

# 基於核心網網元KPI指標的網路異常預測

## Anomaly Detection in Cellular Networks based on Network Function KPI

專題學生：鄭家宇、張昱威 指導教授：張宏慶

### 簡介

- 隨著互聯網的高速發展，保障網路的穩定性非常重要，網路服務的穩定性主要靠維運人員監控各式各樣的關鍵性能指標(Key Performance Indicators, 簡稱KPI)來判斷網路是否穩定。
- KPI異常檢測指的是透過演算法分析KPI的時間序列數據，並標記網路在何時發生異常，然後利用其時間序列的規律性進行異常預測。
- 本研究的實驗數據來自2021華為某次學習賽提供之某網路營運商之部分網元的KPI真實數據，數據採樣間隔為一小時，我們根據歷史一個月的數據訓練各種模型並預測隨後一周內KPI中的異常情況。
- 本研究從多維度特徵提取和基於機器學習及深度學習模型建構等方面對單維度的KPI數據進行研究，目的是找到有效的特徵提取方式和異常預測方法，從而增強維運人員對網路服務故障的應對能力。

### 資料前處理

- 本研究將102個KPI指標劃分成以下三種類型：

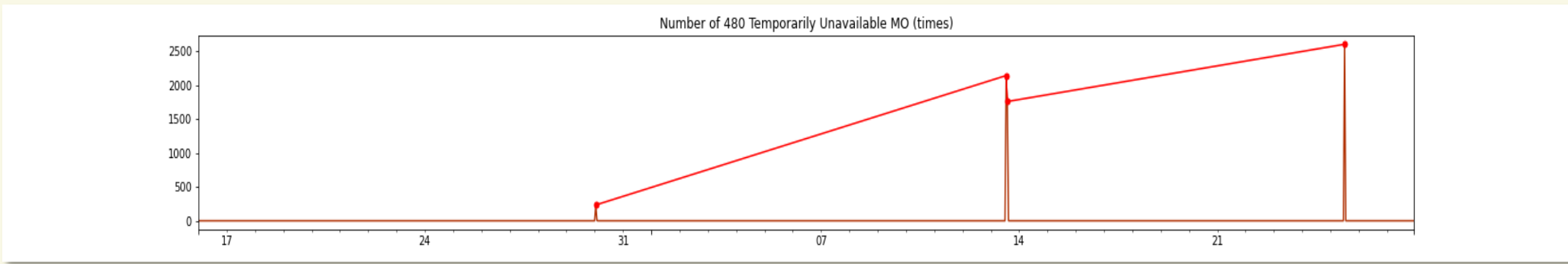


圖1：邊界型異常示意图

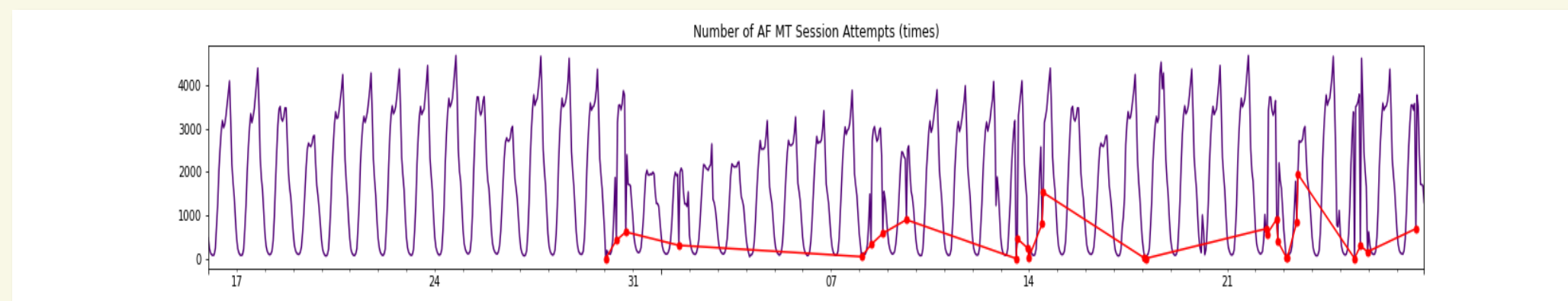


圖2：週期型異常示意图

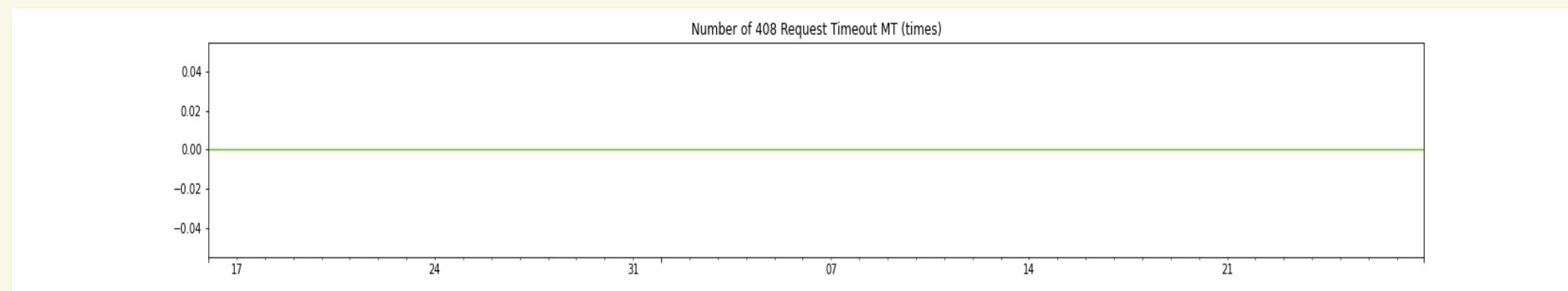


圖3：無異常型異常示意图

- KPI指標分類步驟：

- 決定邊界型異常：將102個KPI依序用Decision tree遍歷，假如只依靠KPI中原生的單一變量得出F1-score大於0.9，則將它分類為邊界型異常KPI。
- 決定無異常型：將KPI中所有Label皆為0(正常樣本)的通通歸納為此類型。
- 決定週期型異常：沒有分類在前述兩種類型的剩餘KPI，皆把它歸類為此類型。

- 最終得到30個邊界型異常，29個無異常型和43個週期型異常KPI。

- 由於原始數據中KPI指標的特徵只有一個，即各時間點上的KPI值，若只使用此單一特徵來進行模型訓練，得到的結果將不符合期待，於是我們需要對其生產出更多特徵，使往後步驟中我們有足夠的上限可以使模型和演算法去逼近。

特徵工程

基礎特徵：  
時間、日期...

