國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering
College of Engineering
National Taiwan University

Master's Thesis

應用機器學習於動態三軸試驗資料預測液化循環剪應力比

Application of Machine Learning for Predicting Cyclic Stress Ratio for Liquefaction from Cyclic Triaxial Test Data

蘇筱丰

Hsiao-Feng Su

指導教授：葛宇甯 博士

Advisor: Louis Ge, Ph.D.

中華民國 114 年 7 月

July, 2025

doi:10.6342/NTU202502436

國立臺灣大學碩士學位論文
口試委員會審定書

NATIONAL TAIWAN UNIVERSITY
MASTER'S THESIS ACCEPTANCE CERTIFICATE

應用機器學習於動態三軸試驗資料預測液化循環剪應力比

Application of Machine Learning for Predicting Cyclic Stress Ratio for Liquefaction from Cyclic Triaxial Test Data

本論文係 蘇筱丰 (R12521104) 在國立臺灣大學土木工程學系大地工程組完成之碩士學位論文，於民國114年07月04日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department of Civil Engineering Geotechnical Engineering on 0704,2025 have examined a Master's Thesis entitled above presented by Hsiao-Feng Su (R12521104) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

葛宇甯 (指導教授 Advisor)	葛宇甯
陳柏華	陳柏華
黃郁惟	黃郁惟
朱民虔	朱民虔

系主管Director： 葛宇甯 葛宇甯

doi:10.6342/NTU202502436

誌謝

兩年前戰戰兢兢地來到了國中時期的夢想學校，在經歷了一場學術與研究洗禮後，一眨眼精實的兩年就過去了，隨著此篇誌謝的完稿，也在此畫下了名為碩士的句點。

首先，我想先感謝指導教授--葛宇甯老師的教導，在研究上提供了許多的資源和自由，感謝老師在這兩年間提供了我去泰國參加 34th KKHTCNN 和在成大舉辦的第 20 屆大地工程研討會的機會，在我的研究遇到瓶頸時給予許多的方向與建議，還願意讓我做機器學習的相關研究並如期完成學業，真的非常謝謝老師的支持！我也要感謝葉馥瑄老師、朱民虔老師和黃郁維老師分別在我做振動台、地盤沉陷預測，和機器學習的研究中和我討論並給予指導，讓我能建立穩固的知識基礎，最終順利完成碩論。

另外想感謝其他教導過並給予我幫助的人，包括文迪和昊擎學長教會我如何做振動台試驗；錦德和梁維學長協助 Python code 的 debug；何博 (政恩) 週五團咪後的吃飯時間；佳榮和品秀滿滿的建議、支持和情緒價值，幫我度過許多難關。還有 R12 的大家，成大朋朋郁潔、會打球有腹肌又好笑的辛哥 (奕呈)、坐對面的加班好夥伴晟祐、機器學習大師彥翔、鼓勵滿滿的永詮、會煮飯又會打球的津津、會打球又會去看展的宥安、唱歌超好聽的團康股長大元、很靦腆又很好笑的李昺、肩膀很寬講話很溫柔的丫祐 (承祐)、跑步好夥伴靖瑀、好笑又可愛的陳心、皮膚白皙做事果斷的昕儀、軟網女王兼新店與花蓮娛樂空間的提供者琬潔，還有葛門的學弟妹和振動台好夥伴均侑、做事有效率的詒茹、在日本讀博的岡野翔大、PLAXIS 和機器學習的好夥伴小汪 (志穎)、品學兼優的琮淯和同桌一起共患難的靜穎，感謝各位在這兩年帶我體驗了好多，包括小巨蛋溜冰、排球組際盃 (*2)、琮淯家跨年 (*2)、六福村、過年的彰化、嘉義和台中遊、花蓮畢旅加慶生、走路環島、兒童新樂園、擎天崗等，以及各式各樣的運動，包含羽球、網球、排球、射箭、游泳，還有路跑 (曾經一次 10k 不間斷)，甚至是每次走出學校吃飯 (公館或 118) 的聊天時光，這些美好的回憶我會心懷感激的收藏好放在心底。

最後，想特別感謝澄毓和煥睿，在神經緊繃的碩二下學期，不分平日假日地在研究室的三個小角落與自己的研究奮戰到隔天凌晨，再以各種小故事和甜點療癒並陪伴著彼此，最後在一起於系館側門舉著大拇指說「你很棒!掰掰，明天見~」，有時還會執行我們的各種瘋狂想法騎著機車馳騁在大台北的各個角落。這或許是整個碩士生涯中最痛苦卻也是最快樂的一段回憶，謝謝你們的鼓勵與陪伴，讓我能以健康的心態走到最後，完成這段旅程。

蘇筱丰 2025/07/13

doi:10.6342/NTU202502436

摘要

土壤液化一直是大地工程領域中長期關注的重要議題，而動態三軸試驗 (Cyclic Triaxial Test) 作為實驗室中模擬地震載重下，研究土壤行為主要的土壤單元試驗。其中，作為施加荷載能量的關鍵參數的循環剪應力比 (Cyclic Stress Ratio, CSR) ，其設定往往依賴試體的物理性質、試驗條件和試驗者的經驗進行推估。然而，不同參數組合將導致 CSR 值有所變化，使得在試驗設計階段難以事先準確掌握合適的 CSR 值，進而影響試驗成功率與數據品質。為解決此問題，本研究蒐集並整理來自 23 篇文獻，共近千筆的動態三軸試驗資料，建立多種機器學習模型以協助預測不同試體參數所對應之 CSR 值，希望提供一套可用應於試驗設計階段之預測工具。研究流程分為三階段，第一階段針對資料進行前處理與特徵工程，包含缺失值填補、數值變數標準化與類別變數做編碼等；第二階段使用七種監督式學習方法，包含五種傳統機器學習模型 (Random Forest、CatBoost、XGBoost、Explainable Boosting Machine (EBM)，與 Support Vector Machine (SVM)) 以及兩種深度學習模型 (Artificial Neural Network (ANN) 與 Bayesian Neural Network (BNN))，並使用決定係數 R^2 和誤差指標 (如 MAE、MSE 與 RMSE) 做為迴歸預測表現之評估依據；第三階段則進一步 CSR 與循環剪切次數 (N) 之關係建立下限曲線作為二元分類基準，將迴歸問題轉化為液化與非液化樣本之分類任務，並加入特徵解釋性與預測不確定性分析。研究結果顯示，CatBoost、XGBoost 與 Random Forest 整體表現最佳。主要特徵 (如初始孔隙比、循環剪切次數與細粒料含量等) 即可支撐預測能力。MICE 為最穩定之補值方式。於深度學習模型中，手動調參表現優於 Grid search，可有效避免測試集表現劣化的問題。EBM 模型之變數貢獻圖可提供解釋性，BNN 模型則能提供預測結果的信賴區間，能量化不確定性。

