

運用機器學習於加護病房肝硬化重症病人之死亡預測

馬瑞菊¹ 曾葭茵² 李佳欣¹ 鄭婉如¹ 蕭嘉瑩³ 蘇珉一³ 鮑永誠^{4,5}

台東馬偕紀念醫院 ¹護理部 ²醫學教育研究部 ³內科部 ;
⁴高雄醫學大學醫務管理暨醫療資訊系所非線性分析及最優化理論研究中心 ;
⁵高雄醫學大學附設醫院臨床醫學研究部

摘 要

機器學習已被應用於多個領域且在醫療預測方面亦已取得成效，目前臨床上尚無針對肝硬化病人在重症環境中的評估量表，因此本研究期望能利用機器學習方法建構肝硬化重症病人之死亡預測模型。研究採電子病歷回溯性調查，以東部某區域醫院 2013 年 8 月 1 日至 2015 年 12 月 31 日入住加護病房之肝硬化 287 位個案進行分析，透過 R 語言使用五種學習方法：Bagging、GLM、AdaboostNeyman-Pearson、Stacking 分別建構五個模型，並且和臨床傳統常用的肝病病人嚴重度評估量表 CTP score、MELD score 以及常用於重症加護病房疾病的評估量表 APACHE II 進行比較。本研究建構的五個機器學習模型準確率達八成以上，其中 Neyman-Pearson 預測效果最佳準確率 86.2% (AUC=0.871)，且皆優於以傳統臨床評估量表 APACHE II (AUC=0.806)、MELD score (AUC=0.755)、CTP score (AUC=0.747) 建構的死亡預測模型，其預測結果期望能夠提供醫療人員決策參考。

關鍵詞：機器學習 (Machine learning)
肝硬化 (Cirrhosis)
加護病房 (Intensive care unit)
死亡預測 (Predictors of death)

前 言

臺灣慢性肝病及肝硬化在 2019 年十大死因排行第 10 名¹。肝硬化是一種常見的慢性肝病，會產生多種併發症以及存在高死亡率²。病理生理特徵包括進行性肝損傷、纖維化、門脈高壓和失償性肝硬化、腹水、自發性細菌性腹膜炎、肝性腦病變、靜脈曲張出血、肝腎症候群和肝細胞癌等³。

肝硬化晚期的重症病人經常因併發症而入住加護病房⁴，在美國每年約有 26,000 例肝硬化病人需要重症照護且院內死亡率大於 50%⁵，因此即時識別嚴重程度較高的病人並介入治療防止進一步惡化就顯得十分重要⁶。由於目前對於末期肝病 (End stage of liver disease, ESLD) 病人有效的根本治療方法唯有器官移植，但肝臟移植供不應求⁷，據 2019 年肝臟捐贈數量僅有 118 例⁸，等待期對病人而言是個艱難的時期，充滿

不確定性還伴隨著高發病率及死亡率⁹。

末期肝病目前有兩個臨床上常用的評估量表(Scoring system)，Child氏分級(Child-Turcotte-Pugh score, CTP score)，由總膽色素、白蛋白、凝血功能、腹水以及肝性腦病變組成，得分越高越嚴重。末期肝臟疾病評分系統(Model for End Stage Liver Disease score, MELD score)為歐美國家肝臟移植經常採用的指標，計算公式包含總膽色素、凝血功能和肌酸酐，分數越高表示死亡率越高¹⁰⁻¹¹。在入住加護病房(Intensive Care Unit, ICU)順序以及嚴重程度判斷嚴重程度評估方面，多採用急性生理性健康評估II(Acute Physiology And Chronic Health Evaluation II, APACHE II)，計分項目為入院24小時內初次12項常規生理測量值、年齡和慢性疾病狀態，分數越高其嚴重程度越高¹²。目前重症肝硬化病人死亡預測尚未發展出特定適用指標¹³，故臨床預測程度有限。

機器學習是從資料的中學習並改進規則達到人工智慧的方法¹⁴，近年來逐漸被應用於醫學診斷和預測，國內外已有研究使用機器學習模型進行預測並指出其結果優於傳統評估量表，雖然利用機器學習方法用於預測肝硬化病人的死亡模型已被提出且具備可用性¹³，但與肝硬化重症傳統評估量表(CTP score、MELD score、APACHE II)預測能力之比較尚未被詳細探討。

綜上所述，本研究目的為利用機器學習方法與傳統評估量表對肝硬化重症病人死亡預測模型的比較與探討，期望模型結果能作為醫療決策輔助工具。

研究材料與方法

一、研究對象及場所

研究計畫經由醫學中心人體試驗委員會同意後執行(編號：18MMHIS0006)，使用來自某東部某區域教學醫院做研究地點(約500多床，員工約700多名)，以電子病歷進行回溯性分析，資料收集期間及對象為2013年8月1日至2015年12月31日入住內科加護病房所有之肝硬化個案(N=287)。

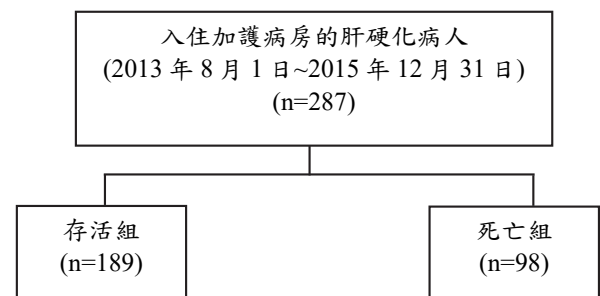
二、內科加護房肝硬化病人定義

肝硬化是由壞死性炎症和肝纖維形成的各種肝損傷引起的肝臟硬化現象¹⁵。本研究收案對象為入住加護病房且診斷為肝硬化之病人，並且排除18歲以下、轉院病人，收案流程如圖一。