統計特徵：  
某段時間範圍的平均值、中位數、  
方差等

平移特徵：  
前一特定時間點的KPI的值

差分特徵：  
當前時間點和上一個時間點或上一個週期中相同時間點的差

趨勢特徵：  
當前時間點的值減去前x時間單位時刻的平均值

- 本研究提取了44種不同特徵供模型訓練使用。

圖4：五種特徵類型

### 實驗流程

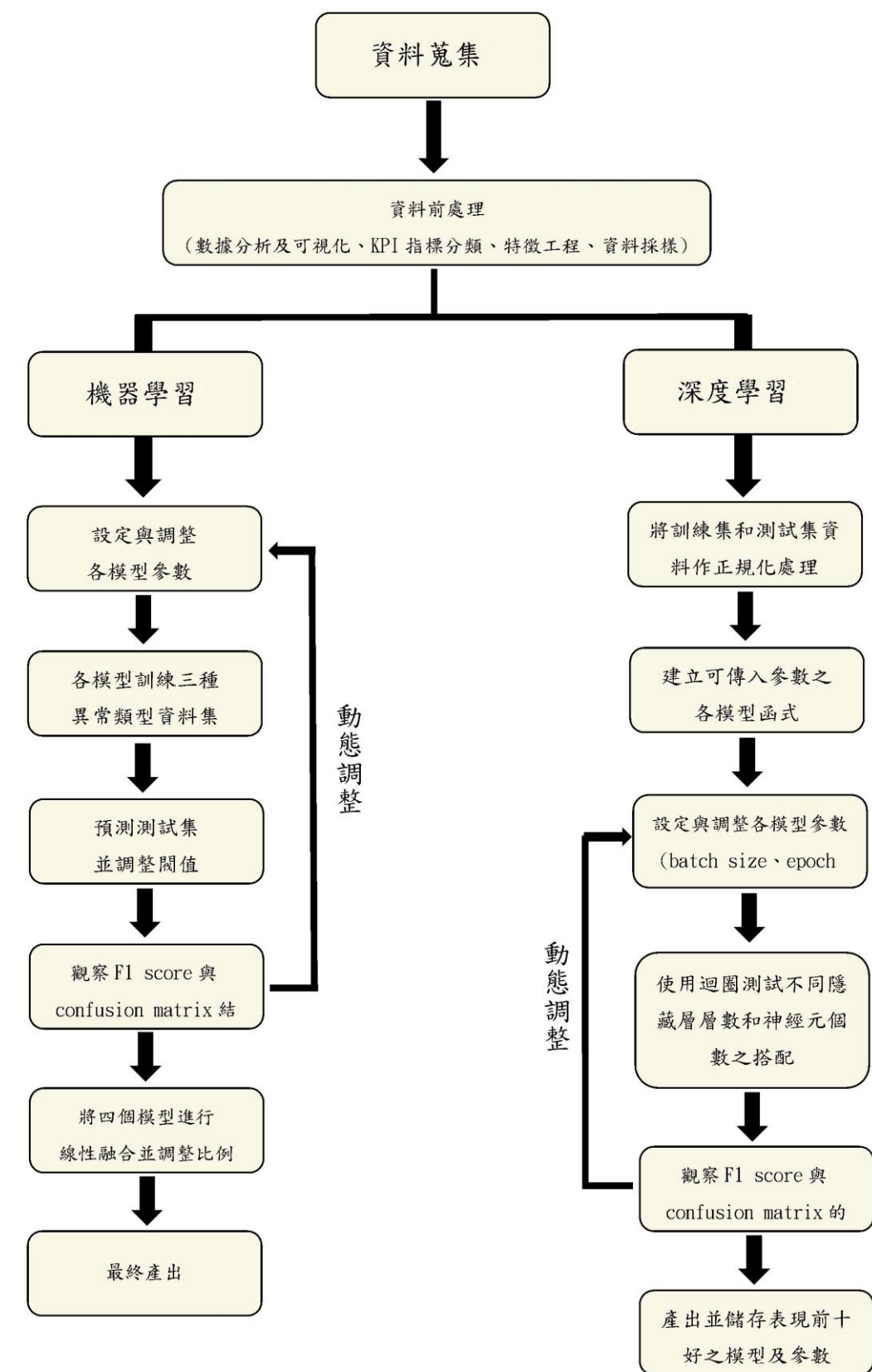


圖5：本研究架構圖

### 實驗結果

- 最終以25:10:25:40的比例融合：XGBOOST、LIGHTGBM、CATBOOST，及NGBOOST並將閾值設定在0.4，獲得將近0.8的F1-score。