關鍵字: 土壤液化、動態三軸試驗、循環剪應力比、機器學習、深度學習

doi:10.6342/NTU202502436

Abstract

Soil liquefaction has long been a critical issue in geotechnical engineering. Among the laboratory methods developed to investigate this phenomenon, the cyclic triaxial test is one of the most important experiments for simulating soil behavior under earthquake loading. A key parameter in this test is the Cyclic Stress Ratio (CSR), which reflects the magnitude of loading energy applied to the soil specimen. However, the appropriate CSR value often depends on the specimen's physical properties, the experimental conditions, and the experience of the test operator. As a result, accurately determining the CSR beforehand is challenging, which may affect the success rate of the test and the quality of the resulting data.

To address the problem, this study collected nearly 1000 cyclic triaxial test records from 23 academic papers and theses, and establish multiple machine learning models to predict the target variable –CSR, aiming to provide a useful reference during the test design stage. The study is divided into three main parts. The first part involves data preprocessing and feature engineering, including missing value imputation, numerical variable standardization, and categorical variable encoding. The second part focuses on model development using seven supervised models, including five machine learning models (Random Forest, CatBoost、XGBoost、Explainable Boosting Machine (EBM), and Support Vector Machine (SVM)) and two deep learning models (Artificial Neural Network (ANN) and Bayesian Neural Network (BNN)). Model performance was evaluated using the coefficient of determination (R^2) and error metrics. In the third stage, the study further converted the regression task into a binary classification problem by defining a trial-and-error lower bound curve based on the CSR-*N* relationship to distinguish between liquefied and non-liquefied samples. This stage also incorporated model interpretability and prediction uncertainty analysis. The results show that

doi:10.6342/NTU202502436

CatBoost, XGBoost, and Random Forest achieved the best overall performance. Moreover, the models were able to make accurate predictions using only a set of primary features (e.g., initial void ratio e_0, number of loading cycles N, and fine content f_c). Among the imputation methods tested, Multiple Imputation by Chained Equations (MICE) produced the most stable outcomes. Manual hyperparameter tuning outperformed Grid search techniques in deep learning models. Finally, EBM provided interpretable visualizations of individual feature contributions, while BNN offered prediction intervals that helped quantify uncertainty and enhance confidence in the results.

Key words: Soil liquefaction, Cyclic triaxial test, Cyclic stress ratio (CSR), Machine learning, Deep learning.

doi:10.6342/NTU202502436

目次

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

圖次

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

表次

doi:10.6342/NTU202502436

doi:10.6342/NTU202502436

第一章 緒論

1.1 研究背景與目的

在動態三軸試驗中循環剪應力比 (Cyclic Stress Ratio, CSR) 為評估土壤抗液化潛能之重要參數。然而，於實務操作上，CSR 並非一組可直接量測之固定值，而是須由試驗者根據試體之物理性質、試驗條件與過往經驗推估設定。由於試體在尚未施加剪力前，其抗剪強度具有不確定性，使得試驗中所設定之 CSR 值難以精準掌握。若設定過大，試體可能於初期循環荷載下即發生液化，導致試驗過程過早終止，循環剪切次數無法完整記錄；若設定過小，試體則可能無法達到液化狀態，試驗時間拉長且無有效結果。此不確定性不僅影響數據品質，更可能導致資源與時間的浪費，對於試驗效率與成本皆造成負擔。此外，動態三軸試體之準備過程繁複，從配比、填砂、試體飽和至壓密皆需不少人力與時間。若於最終階段的反覆剪切步驟因 CSR 設定不當而失敗，需重新配製試體並操作試驗，將降低整體試驗效率。因此，若能於試驗前提供一套具有參考價值的 CSR 預測方式，將有助於提升試驗成功率與數據品質，亦可作為經驗法則以外之補充參考依據。

為解決上述問題，本研究提出以機器學習模型進行預測 CSR 之方法，藉由整理過去文獻中之試驗資料，訓練模型學習不同試體條件與 CSR 之間的非線性關係，期許能於試驗設計階段提供較具依據性的參考建議。此外，透過模型所建構之特徵重要性排序與解釋性分析，亦可輔助試驗者理解各項物理與試驗參數對 CSR 的影響程度，作為工程應用與研究判讀之輔助依據。

本研究所建構之完整分析流程，包括資料處理、模型比較、解釋性分析與不確定性評估，具高度彈性與應用延伸性。除可應用於動態三軸試驗之 CSR 預測，亦具備延伸至其他土壤工程相關問題之潛力，有利於拓展機器學習在大地工程領域中之應用廣度。

doi:10.6342/NTU202502436

1.2 研究方法

本研究主要在建立一套具多種模型比較與解釋能力之分析架構，以協助理解動態三軸試驗資料中循環剪應力比 (Cyclic Stress Ratio, CSR) 與土壤液化行為之間的潛在關聯性。整體研究方法共分為三大階段，分別為資料蒐集與前處理、模型建構與表現評估、進階分析與應用探討，各階段內容說明如下。