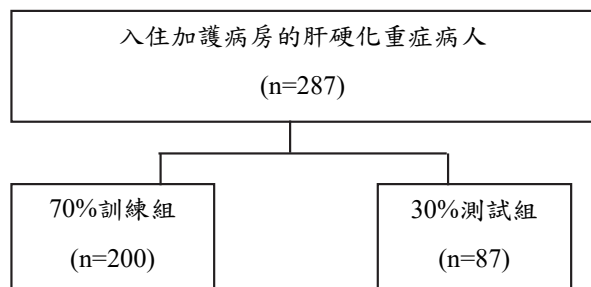
三、資料處理與分析

資料庫收集變項參照於國內外研究文獻和醫師臨床判斷，生理數值變項採用入住加護病房24小時內第一筆數據。個人背景變項包含：年齡、性別、有無急救(cardiopulmonary resuscitation, CPR)、過去病史：肝硬化(Cirrhosis)、吸菸(Smoke)、酒精(Alcohol)；肝硬化相關合併症包含：酒精性肝炎、食道靜脈曲張、腹水、肝性腦病變、病毒性肝炎；入院24小時初次測量生理數值包含：血色素(Hemoglobin)、白血球(WBC)、血小板(Platelets)、白蛋白(Albumin)、總膽紅素(bilirubin)、凝血時間(international normalized ratio, INR)、肌酸酐(creatinine)、腎絲球過濾率(glomerular filtration rate, GFR)、鈉(Sodium, Na)；傳統臨床評估量表包含：APACHE II、MELD score、CTP score；以及照護結果：本次入住加護病房中為死亡或存活。

首先將資料進行10折k折交叉驗證(k-fold cross-validation, K-fold)(如圖二)，再將資料庫(N=287)以隨機亂數切割並設置亂數種子固定結果，為避免選擇偏差或過度擬合再以交叉驗證確認模型性能，分為訓練集70%(N=200)及測試集30%(N=87)，訓練集用作建構模型，測試集驗證模型效果。



圖一：加護病房肝硬化病人收案流程。



圖二、資料切割流程。

四、研究工具及方法

研究方法工具為 R STUDIO 3.6.0 版本，使用的機器學習方法包括 Bagging、GLM、Ada-boost、Neyman-Pearson、Stacking。

(一) Bootstrap aggregating (Bagging)

重複從訓練資料中隨機抽取資料用以建立多個分類器，分類器多半使用相同方法，屬同質整合¹⁶。最終以投票或平均得出結果，這個抽樣模式在統計上稱為 bootstrap 可以降低模型的不穩定性¹⁷。

(二) Generalized linear model (GLM)

當預測值分布為二項式，模型用的鏈結函數 (link function) 是羅吉特函式 (logit)，以此建模的便稱為羅吉斯迴歸 (Logistic regression) 模式。

羅吉斯迴歸 (Logistic regression) 基礎是建立在結果和自變項之間呈現直線性關聯假設¹⁸。目前廣泛應用於不同的研究中。透過建立模型可以獲得事件發生的機率，而在二分類預測中，根據其預測出的機率可歸類為“1”和“0”¹⁹。

(三) Adaptive Boosting (Adaboost)

即 Adaptive Boosting (自適應增強)，是一種透過決策樹改進 Boosting 的算法²⁰，首先給予訓練集中每個樣本相同權重，接著每次疊代對分類錯誤的樣本增加權重，使下一次的疊代更重視學習這些樣本，是為提高學習算法通過結合“弱分類器”的一種算法²¹⁻²²。主要理念為和多個弱學習器 (學習效果較差的模型) 互相

學習以提高效果¹³。通過改變分類錯誤樣本的權重來改變下一個弱分類器的分類邊界，從而提升分類演算法的準確度。

(四) Neyman-Pearson (NP)

疾病診斷會通常期望能限制統計 I 型錯誤 (將類別 0 觀察錯誤分類為類別 1 的條件機率，即將患病患者誤診為健康患者)，並使其維持在閾值以下。Neyman-Pearson (NP) 分類不僅能夠能夠減少統計 II 型錯誤 (錯誤分類的條件機率，將健康患者誤診為患病患者) 並控制統計 I 型錯誤在一定的範圍內²³。

儘管 NP 在假設檢驗理念已存在超過一世紀，但使用何者方法實踐其概念仍是挑戰²⁴，2018 年 Yang 等人發表一種傘型算法可利用邏輯回歸、支持向量、隨機森林等實現 Neyman-Pearson 概念²⁴。

(五) Stacking

Stacking 堆疊是集成多個分類器的方法，訓練初級學習器用以訓練次級模型，意即能夠合併使用不同的學習算法生成的分類器，從而實現更高的預測精度²⁵。堆疊使用的基礎模型通常是不同的學習演算法，屬於異質整合，藉此獲得最佳的方式來整合預測結果並確保其多樣性²⁶。