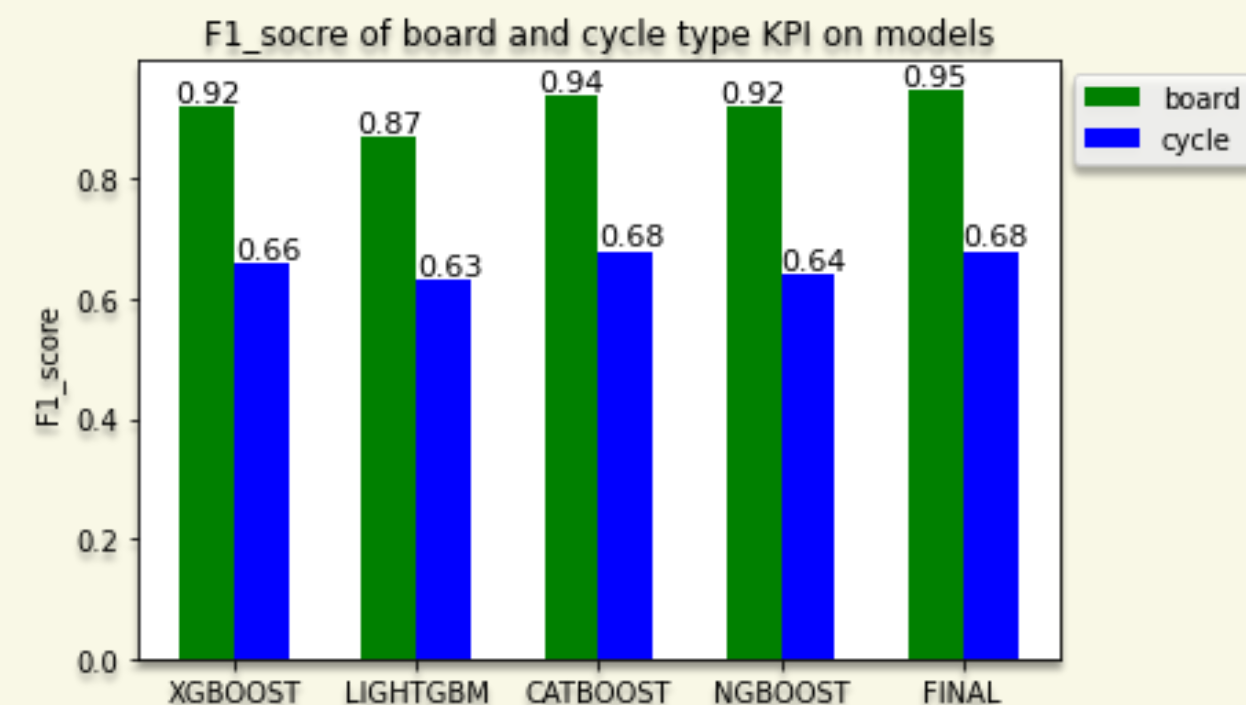


圖6：四個模型在邊界型異常和週期型異常上之表現

- 實際場景中，比較著重於較低的FN因為我們希望只要有異常狀況都可以被預測出來並盡早採取預防措施，才不會使其危害整個網路運作，影響用戶體驗。而FP相對較高雖然會導致誤報數量較多，但考量其較不會危害到整體網路品質，於是我們最終選擇FN相對低的此種融合比例作為最終結果。

