

# 以基於卷積神經網路之自動編碼器與機器學習模型配合結構磁共振影特徵區分雙極性疾患

## A convolutional neural network-based autoencoder and machine learning model for identifying bipolar disorder patients using structural MRI features

學生：鄭茹云

指導老師：翁駿程

### 研究目的

本專題主要在研究雙極性疾患（Bipolar Disorder）與正常大腦之間的影像差異預測，雙極性疾患是一種情緒障礙症，主要依賴於臨床問診和症狀觀察，因此對腦部變異的檢測效果有限，所以我們開發了一個程式系統，利用非監督式學習的方法搭配機器學習與堆疊集成模型來預測大腦是否出現雙極性疾患症狀，期望研究能準確預測雙極性疾患並輔佐醫師做出更為客觀的診斷，提高診斷的準確性和效率。

### 材料與方法

收案：雙極性疾患患者62位(男23,女39,成大醫院)及正常62位(男26,女36,社區招募)，共124位

機器學習：隨機森林(random forest, RF)、極限樹(ExtraTreesClassifier)、

高斯過程分類 ( GaussianProcessClassifier, GPC)、

邏輯迴歸 (logistic regression,LR)、極限梯度提升(extreme gradient boosting, XGB)

模型評估： Hold-Out method

影像參數：

(1)磁場大小：3T

(2)Repetition time (TR) / echo time (TE) = 8000/115 ms

(3)Field of view (FOV)：250 × 250 mm<sup>2</sup>

(4)Matrix size = 128 × 128

(5)Slices = 40, Slice thickness = 3 mm

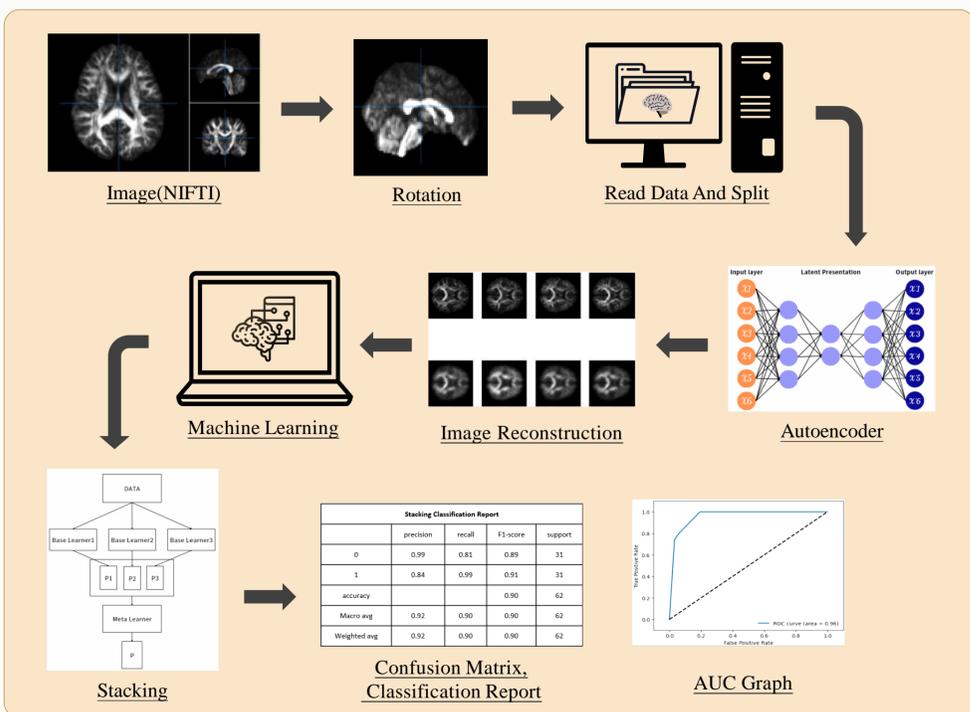
(6)In-plane resolution = 1.95 × 1.95 mm<sup>2</sup>

(7)Signal average = 1

(8)96 noncollinear diffusion weighting gradient directions with b = 2500 s/mm<sup>2</sup>

(9)1 null image without diffusion weighting (b = 0 s/mm<sup>2</sup>)

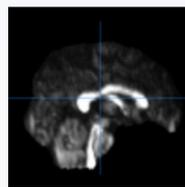
### 系統流程圖：



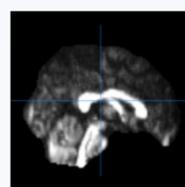
研究中，我們利用腦部磁共振造影影像，結合基於機器學習的方法來進行分析，首先我們進行了資料增補以擴充我們的樣本集，並使用流出法將資料集劃分為訓練集(80%)、驗證集(10%)和測試集(10%)，以進行模型的訓練和評估。