結 果

一、描述性統計

本研究資料庫描述性統計，加護病房整體死亡率為 34.1% (98/287)；年齡平均為 54.98 (標準差 14.25)、性別以男性佔多數 214 筆 (74.6%)、有進行急救的 4 筆 (1.4%)、患有肝硬化病史病人居多 231 筆 (80.5%)；合併有酒精性肝炎 152 筆 (53.0%)、食道靜脈曲張 121 筆 (42.2%) 及腹水 137 筆 (51.3%) 約佔一半、有肝性腦病變 163 筆 (56.8%) 和病毒性肝炎 98 筆 (34.1%)；生理檢驗數值：血色素平均值 9.59 (±2.77)、白血球平均值 11.23 (±7.42)、白蛋白平均值 2.61 (±0.59)、總膽紅素平均值 5.23 (±6.72)、凝血時間平均值 1.65 (±1.31)、

表一：肝硬化重症病人加護病房住院中死亡與否之比較

變項名稱	整體 (n=287)	存活組 (n=189)	死亡組 (n=98)
年齡 (Year) (Mean ± SD)	54.98 (± 14.24)	54.80 (± 13.02)	55.32 (± 16.34)
性別 (Gender)			
男 n(%)	214(74.6%)	140 (74.1%)	74(75.5%)
女 n(%)	73(25.4%)	49 (25.9%)	24(24.5%)
有無急救 (CPR)			
無 n(%)	283(98.6%)	187(98.9%)	96(98.0%)
有 n(%)	4(1.4%)	2 (1.1%)	2(2.0%)
肝硬化病史 (Cirrhosis)			
無 n(%)	56(19.5%)	40(21.2%)	16(16.3%)
有 n(%)	231(80.5%)	149(78.8%)	82(83.7%)
吸菸病史 (Smoke)			
無 n(%)	72(25.1%)	43(22.8%)	29(29.6%)
有 n(%)	154(53.7%)	100(52.9%)	54(55.1%)
戒菸 n(%)	61(21.2%)	46(24.3%)	15(15.3%)
酒精病史 (Alcohol)			
無 n(%)	59(20.6%)	36(19.0%)	23(23.5%)
有 n(%)	135(47.0%)	85(45.0%)	50(51.0%)
戒酒 n(%)	93(32.4%)	68(36.0%)	25(25.5%)
酒精性肝炎			
無 n(%)	135(47.0%)	86(45.5%)	49(50.0%)
有 n(%)	152(53.0%)	103(54.5%)	49(50.0%)
食道靜脈曲張			
無 n(%)	166(57.8%)	110(58.2%)	56(57.1%)
有 n(%)	121(42.2%)	79(41.8%)	42(42.9%)
腹水			
無 n(%)	150(52.3%)	107(56.6%)	43(43.9%)
少量 n(%)	61(21.3%)	43(22.8%)	18(18.4%)
中量 n(%)	52(18.1%)	26(13.8%)	26(26.5%)
多量 n(%)	24(15.3%)	13(6.9%)	11(11.2%)
肝性腦病變			
無 n(%)	163(56.8%)	127(67.2%)	36(36.6%)
混亂 n(%)	42(14.6%)	24 (12.7%)	18(18.4%)
嗜睡 n(%)	22(7.7%)	12(6.3%)	10(10.2%)
半昏迷 n(%)	32(11.1%)	15 (7.9%)	17(17.4%)
昏迷 n(%)	28(9.8%)	11(5.8%)	17(17.4%)
病毒性肝炎			
無 n(%)	189(65.9%)	126(66.7%)	63(64.3%)
B 型 n(%)	56(19.5%)	37 (19.6%)	19(19.4%)
C 型 n(%)	37(12.9%)	22(11.6%)	15(15.3%)
B 型+C 型 n(%)	5(1.7%)	4(2.1%)	1(1.0%)
血色素 (Hemoglobin)(g/dL)(Mean ± SD)	9.59(± 2.77)	9.53(± 2.77)	9.70(± 2.78)
白血球 (WBC)(10 ³ /uL)(Mean ± SD)	11.23(± 7.42)	10.74(6.84±)	12.71(± 8.38)
血小板 (Platelets)(10 ³ /uL)(Mean ± SD)	115.31(± 94.50)	111.99(94.06±)	121.72(± 95.49)
白蛋白 (albumin)(g/dL)(Mean ± SD)	2.61(± 0.59)	2.65(± 0.53)	2.53(± 0.68)
總膽紅素 (bilirubin)(mg/dL)(Mean ± SD)	5.23(± 6.72)	3.64(± 4.10)	8.32(± 9.29)
凝血時間 (INR)(Mean ± SD)	1.65(± 1.31)	1.57(± 1.49)	1.82(± 0.85)
肌酸酐 (creatinine)(mg/dL)(Mean ± SD)	2.38(± 2.36)	1.90(± 1.69)	3.31(± 3.08)
腎絲球過濾率 (GFR)(mL/min)(Mean ± SD)	51.67(± 38.93)	61.34(± 41.36)	31.53(± 23.69)
鈉 (Sodium,Na)(mEq/L)(Mean ± SD)	137.77(± 7.31)	137.1(± 6.36)	139.09(± 8.74)
APACHE II *1(Mean ± SD)	21.96(± 9.11)	18.92(± 7.55)	27.86(± 9.02)
MELD score *2(Mean ± SD)	21.69(± 8.98)	18.99(± 7.95)	26.92(± 8.59)
CTP score *3(Mean ± SD)	9.13(± 2.05)	8.65(± 1.87)	10.06(± 2.07)