第一階段為資料準備與特徵處理。本研究所使用之數據來自 23 篇已發表之文獻及學位論文所紀錄之動態三軸試驗資料，總計 997 筆試驗紀錄，涵蓋細粒料含量與相對密度等常見土壤物理性質與試驗控制變數。為提高模型學習效率並達到資訊最大化之利用，資料處理方面，本研究首先依特徵缺失率與工程意義進行初步分類，採取適當方法補齊資料缺漏，並針對類別與數值型變數進行一致性調整與轉換，以確保後續模型輸入格式之統一性與可比較性。

第二階段為模型建立與預測分析。為驗證不同模型對 CSR 預測之適用性與穩定性，本研究採用七種監督式學習模型，涵蓋機器學習模型 (Random Forest、CatBoost、XGBoost、EBM 與 SVM) 與深度神經網路 (ANN 與 BNN)。各模型皆使用相同資料組合，並於訓練前依固定比例分割為訓練集與測試集作為輸入資料。在模型訓練完成後，統一以決定係數 R^2 與誤差指標作為迴歸表現評估依據。

第三階段則為進階分析與模型應用探討。此階段加入分類分析，透過自行擬定之 CSR-*N* 下限曲線作為判別界線，將迴歸輸出轉換為液化與否之分類結果，並以正確率 (Accuracy)、精準率 (Precision)、召回率 (Recall) 與 F1 分數 (F1 score) 進行評分。同時，本研究亦於 EBM 模型中加入視覺化與解釋性分析，包含 SHAP 總結圖與單變數貢獻函數曲線等，此外，對於 BNN 模型則是以參數分佈推論建立預測之不確定性，提供模型於實務應用中信心量化之依據。

doi:10.6342/NTU202502436

1.3 論文架構

本研究內容共分為六章，各章節內容如下

第一章 緒論

說明研究之背景與動機、研究目的與研究方法。

第二章 文獻回顧

首先說明土壤液化發生原因與循環剪應力比 (CSR) 之意義，並介紹動態三軸試驗中試體準備流程、液化判定方式與 CSR-*N* 曲線之應用。其次，整理常見機器學習與深度學習方法，說明其基本架構與訓練原理，並探討其於大地工程領域中之應用發展，作為後續建模分析之理論依據。

第三章 研究方法

說明本研究所蒐集之試驗資料來源、前處理方式與特徵工程內容，並依據不同分析目標建立迴歸與分類模型，說明各類機器學習與深度學習演算法之架構。

第四章 模型表現與預測結果比較

模型預測結果與表現比較，呈現各模型在不同調參方式、特徵組合、缺失值填補方式下之迴歸表現，並透過多項評估指標進行系統性比較分析。

第五章 EBM 與 BNN 於液化數據之應用

聚焦於兩種具代表性之模型 (EBM 與 BNN) 進行深入探討，分別就其迴歸預測結果、分類表現、特徵解釋性與不確定性評估等層面進行說明與比較。

第六章 結論與建議

統整本研究之主要發現，並根據分析結果提出後續研究可努力之方向建議。

doi:10.6342/NTU202502436

第二章　文獻回顧

2.1 液化行為、循環剪應力比與動力三軸試驗

本節旨在介紹土壤液化之基本機制與循環剪應力比 (Cyclic Stress Ratio, CSR) 的理論基礎。首先將說明液化發生的條件與種類、CSR 之定義，以及在動態三軸試驗中的 CSR 值，並說明 CSR 與剪切循環次數 N 之關係。

2.1.1 液化機制

土壤液化 (soil liquefaction) 是指在地震作用下，剪力波於土層中向上傳播，如圖 2.1，飽和無凝聚力土壤 (saturated cohesionless soils) 顆粒原本的排列重新組合為更緊密的狀態，排水來不及發生，孔隙水壓承受增加的體積壓力而導致超額孔隙水壓上升。一旦超額孔隙水壓升高至接近或等於總應力，有效應力便趨近於零，導致土壤喪失乘載能力而呈現類似流體的狀態。(Seed, 1976; Obermeier, 1996)

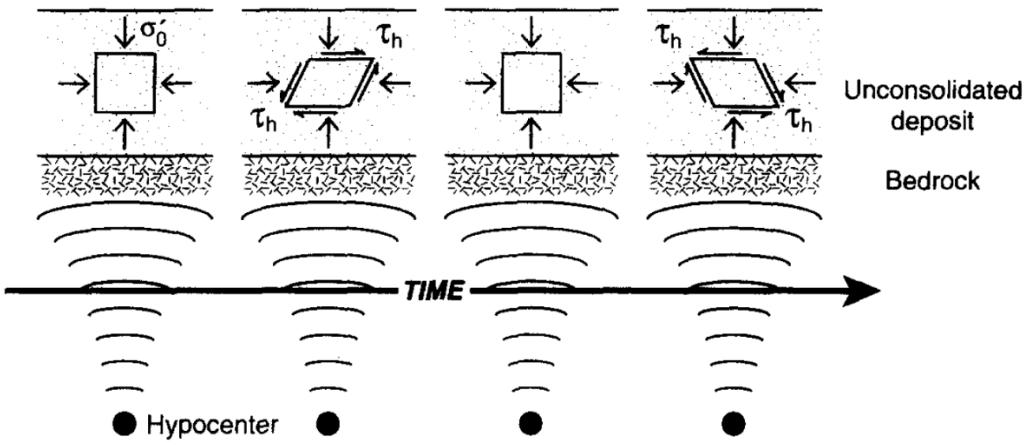


圖 2.1 土層受反覆剪應力歷時示意圖 (Obermeier, 1996)

doi:10.6342/NTU202502436

土壤液化相關的破壞機制分成以下三種類型 (Seed, 1976; Pan et al., 2019; Pan et al., 2020; Wei et al., 2023; Zhou et al., 2023):

1. 流動液化 (Flow liquefaction)