接下來採用Autoencoder模型的降維和升維來抓取特徵，以及後續影像重建的數據，並使用多種機器學習方法來訓練模型，調整超參數以尋找最適合的模型參數，最後再使用堆疊的方式搭配邏輯迴歸對所有機器學習模型進行最終評估，並使用混淆矩陣、分類報告和 Area Under Curve (AUC) Graph來呈現結果。

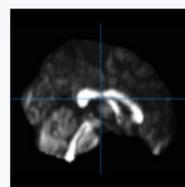
### 資料增補：影像旋轉(範例圖)



逆時針旋轉5度



原始影像



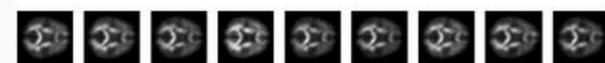
順時針旋轉5度

順時針&逆時針各旋轉1度和2度進行資料增補

• 雙極性疾患患者62筆影像 -> 310筆影像

• 健康對照組62筆影像 -> 310筆影像

### 影像重建

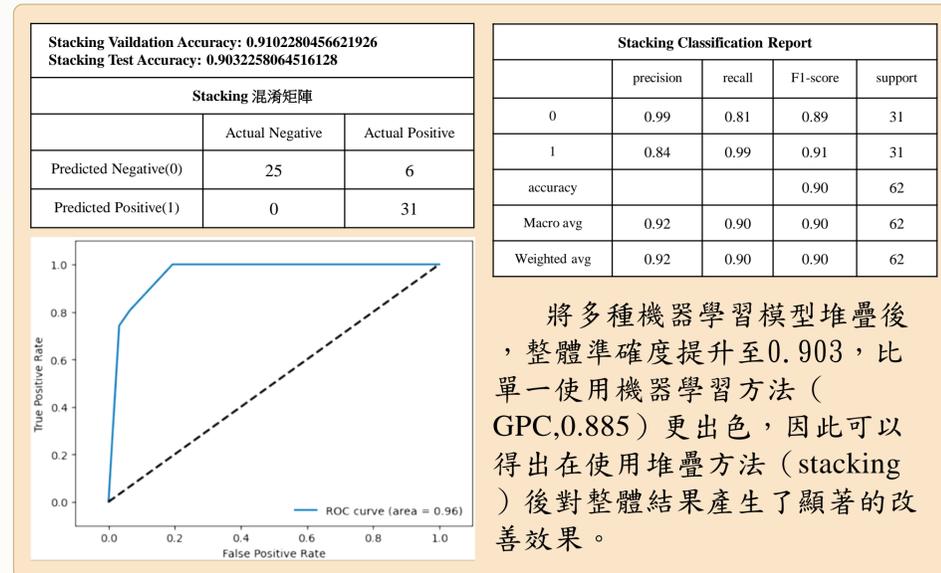


利用Autoencoder，在升維時得取數據並開始訓練影像最後顯示重建後的大腦切面圖並確認影像是否完整。

### 研究結果

	Validation_accuracy	Test_accuracy	ROC curve
RF	0.903	0.871	0.94
ExtraTreesClassifier	0.870	0.854	0.97
GPC	0.887	0.885	0.98
XGB	0.887	0.806	0.94

使用不同分類法在預測雙極性疾患的大腦影像方面取得了不同的性能指標，再透過堆疊搭配邏輯迴歸對指定機器學習模型進行最終評估。



將多種機器學習模型堆疊後，整體準確度提升至0.903，比單一使用機器學習方法 (GPC,0.885) 更出色，因此可以得出在使用堆疊方法 (stacking) 後對整體結果產生了顯著的改善效果。

### 結論

本研究表示，透過自動編碼器和堆疊機器學習分類演算法的組合，可以有效預測雙極障礙患者的大腦影像與正常人之間的差異，但仍需要進一步考慮樣本數、分類法參數、以及不同族群特徵等因素對模型表現的影響，這些因素有望提高我們的預測模型的精確性和可靠性，為雙極性疾患的早期診斷和治療提供更可靠的輔助工具。

### 文獻

1. M.W. Gardner, S.R Dorling (1997) International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, N°1 Artificial neural networks (the multilayer perceptron)
2. J.A. Hamley (1983) A Method of Comparing the Areas under Receiver Operating Characteristic Curve Derived from the Same Case
3. P. Baldi (2012) Autoencoder, Unsupervised Learning, and Deep Architectures
4. Yash Khandelwal (2021) Ensemble Stacking for Machine Learning and Deep Learning