備註：*1. APACHE II：臨床上重症疾病嚴重程度評估，用於評比入住加護病房的優先順序，分數越高代表越嚴重。

*2. MELD score：歐美肝移植評估指標，介於 5-10 分之間，分數越高代表越嚴重。

*3. CTP score：臨床上用於評估肝硬化程度，分數越高代表越嚴重。

肌酸酐平均值 2.38 (± 2.36)、腎絲球過濾率平均值 51.67 (± 38.93)、鈉平均值 137.77 (± 7.31)；傳統評估量表 APACHE II 平均分數 21.96 (± 9.11)、MELD score 平均值分數 21.69 (± 8.98)、CTP score 平均分數 9.13 (± 2.05) 在死亡組平均明顯較高；詳見表一。

二、機器學習模型

(一) 機器學習模型設定

在本研究中利用 R 語言套件實踐 Bagging、GLM、Adaboost、Neyman-Pearson、Stacking 算法，設定參數如表二。

(二) 模型效果

本研究交叉驗證採用 10 次 K-fold，再將其分為 70% 訓練資料與 30% 測試資料。ROC 曲線下的面積 AUC (Area Under Curve) 可以代表

分類器預測能力，測試組交叉驗證 AUC 值結果如表三，本研究選用多數模型 AUC 值最佳的第 1 折模型進行比較。

接者利用混淆矩陣的四個指標 (如表四) 真陽性 (True Positive, TP)：預測值為正樣本，實際值為正樣本；真陰性 (True Negative, TN)：預測值為負樣本，實際值為負樣本；偽陽性 (False Positive, FP)：預測值為正樣本，實際值為負樣本；統計 I 型錯誤；偽陰性 (False Negative, FN)：預測值為負樣本，實際值為正樣本；統計 II 型錯誤)。所計算出的準確率 (Accuracy)、敏感性 (Sensitivity)、特異性 (Specificity)、F1 分數 (F1 Score) 呈現模型預測效果。

準確率 (Accuracy) = $(TP+TN)/(TP+TN+FP+FN)$

敏感性 (Sensitivity) = $TP/(TP+FN)$

特異性 (Specificity) = $TN/(FP+TN)$

精準率 (Precision) = $TP/(TP+FP)$

F1 分數 (F1 Score) = $(2 \times Precision \times Sensitivity) / (Precision + Sensitivity)$

表二：模型參數設定

模型	設定參數
Bagging	Nbagg = 100
GLM	無調整
Adaboost	n_rounds = 300
Neyman-Pearson	method = nb, alpha = 0.2
Stacking	glm, ranger

表四：混淆矩陣

	預測值		
	正樣本	負樣本	
實際值	正樣本	TP	FN
	負樣本	FP	TN

表三：模型交叉驗證之 AUC 值比較

	Bagging	GLM	Adaboost	Neyman-Pearson	Stacking
1-fold	0.866	0.868	0.818	0.871	0.863
2-fold	0.815	0.811	0.801	0.837	0.825
3-fold	0.784	0.832	0.768	0.846	0.827
4-fold	0.734	0.770	0.693	0.738	0.743
5-fold	0.827	0.825	0.772	0.826	0.814
6-fold	0.818	0.745	0.772	0.782	0.800
7-fold	0.809	0.795	0.755	0.835	0.816
8-fold	0.815	0.758	0.792	0.823	0.805
9-fold	0.863	0.869	0.800	0.845	0.846
10-fold	0.805	0.815	0.786	0.804	0.818
平均	0.813	0.808	0.775	0.820	0.815

表五呈現五個模型在測試集中對重症肝硬化病人的死亡預測的性能。由表得知模型效果由高到低分別為 Neyman-Pearson (準確率：0.862)，接著是 GLM、Adaboost (準確率：0.827)，最後是 Bagging、Stacking (準確率：0.804)。