流動液化為液化破壞機制中最嚴重的一種類型，常發生在鬆散且低密度的飽和砂土中。在地震等動態載重 (或靜態載重改變) 作用下，土體顆粒重組 (rearrangement) 而發生類似於壓密的行為，孔隙水壓迅速上升至等於或接近有效應力，使得有效應力幾乎歸零，導致土體完全喪失抗剪強度，進而產生突然的塑性流動 (plastic flow)，類似滑動或崩塌。

在含細粒料之砂土中，若細粒料含量較低 ($f_c \approx$ 5~15%)，顆粒骨架仍以砂為主，且初始剪應力偏低或不存在時，試體仍可能會快速累積孔隙水壓力並失去穩定性，出現類似純砂之流動型破壞。當 f_c 提升至中高範圍，細粒會干擾顆粒間接觸與排水性，使得孔壓上升速率降低，導致破壞時間推遲、程度減緩，甚至轉變為其他破壞形式。Zhou et al. (2023) 指出，隨著初始剪應力增加，流動破壞需更高之能量輸入才能達成。

2. 反覆流動 (Cyclic mobility)

反覆流動是發生於中密至密實砂土，或含中等細粒料 ($f_c \approx$ 15~30%) 之砂土中的一種液化型態。與流動液化不同，此類破壞不會導致土體立即喪失全部剪力強度，而是經歷多次剪應力循環後，逐漸累積孔隙水壓與不可逆變形。在地震或類似的週期載重下，土體內部孔隙水壓逐步升高，但因其結構相對緊密而發生剪脹 (dilatancy)，使得每次循環後有效應力略為回升，變形受到一定程度的抑制。此過程中，土體會出現顯著的週期性剪應變峰值與殘餘變形，剪應力–應變路徑呈現穩定擴展的循環模式，如圖 2.2 所示。其能量累積模式呈階梯式上升，反映每一次循環僅轉化部分剪力能量，尚未達到臨界破壞能量的釋放條件。此一行為具有明顯的週期性與漸進性，介於典型流動液化與殘餘應變累積破壞之間，屬於過渡型態。

doi:10.6342/NTU202502436

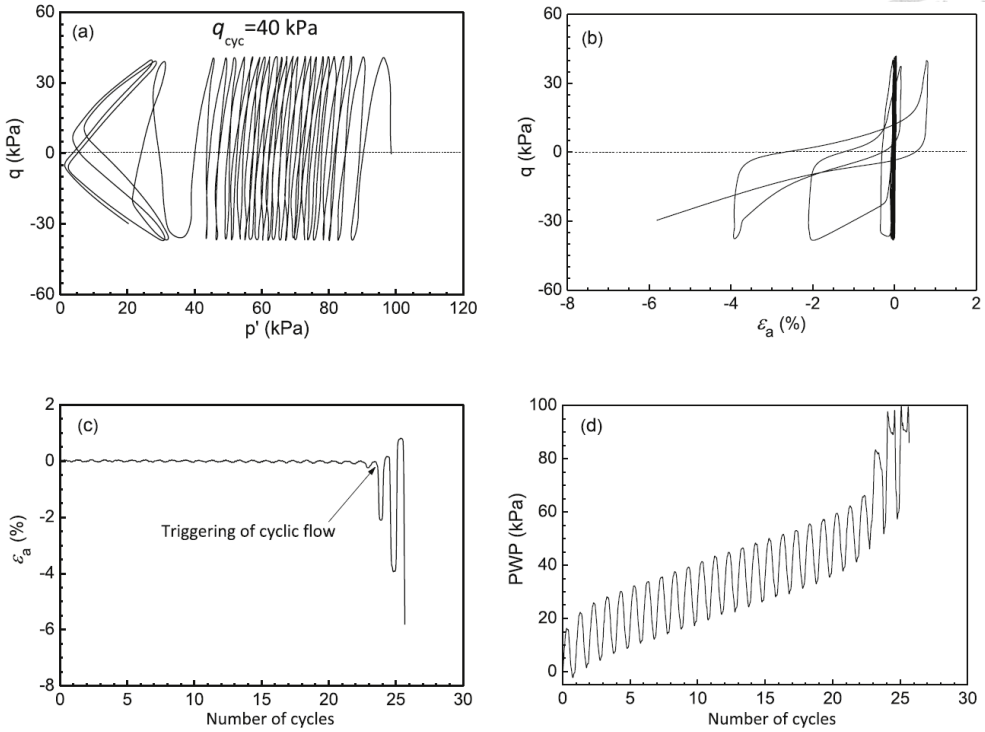


圖 2.2 砂土在單向度壓密下之循環反應 (Pan et al., 2019)

3. 殘餘變形累積 (Residual deformation accumulation)

殘餘變形累積是一種進行緩慢且穩定的漸進型液化相關破壞型態，主要發生於未經剪力反轉 (stress reversal) 之情境下，如圖 2.3，即初始靜態剪應力（static shear stress）與循環剪應力方向一致。於此機制下，土體雖不會瞬間失去全部抗剪強度，但在每次剪應力循環中皆產生小幅度、不可逆的塑性變形，孔隙水壓亦逐步升高，但未達液化臨界值。有效應力路徑通常維持在單側穩定變化，長期累積則可能導致顯著的軸向或側向變形。