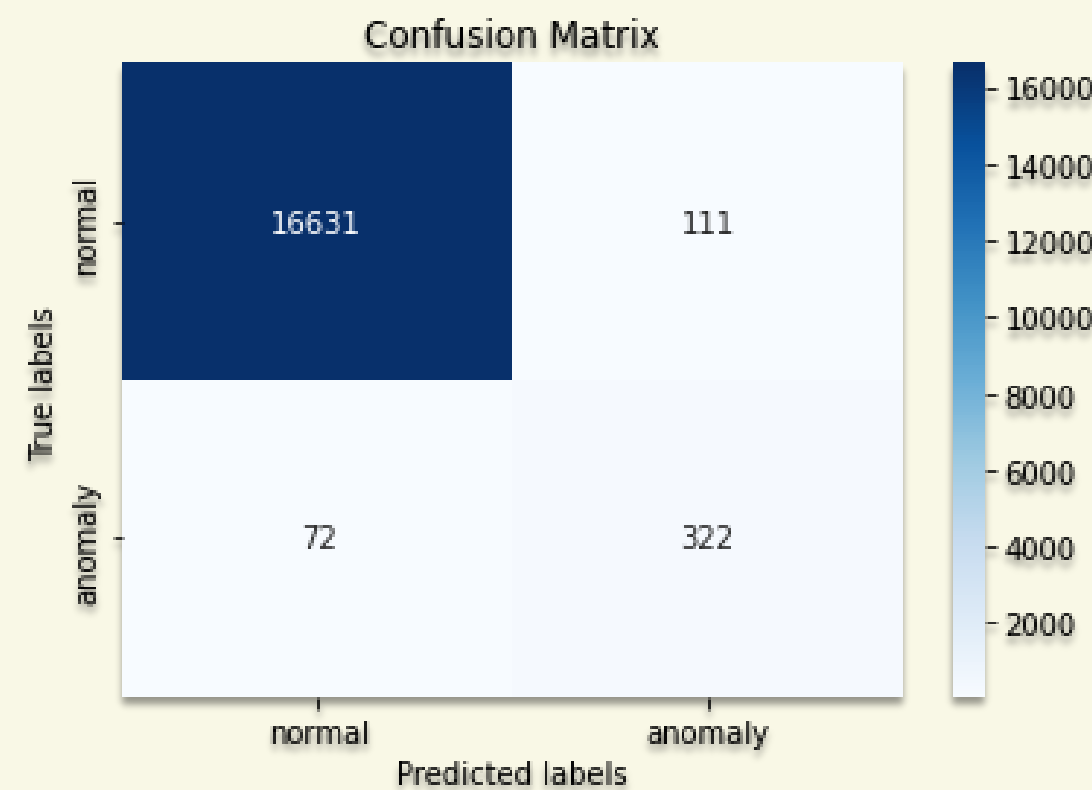


圖7：融合後模型的混淆矩陣

- 從圖中能發現，紅點分布在少隱藏層數、多神經元個數，以及多隱藏層數、少神經元個數的位置。從上述情況推測，過多的隱藏層層數搭配過多的神經元個數，會導致模型複雜度太高，間接導致過度擬合(overfitting)的情況。