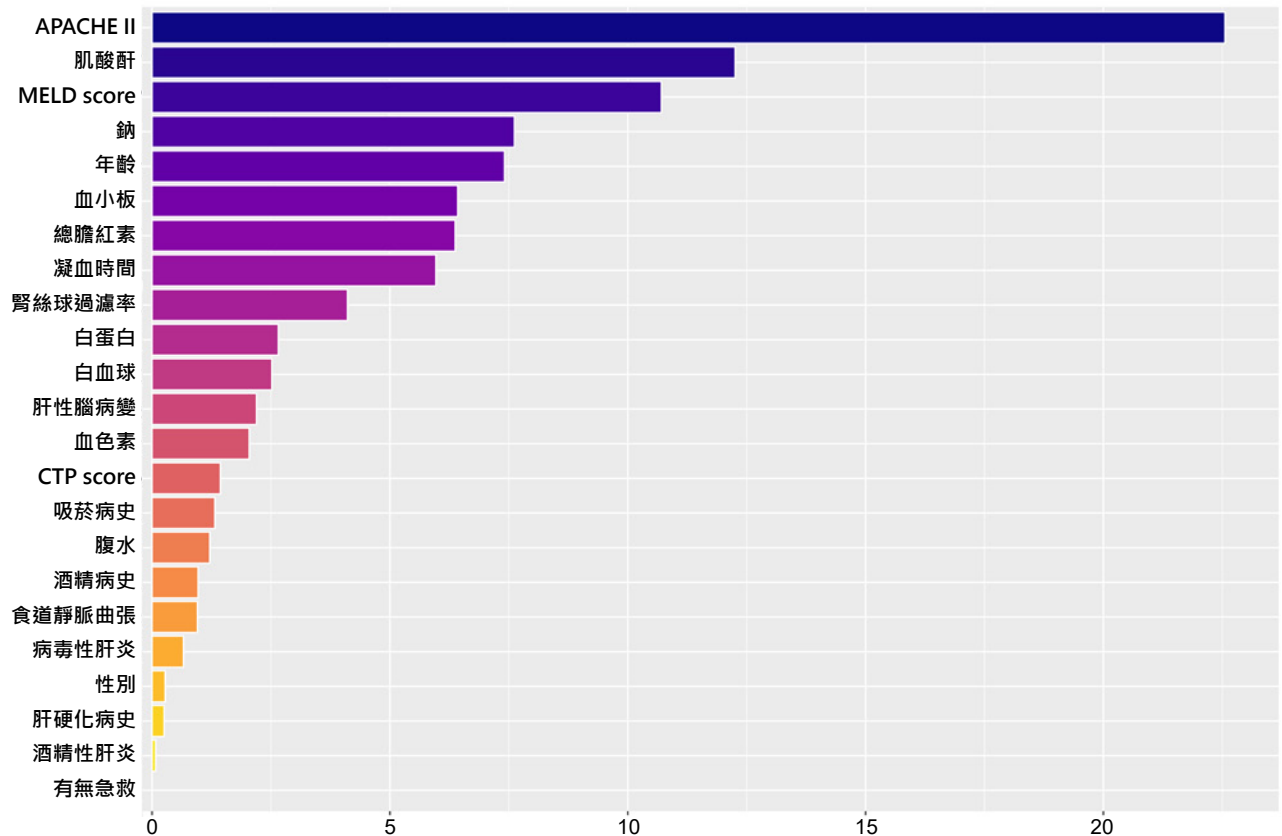
圖三為 Bagging 模型重要性變量圖，在 Bagging 模型中，認為 APACHE II 是預測死亡的最重要因素，接著是肌酸酐、MELD score、鈉。

(三) 模型之 ROC 曲線、ROC 曲線下的面積 AUC (Area Under Curve)

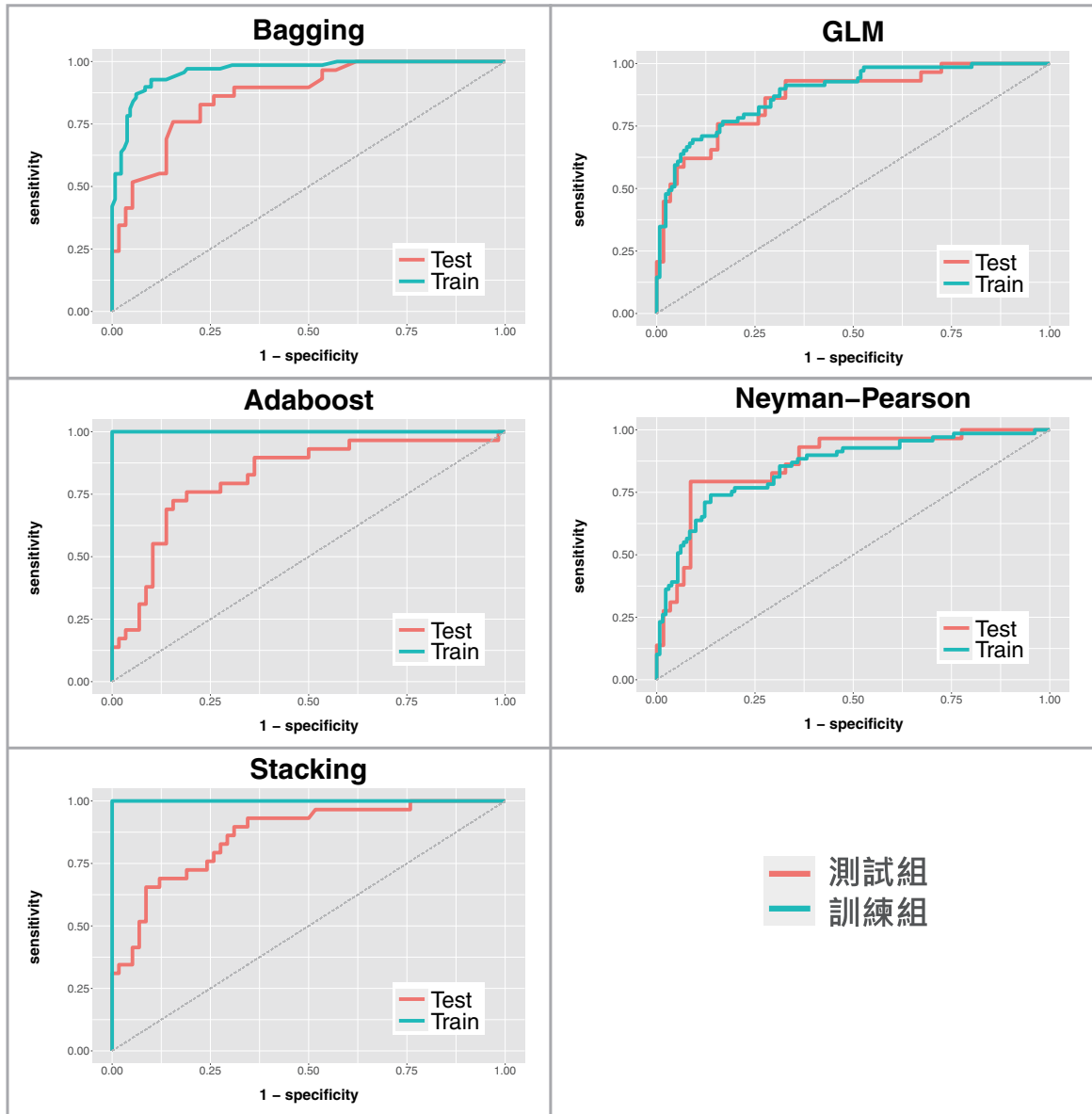
ROC 曲線為真陽性率 (True positive rate, TPR) 與偽陽性率 (False positive rate, FPR) 之間的相對關係，越靠近左上越佳。五個模型訓練組與測試組 ROC 曲線如圖四。圖五為合併五個模型測試組 ROC 比較圖，得知使用 Neyman-Pearson 模型效果最佳 (藍色，AUC=0.871)。