Wei et al. (2023) 指出，當初始剪應力存在時，含細粒料砂的破壞模式會由流動液化轉為殘餘應變累積。該研究稱此行為為「延遲破壞模式 (delayed failure)」，由於細粒提升了土體的排水阻力，抑制孔隙水壓迅速發展，致使試體雖歷經多次循環亦不立即破壞，但卻逐步產生不可逆的殘餘變形。此外，研究強調，若分析中未考慮初始剪應力與細粒含量的交互效應，可能會高估土體穩定性與抗液化能力

doi:10.6342/NTU202502436

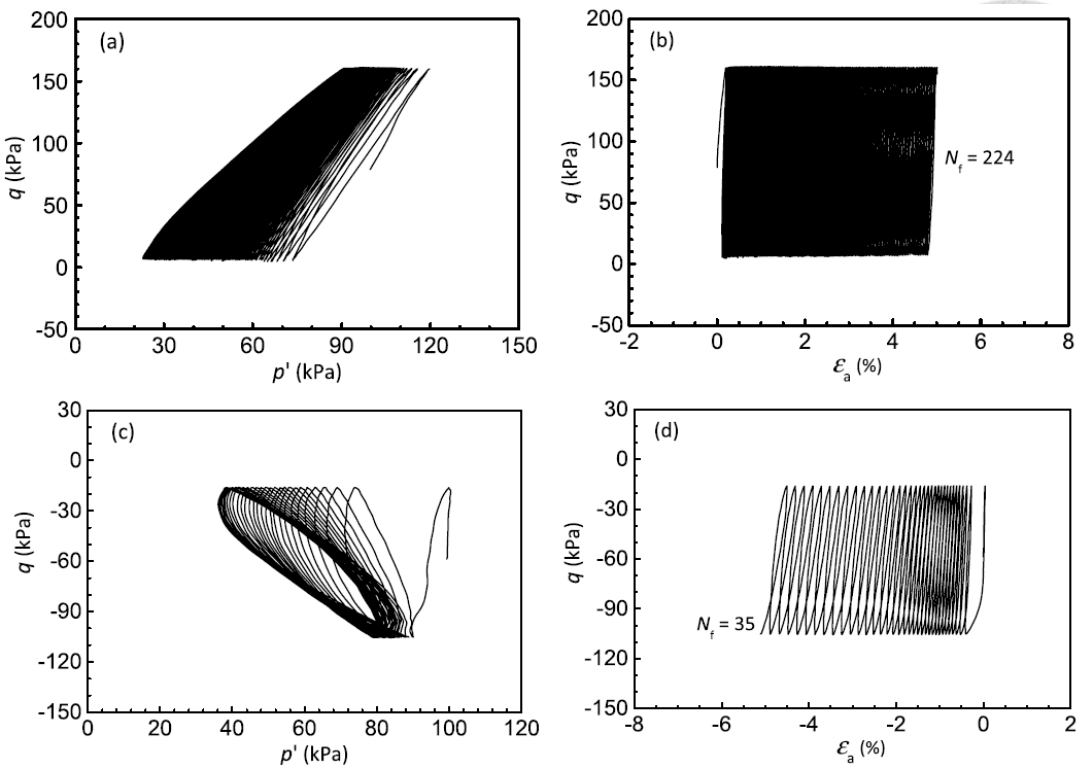


圖 2.3 含細粒料砂於「無應力反轉」條件下之有效應力路徑與應力–應變曲線 (Pan et al., 2020)

此外，針對黏土層在動態載重下的強度弱化現象，Tonyalı et al. (2024) 指出，反覆軟化 (Cyclic softening) 是一種常見於飽和黏性土壤 (如低塑性黏土) 中的破壞機制，本質為土壤在反覆剪應力作用下，剛性 (stiffness) 與強度 (strength) 逐漸退化的過程。與典型砂土液化不同，反覆軟化不一定伴隨有效應力完全消失，但會導致土體持續軟化、承載力下降與地盤過量變形等，進而引發基礎傾斜或沉陷。此破壞機制雖不劇烈，但在高層建築或偏軟地層中，仍可能造成顯著沉陷與安全風險。

doi:10.6342/NTU202502436

2.1.2 循環剪應力比 (CSR) 之定義與於液化評估中的角色

循環剪應力比 (Cyclic Stress Ratio, CSR) 是用以表示地震或受震期間，土層所受到的剪應力需求 (demand)。其基本定義為地震或受振期間土層中產生的最大剪應力與初始有效覆土壓力之比值，表示如下 (Seed & Idriss, 1971) :

$$CSR = \frac{\tau_{av}}{\sigma'_{v0}} = 0.65 \cdot \frac{a_{max}}{g} \cdot \frac{\sigma_{v0}}{\sigma'_{v0}} \cdot r_d \tag{2.1}$$

其中，

τ_{av}: 地震作用下之平均剪應力

a_{max}: 地震引致的，於地表之最大 (peak) 水平加速度

g: 重力加速度

σ_{v0}: 覆土總壓力

σ'_{v0}: 有較覆土壓力

r_d: 剪應力衰減係數 (stress reduction coefficient)，反映地震減力隨深度而減弱的效果

此公式源自 Seed and Idriss (1971) 所提出之簡化法 (simplified procedure)，並在後續多篇研究中逐步修正與優化。根據 Youd et al. (2001) 中 NCEER 工作坊的共識建議，該公式仍為工程實務中常使用的 CSR 估算方法，適用於離地表約 15 公尺深的飽和砂質地層。

CSR 為評估液化的重要參數，與循環阻抗比 (Cyclic Resistance Ratio, CRR) 相對應，兩者之比值用來計算液化安全係數 (Factor of safety, FS) :

$$FS = \frac{CRR}{CSR} \tag{2.2}$$

當$FS < 1$時，即表示該土層在地震期間可能發生液化現象。為提升精度，CSR 計算中常考慮修正因子，如地震規模調整係數 (Magnitude scaling factors, MSFs) 與剪應力修正係數 (correction factor for soil layers subjected to large static shear

doi:10.6342/NTU202502436

stresses, K_α) 等，可進一步用於評估各類地層在不同地震環境下的液化潛勢。

整體而言，CSR 提供了對地震載重下土壤行為的量化方式，是連結地震輸入與土壤反映的核心參數，亦為後續動態試驗、曲線分析與液化評估模型建構的重要依據。

在動態三軸試驗中，CSR 常用於表示施加於試驗之剪應力比，反映外力對土體造成之循環載重程度，而循環阻抗比 (Cyclic Resistance Ratio, CRR) 則為土體抵抗液化的能力指標，通常需藉由試驗結果來判定。由於本研究蒐集之試驗資料均以 CSR 值作為輸入試驗之參數，故本研究以該 CSR 值作為目標變數進行預測。雖名為 CSR，實際上在概念更接近於 CRR，亦即代表土體在特定條件下的抗剪強度。因此，後續文中所稱目標 CSR，可視為樣本的「等效 CRR」，兩者意義雖不同，於本研究中應予以區分理解，避免混淆。

