, " *Journal of Biomedical Informatics*, Vol. 42, February 2009, pp. 82-89.

[9] H. Liu, and R. Setiono, "A Probabilistic Approach to Feature Selection - A Filter Solution," *Proc. of 13th International Conference on Machine Learning(ICML)*, Bari, Italy, July 1996, pp. 319-327.

[10] I. H. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, USA, 2005, pp. 365-368.

[11] M. A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning," *Proc. of 17th International Conference on Machine Learning(ICML)*, California,USA, June 2000,pp. 359-366.

[12] N. C. Hsieh, and L.P Hung, and C.C Shih, and H.C Keh, and C.H Chan, "Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques, " *Journal of Medical Systems*, December 2010, pp. 1-12.

[13] NIST. *Exploratory Data Analysis*. SEMATECH e-Handbook of Statistical Methods 2010; Available from: http://www.itl.nist.gov/div898.

[14] P. K. Chan, and S.J Stolfo, "On the Accuracy of Meta-learning for Scalable Data Mining, " *Journal of Intelligent Information Systems*, Vol. 8, February 1997, pp. 5-28.

[15] Purwanto, and C. Eswaran, and R. Logeswaran, and A. Rahman, "Prediction Models for Early Risk Detection of Cardiovascular Event, " *Journal of Medical Systems*, May 2010, pp. 1-11.

[16] R. Caruana, and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," *Proc. of 23rd International Conference on Machine Learning(ICML)*, Pittsburgh, USA, June 2006, pp. 161-168.

[17] R. Kohavi, and G.H John, "Wrappers for feature subset selection, " *Artificial Intelligence*, Vol. 97, December 1997, pp. 273-324

[18] R. M. Centor, "Signal Detectability:The Use of ROC Curves and Their Analysis, " *Med. Decis. Making*, Vol. 11, June 1991, pp. 102-106.

[19] S. Lee, "Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C

controls," *Decision Support Systems*, Vol. 49, Novembwe 2010, pp. 486-497.