表五：五個模型的使用測試集在預測加護病房肝硬化病人死亡率之準確率、敏感性、特異性、F1 Score、AUC

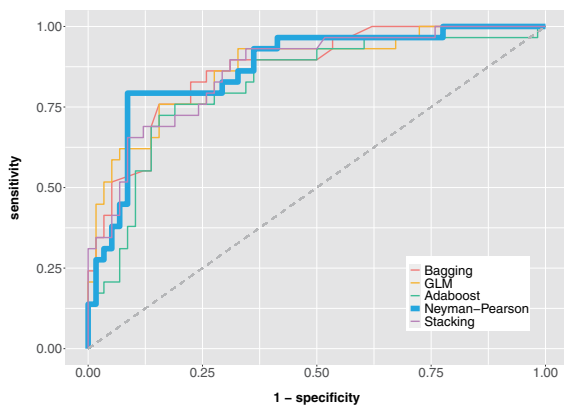
	Accuracy	Sensitivity	Specificity	F1 Score	AUC
Bagging	0.804	0.517	0.948	0.638	0.866
GLM	0.827	0.620	0.931	0.705	0.868
Adaboost	0.827	0.689	0.896	0.727	0.818
Neyman-Pearson	0.862	0.793	0.896	0.793	0.871
Stacking	0.804	0.586	0.913	0.666	0.863



圖三：Bagging 模型重要性變量圖。



圖四：五個模型：Bagging、GLM、Adaboost、Neyman-Pearson、Stacking 訓練組 (Train) 與測試組 (Test) 模型之 ROC 曲線。

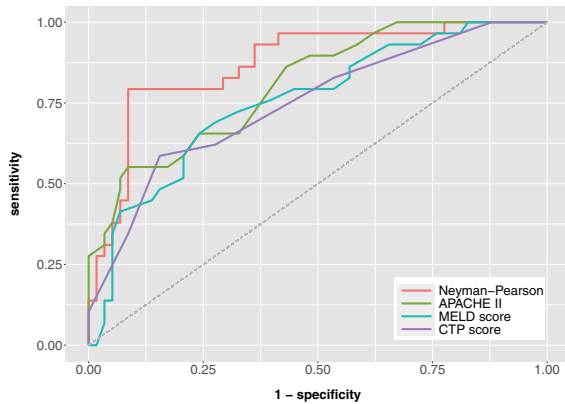


圖五：五個模型：Bagging、GLM、Adaboost、Neyman-Pearson、Stacking 之 ROC 曲線。

TPR= Sensitivity

FPR= 1- Specificity

圖六為準確率最高的 Neyman-Pearson 模型與傳統臨床三種評估量表 (CTP score、MELD score、APACHE II) 之 ROC 比較圖，皆採用測試組數據進行比較。由圖可得知 Neyman-Pearson 的效果最佳且優於所有傳統指標 (紅色, AUC = 0.871)，接著依序為 APACHE II (綠色, AUC = 0.806)、MELD score (藍色, AUC = 0.755)、CTP score (紫色, AUC = 0.747)。



圖六：Neyman-Pearson 與 APACHE II、MELD score、CTP score 指標模型之 ROC 曲線。

討論

本研究建構的五個機器學習模型的 AUC 都在 0.81 以上，皆優於傳統臨床評估量表 (APACHE II 之 AUC 值：0.807、MELD score 之 AUC 值：0.755、CTP score 之 AUC 值：0.747)，APACHE II 優於 MELD score、CTP score 最低，在五個模型中 Neyman-Pearson 具有最佳性能 AUC 高達 0.871 (準確率 86.2%)。

為驗證本研究機器學習模型在死亡預測能力方面比傳統臨床評估量表現更佳，選用 APACHE II、MELD score、CTP score 作比較對象。Peng 等人對 MELD score、CTP score 肝硬化的預後進行系統評價和整合分析，發現當評估對象為 ICU 病人，MELD score 能力優於 CTP score²⁷，不過 MELD score 卻缺少肝硬化常見的不良併發症：例如肝性腦病變、食道靜脈曲張破裂，且不適用於急性重症情況^{28,29}，因此本研究將肝硬化常見併發症加入預測因子。而 APACHE II 主要由各種器官系統的生理參數組成，此與 Ho 等人研究認為雖然 CTP score 可準確預測重症肝硬化病人在醫院死亡率，但在屬於急性重症的加護病房環境下，APACHE II 預測死亡能力仍舊比 CTP score 更佳⁴，因本研究對象為加護病房重症病人故以該結果為參考加入 APACHE II 進行比較，期望能獲得更精準的預測結果。