2.1.3 液化判定與 CSR 分析

液化判定準則

動態三軸試驗 (Cyclic Triaxial Test) 之主要目的在於透過試驗控制與條件設定，模擬飽和土體在地震等動態荷載作用下於不排水 (undrained) 狀態中所產生之超額孔隙水壓上升、有效應力降低、應變發展與最終破壞行為，藉以獲得試體是否發生液化之關鍵參數，並量化其抗液化能力，而其判定液化發生的標準主要包括:

1. 超額孔隙水壓比達 1 ($r_u = \frac{\Delta u}{\sigma'_{vo}} = 1$)，代表有效應力趨近於零
2. 雙震幅軸向應變達 5% (double-amplitude axial strain=5%)
3. 有時亦參考最大剪應變 $\gamma_{max} \approx 3.75\%$ 或累積塑性應變作為輔助，如圖 2.5 所示

doi:10.6342/NTU202502436

上述標準均反映土體已失去抗剪能力，進入接近流動狀態的液化條件 (Tsukamoto et al., 2004)。

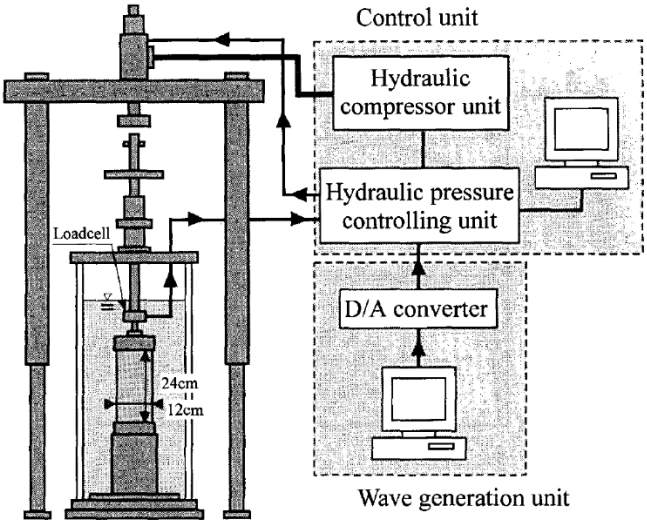


圖 2.4 動態三軸試驗設備配置示意圖 (Tsukamoto et al., 2004)

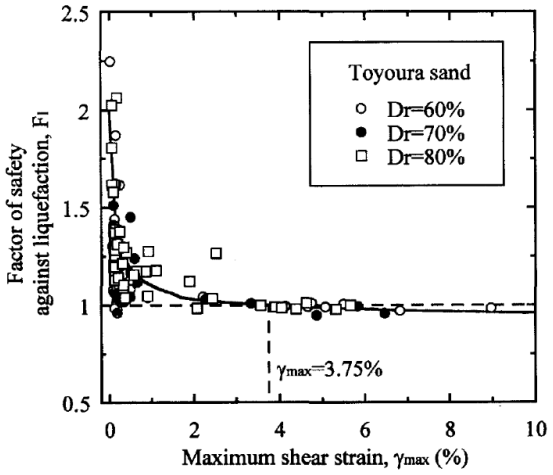


圖 2.5 抗液化安全係數（F_l）與最大體積應變（γ_{max}）總結圖 (Tsukamoto et al., 2004)

CSR 與 CSR-*N* 曲線

循環剪應力比 (Cyclic Stress Ratio, CSR) 為動態三軸試驗中用以量化載重的試驗設定參數，可用來模擬不同地震條件下土層實際承受的剪應力，試驗中常設定不同 CSR 值 (如 0.15、0.2 或 0.25 等)，以觀察其對試體在動態剪切過程中的影響。

doi:10.6342/NTU202502436

在不同 CSR 條件下，試體所能承受的剪切循環次數 (N) 可繪製成 CSR-N 曲線，如圖 2.6 所示，作為評估土壤抗液化能力之依據。該曲線一般呈反比趨勢，顯示 CSR 越高時，試體達液化所需之循環次數越少。實務應用中，常以循環剪切次數 $N = 15$ (Papadopoulou & Tika, 2008; Stamatopoulos, 2010; Tsai et al., 2010; Xu et al., 2024) 或 $N = 20$ (Pan et al., 2020; Xu et al., 2024) 為參考基準，取其對應之 CSR 值定義為該土壤試體之抗液化能力 (Cyclic resistance ratio, $CRR_{N=15}$ 或 $CRR_{N=20}$)。