反之，過少的隱藏層層數搭配過少的神經元個數，可能會導致低度擬合(underfitting)的情況。

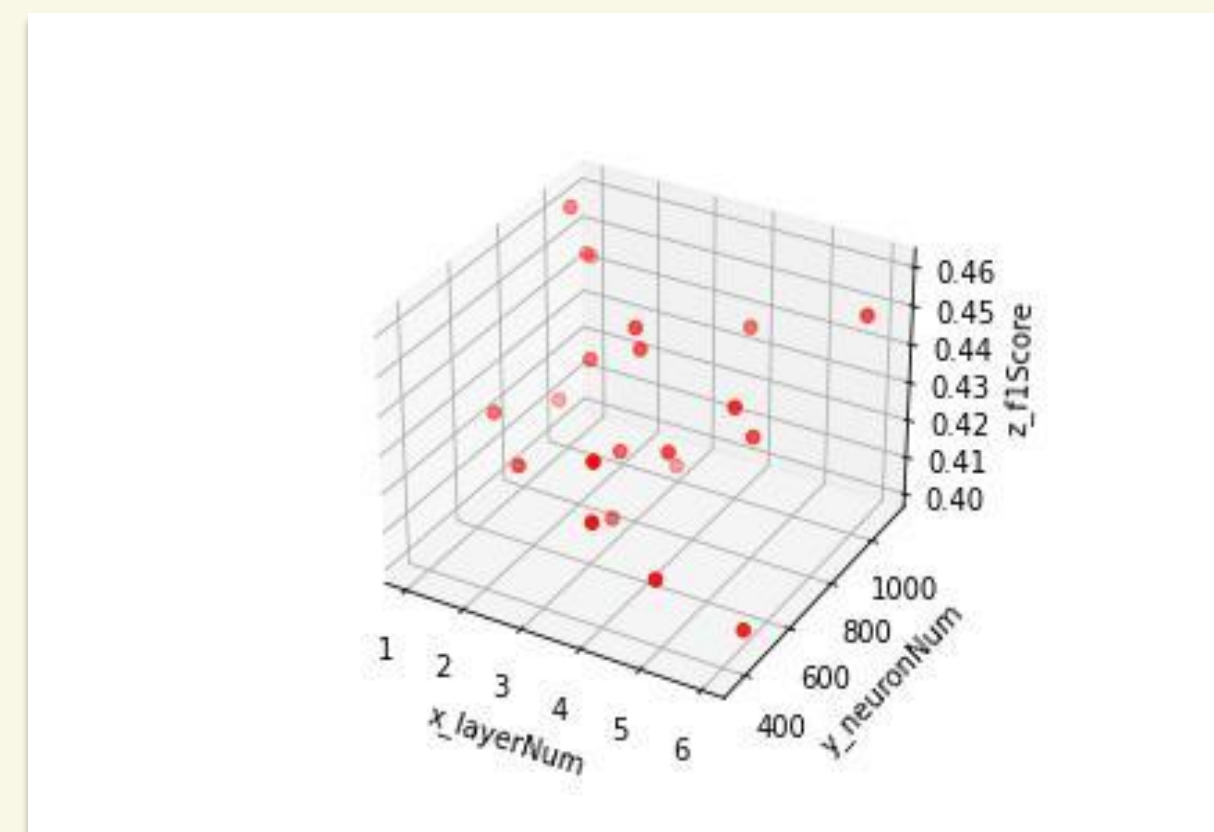


圖8：實驗參數成果F1-score為0.4以上之三維空間顯示圖