機器學習具有預測能力並且能夠應用在醫學診斷，過去已有研究指出機器學習在肝硬化

病人死亡預測上具可行性，例如 Lin 等人使用 1214 例數據並利用多種學習方法：線性判別分析 (Latent Dirichlet Allocation, LDA)，支持向量機 (Support vector machine, SVM)、樸素貝葉斯分類器 (Naive Bayes Classifiers)、決策樹 (Decision Tree)、隨機森林 (Random Decision Tree) 和 Adaboost 預測死亡率及對 ESLD 病人進行分類，其中隨機森林有最佳性能 (AUC: 0.85) 比傳統臨床評估量模型 MELD score 效果佳 (AUC: 0.76)¹³，Kanwal 等人 (2020) 透過 107,939 筆數據開發肝硬化死亡預測模型：CiMM，能夠準確預測 1 年，2 年和 3 年死亡率 (AUC 0.78、0.76 和 0.71)³⁰。

然本研究建構的預測模型無法保證準確預測病人的生存或死亡³¹，醫學臨床的環境瞬息萬變，在重症環境中會遇到的變化與突發狀況更是難以預料，模型中的規則與現實條件相比總會有例外。雖然機器學習通常具有很高的準確率，但是其複雜黑盒結構卻限制可用性，難以被廣泛運用，最佳方法是同時使用高準確率的機器學習模型與兼備易衡量性和透明度的傳統評估量表^{14,30}。因此本研究的死亡預測模型不適合作為主要決策工具，而是做為臨床輔助工具，期望模型能夠輔助醫療人員提早篩出可能死亡的病人以便提早提供醫療照顧。

研究限制

本研究收集資料範圍僅為東區某區域教學醫院內科加護病房，受限於樣本數量不足，其結果可能無法外推至其他醫院。資料庫重症相關變項未考量病人該次入住加護病房的主要因素及加護病房常見疾病，例如：敗血性休克、呼吸衰竭等，由於其定義較為複雜難以於病人入住加護病房 24 小時內判斷，而本研究數據以入住加護病房院 24 小時內為主，該變項難以在收集時限內獲取完整且正確的資料。

機器學習可解釋性往往較為薄弱，後續研究可考慮增加入住加護病房相關疾病因素後，利用重要性變量分析，進一步篩選出關鍵變量，同時利用統計學探討各變數之間的關係，期望日後能與其他層級醫療院所合作擴大收集