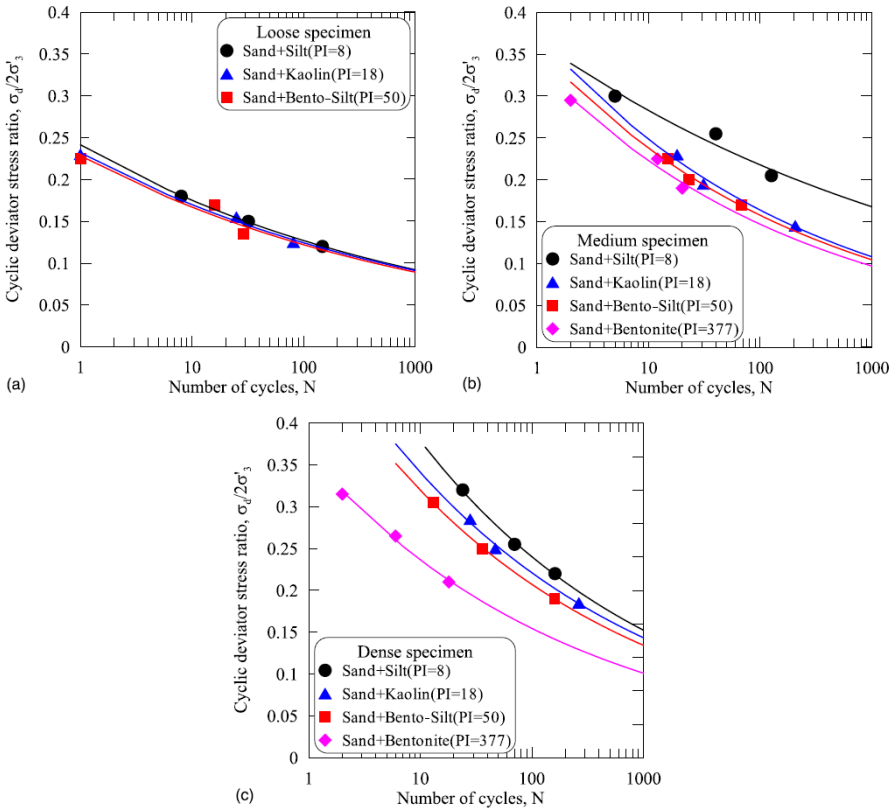


圖 2.6 不同密度砂之抗液化曲線 (Park & Kim, 2013)

此外，CSR-N 曲線可反映不同因素對液化行為之影響。其中，試體相對密度 (D_r) 越高，曲線將向右偏移，表示土體需承受更多次循環剪力才能達到液化，具有較高的抗液化能力 (Tsai et al., 2010)。另外，於相同 CSR 條件下，非塑性細粒料 (non-plastic fines) 會降低試體抗剪強度並增加變形；反之，具塑性之細粒料 (plastic fines) 則有助於提升試體之抗液化能力 (Park & Kim, 2013)。不同的試體

doi:10.6342/NTU202502436

準備方式 (如乾式沉積、霣降法) 所形成之土體顆粒排列、接觸力與孔隙分布也會導致相同 CSR 值下，試體對剪力出現顯著反應差異。Ishihara et al. (1980) 指出以漿體壓密 (slurry consolidation) 準備之試體，其孔隙結構較均勻，抗液化能力較高，反映於 CSR-*N* 曲線呈現較緩的斜率。CSR-*N* 曲線不僅為土壤液化評估的工具，更能呈現不同材料性質所造成的行為差異。

2.2 機器學習

隨著資料科學與運算技術的發展，機器學習 (Machine Learning) 已逐漸成為土木與大地工程領域中重要的分析工具之一。在液化潛能評估等議題中，機器學習能透過大量試驗或場址資料，自動學習輸入變數與目標變數間之非線性關係，並建立具預測力之模型。

2.2.1 機器學習種類

本節將接續介紹機器學習之基礎觀念，以及其於大地工程領域中的應用案例與發展趨勢。

機器學習方法

機器學習主要有監督式學習 (Supervised Learning)、非監督式學習 (Unsupervised Learning) 與半監督式學習 (Semi-supervised Learning) 等不同方法，以下分別說明 (van Engelen & Hoos, 2020; Jafari-Marandi, 2021; Rani et al., 2023)。

監督式學習 (Supervised Learning) 為機器學習中最常見之類型，其基本概念為透過含標籤 (label) 之輸入變數與目標變數之間的對應關係。模型的學習目標是最小化預測值與真實值之間的誤差 (residual) ，進而達到迴歸 (regression) 或

doi:10.6342/NTU202502436

分類 (classification) 的目的。在實務應用中，監督式學習常被用於據明確定義之問題情境，例如疾病診斷、災害發生與否之預測等。

相較於監督式學習，非監督式學習 (Unsupervised Learning) 則不需資料標籤，僅使用未標記之資料進行識別。其主要目的是尋找資料中潛藏的結構與分部特性，例如資料分群 (clustering)、特徵提取 (feature extraction) 與降維 (dimensionality reduction) 等。非監督式學習更適合應用於缺乏明確學習目標，或需模型自行分析資料的情境，例如地質剖面分類或材料性質歸類等。

半監督式學習 (Semi-supervised Learning) 介於監督式與非監督式學習之間，其基本理念為利用少量含標籤資料與大量無標籤資料共同參與模型訓練，以提升在標籤資料有限的情境下，模型的預測能力。此類方法在標註成本高昂或資料龐大但缺乏人工標籤的領域中具有實用價值，如場址試驗資料或歷史震災資料分析。

機器學習 (Machine Learning) 與深度學習 (Deep Learning)

機器學習 (Machine Learning) 為人工智慧領域中的一項核心技術，其主要目的是透過資料訓練模型，從中學習輸入變數與輸出目標之間的規律，進而進行預測或分類。傳統機器學習方法，如決策樹 (Decision Tree, DT) 或隨機森林 (Random Forest, RF) 等，通常需進行特徵工程 (feature engineering)，由人工選取或轉換輸入資料中具判別力的特徵，以提升模型表現。

相較之下，深度學習 (Deep Learning) 則是機器學習的一種延伸，其基礎為類神經網路 (Neural Networks)，並透過多層非線性結構進行端到端 (end-to-end) 的學習。深度學習模型可自動從原始資料中學習多層次之抽象特徵，尤其適用於影像、語音、自然語言等高維資料的建模分析。與機器學習相比，深度學習具備更強的特徵表達能力與模型靈活性，但同時也伴隨著更高的資料需求與運算資源負擔。

doi:10.6342/NTU202502436

Janiesch et al. (2021) 指出，深度學習在處理結構化與非結構化資料中皆展現出優異的表現，但在可解釋性、訓練穩定性與參數調整方面，對使用者來說具一定挑戰性。因此，機器學習與深度學習在應用選擇上具有互補性，須視資料特性與問題需求靈活選用。

梯度 (Gradient) 與梯度下降法 (Gradient Descent)