樣本，或日後收集更多研究數據，藉此可從資料庫中篩選影響力較大的關鍵變量增加模型可解釋力，再由實證資料來持續驗證模型可信度。

誌謝

向內科加護病房醫療團隊人員以及收案病人致意使本研究得以完成。

參考文獻

- 108年國人死因統計結果。衛生福利部。2020年1月3日，取自<https://www.mohw.gov.tw/cp-16-54482-1.html>。
- Wang X, Wu B. Critical issues in the diagnosis and treatment of liver cirrhosis. *Gastroenterology Report* 2019;7(4):227-30. doi:10.1093/gastro/goz024
- Philip G, Runyon BA. Treatment of patients with cirrhosis. *N Engl J Med* 2016; 375(21):2102-5. doi:10.1056/nejmc1612334
- Ho YP, Chen YC, Yang C, et al. Outcome prediction for critically ill cirrhotic patients: A comparison of APACHE II and Child-Pugh scoring systems. *Intensive Care Med* 2004;19(2):105-10. doi:10.1177/0885066603261991
- Olson JC, Wendon JA, Kramer DJ, et al. Intensive care of the patient with cirrhosis. *Hepatology* 2011;54(5):1864-72. doi:10.1002/hep.24622
- Levesque E, Hoti E, Azoulay D, et al. Prospective evaluation of the prognostic scores for cirrhotic patients admitted to an intensive care unit. *J Hepatol* 2012;56(1):95-102. doi:10.1016/j.jhep.2011.06.024
- Poynard T, Naveau S, Doffoel M, et al. Evaluation of efficacy of liver transplantation in alcoholic cirrhosis using matched and simulated controls: 5-year survival. *J Hepatol* 1999;30(6):1130-7. doi:10.1016/s0168-8278(99)80269-4
- 108年度器官捐贈人數統計表。器官移植登錄中心。2021年1月3日，取自<https://www.torsc.org.tw/docList.jsp?uid=158&pid=9&rn=1636018696>。
- Brown J, Sorrell JH, McClaren J, Creswell JW. Waiting for a Liver Transplant. *Qualitative Health Research* 2006; 16(1):119-36. doi:10.1177/1049732305284011
- 黃欣智、蔡銘鴻、陳邦基。肝硬化的診治概要。內科學誌 2012 ; 23(6) : 392-7。
- Bao Q, Wang B, Yu L, Weng H, Ge J, Li L. A modified prognostic score for critically ill patients with cirrhosis: an observational study. *Gastroenterol Hepatol* 2016;31(2):450-458. doi:10.1111/jgh.13076
- Gall JR. L, Loirat P, Alperovitch A. APACHE II-A Severity of Disease Classification System. *Crit Care Med* 1986;14(8):754. doi:10.1097/00003246-198608000-00027
- Lin YJ, Chen RJ, Tang JH, et al. Machine-Learning Monitoring System for Predicting Mortality Among Patients With Noncancer End-Stage Liver Disease: Retrospective Study. *JMIR Medical Informatics* 2020;8(10):e24305. doi:10.2196/24305
- Michie D, Spiegelhalter DJ, Taylor CC. Machine learning. *Neural and Statistical Classification* 1994;91(433):436-8. doi:10.2307/2291432
- Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. *Lancet* 2014; 383(9930):1749-61. doi:10.1016/s0140-6736(14)60121-5.
- Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning* 2000;40(2):139-57. doi:10.1023/a:1007607513941
- Breiman L. Bagging predictors. *Machine learning* 1996; 24(2):123-40. doi:10.1023/a:1018054314350
- Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med* 2011;18(10):1099-104. doi:10.1111/j.1553-2712.2011.01185.x
- Lee ES. Exploring the performance of stacking classifier to predict depression among the elderly. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) 2017;13-20. doi:10.1109/ICHI.2017.95
- Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 1997;55(1):119-39. doi:10.1006/jcss.1997.1504
- Schapire RE. The strength of weak learnability. *Machine Learning* 1990;5(2):197-227. doi:10.1023/a:1022648800760
- Sagi O, Rokach L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2018;8(4):e1249. doi:10.1002/widm.1249
- Freeman EA, Moisen GG. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling* 2008;217(1-2):48-58. doi:10.1016/j.ecolmodel.2008.05.015
- Tong X, Feng Y, Li JJ. Neyman-Pearson classification algorithms and NP receiver operating characteristics. *Science Advances* 2018;4(2):eaao1659. doi:10.1126/sciadv.aao1659
- Wolpert DH. Stacked generalization. *Neural Networks* 1992;5(2):241-59. doi:10.1016/s0893-6080(05)80023-1
- Zenko B, Todorovski L, Dzeroski S. A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. In *Proceedings 2001 IEEE International Conference on Data Mining* 2001;669-70. IEEE. doi:10.1109/icdm.2001.989601
- Peng Y, Qi X, Guo X. Child-Pugh versus MELD score for the assessment of prognosis in liver cirrhosis: A systematic review and meta-analysis of observational studies. *Medicine* 2016;95(8):e2877. doi:10.1097/MD.0000000000002877
- Rassameehiran S, Woreta T. The use of MELD scores in critically ill cirrhotic patients. *The Southwest Respiratory and Critical Care Chronicles* 2016;4(16):45-50. doi:10.12746/swrccc2016.0416.219
- Kim HJ, Lee HW. Important predictor of mortality in patients with end-stage liver disease. *Clinical and Molecular Hepatology* 2013;19(2):105-15. doi:10.3350/cmh.2013.19.2.105
- Kanwal F, Taylor TJ, Kramer JR, et al. Development, validation, and evaluation of a simple machine learning model to predict cirrhosis mortality. *JAMA Network Open* 2020;3(11):e2023780-e2023780. doi:10.1001/jamanetworko-

pen.2020.23780

31. Zimmerman JE, Wagner DP, Seneff MG, Becker RB, Sun X, Knaus WA. Intensive care unit admissions with cirrhosis:

Risk-stratifying patient groups and predicting individual survival. *Hepatology* 1996;23(6):1393-401. doi:10.1053/jhep.1996.v23.pm0008675156

Machine Learning to Predict the Death of Patient in the Intensive Care Unit with Cirrhosis

Jui-Chu Ma¹, Lang-Yin Zeng², Chin-Hsin Li¹, Wan-Ju Cheng¹,
Chia-Ying Hsiao³, Min-I Su³, and Yeong-Cheng Liou^{4,5}

¹*Department of Nursing, ²Department of Medical Research, ³Department of Internal Medicine, Taitung MacKay Memorial Hospital, Taitung, Taiwan;*

⁴*Department of Healthcare Administration and Medical Informatics, and Research Center of Non-linear Analysis and Optimization Kaohsiung Medical University;*

⁵*Department of Medical Research, Kaohsiung Medical University Hospital*

Machine learning (ML) is used in a wide variety of applications and achieve certain results in medical science. At present, there is no clinical assessment scale to available patients with critical cirrhosis patients. Therefore, this study used ML methods to construct a death prediction model for critical cirrhosis patients. A retrospective investigation is designed with electronic medical records to make analysis on 287 patients with cirrhosis in the intensive care unit of a regional hospital from August 1, 2013 to December 31, 2015. Using R language to achieve five ML methods: Bagging, GLM, Adaboost, Neyman-Pearson, and Stacking to construct five models. The study compare the discriminative ability of traditional clinical scoring systems: CTP score and MELD score, and APACHE II. The accuracy of the five ML models in this research was above 80%. Top-performance is Neyman-Pearson accuracy rate was 86.2% (AUC=0.871), is get ahead of traditional scoring system: APACHE II (AUC=0.806), MELD score (AUC=0.755), CTP score (AUC=0.747), and its prediction results are expected to provide reference for medical decision making. (*J Intern Med Taiwan* 2021; 32: 40-49)