在機器學習模型的訓練過程中，損失函數 (Loss function) 是用以衡量模型預測值與實際值之間誤差的重要指標，常見形式包括均方誤差 (Mean Squared Error, MSE) 與交叉熵損失 (Cross-Entropy Loss) 等。為了提升模型預測能力，訓練過程需透過演算法不斷調整模型參數，以最小化損失函數值。

此一優化過程的核心便是梯度 (gradient)。梯度代表損失函數對模型參數的偏導數，其方向指出當前參數下損失函數上升最快的方向，而梯度下降法 (Gradient Descent) 則反方向沿該方向調整參數，以逼近損失最小值。最基本的梯度更新形式可表達如下 (Monego et al., 2022) :

$$\theta \coloneqq \theta - \eta \cdot \nabla_\theta L(\theta) \tag{2.3}$$

其中，θ 為模型參數，η 為學習率 (learning rate)，而 $\nabla_\theta L(\theta)$ 為損失函數的梯度。此方法廣泛應用於類神經網路訓練中，並結合反向傳播 (Backpropagation) 進行多層權重的遞迴更新。

然而，Shakir Mohamed (2020) 指出，在某些模型架構中，如隱變數模型 (Latent variable model)、生成模型 (Generative model) 或貝葉斯神經網路 (Bayesian Neural Network, BNN) 等，因損失函數之期望 (expectation) 運算往往無法以解析型式 (closed form) 表示，梯度難以求解，需透過蒙地卡羅梯度估計 (Monte Carlo Gradient Estimation) 進行近似。此方法利用隨機抽樣對期望值進行估算，並保留對原始梯度方向的期望性質，進而為值模型的收斂穩定性與泛化能

doi:10.6342/NTU202502436

力。本研究於 BNN 模型即採用蒙地卡羅梯度估計進行目標值之預測。

另外，梯度提升決策樹 (Gradient Boosted Decision Trees, GBDT) 是一種基於決策樹的機器學習集成方法，以加法模型的方式逐步結合，每次學習一個新的樹(通常為淺層決策樹)，用來預測前一次模型的殘差，從而逐步逼近整體預測目標。這一學習過程可視為在函數空間中進行梯度下降，每一棵樹都近似地對應於損失函數梯度的負方向 (Nguegnang et al., 2024)。本研究所使用的分類增強模型 (Categorical Boosting, CatBoost)、極端梯度提升機 (Extreme Gradient Boosting, XGBoost) 與可解釋增強模型 (Explainable Boosting Machine, EBM) 均屬於 GBDT 的延伸模型。

傳播 (Propagation)

在人工神經網路 (Artificial Neural Network, ANN) 中，傳播 (propagation) 指的是輸入資料與誤差訊息在網路各層間的傳遞過程，為模型進行學習與預測的基礎。傳播機制包含兩個主要階段，前向傳播 (forward propagation) 與反向傳播 (backpropagation)，兩者方向相反，卻共同構成神經網路之訓練核心 (Rumelhart et al., 1986; Xinghuo et al., 2002; Balamurugan et al., 2022)。

前向傳播 (Forward propagation) 負責計算神經網路的預測輸出。在每一層中，輸入向量 $\boldsymbol{z}$ 經由加權運算與非線性激活函數 $f(\cdot)$ 處理後，生產出 $\boldsymbol{a}$ ，其計算公式如下:

$$\boldsymbol{z}^{(l)} = \boldsymbol{W}^{(l)}\boldsymbol{a}^{(l)} + \boldsymbol{b}^{(l)}, \boldsymbol{a}^{(l)} = f(\boldsymbol{z}^{(l)}) \tag{2.4}$$

其中，$\boldsymbol{W}^{(l)}$為第 l 層的權重矩陣，$\boldsymbol{b}^{(l)}$為偏差項 (bias)，$\boldsymbol{a}^{(l)}$為該層輸出。此流程會持續進行，模型會對輸入資料進行特徵提取與轉換，但並不改變權重，其目的是建立輸入與輸出之映射關係，直到輸出層產生預測結果 $\hat{\boldsymbol{y}}$。

反向傳播 (Backpropagation) 則是根據預測值 $\hat{\boldsymbol{y}}$ 與真實值 $\boldsymbol{y}$ 所產生的損失

doi:10.6342/NTU202502436

函數 $L(\boldsymbol{y},\hat{\boldsymbol{y}})$，利用鏈式法則 (chain rule) 逐層計算損失對各層權重與偏差的導數(梯度)，並透過優化演算法進行反向參數更新，使預測誤差逐步收斂。以單一權重的更新公式為例，其梯度計算公式如下:

$$\frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} \cdot \left(\boldsymbol{a}^{(l-1)}\right)^T \; with \; \delta^{(l)} = \frac{\partial L}{\partial \boldsymbol{z}^{(l)}} \tag{2.5}$$

其中，$\boldsymbol{a}^{(l-1)}$ 為第 $(l-1)$ 層的激活輸出，$\delta^{(l)}$為第l層的誤差向，表示損失對線性組合輸出 $\boldsymbol{z}^{(l)}$ 的偏導數。

Balamurugan et al. (2022) 進一步說明，前向傳播可視為模型預測流程中的「辨識」階段，用以完成特徵提取與輸出運算；而當模型預測產生偏差時，便須使用反向傳播來調整模型參數以達到「學習」效果。

2.2.2 機器學習方法於大地工程之應用

近年來，隨著地工相關數據的增加與模型建立技術的成熟，機器學習方法逐漸被應用在大地工程領域的不同議題中，作為輔助分析與預測的工具。相較於傳統經驗式或數值模擬分析，機器學習能透過大量資料進行模型訓練，建立輸入變數與結果之間的非線性關係，並提升在複雜條件下的預測效率與彈性。

本節將介紹機器學習於大地工程中三個主題的應用，分別為邊坡破壞預測、地盤沉陷預測與液化潛能評估，並說明其在各主題中的應用方式。

邊坡破壞預測

在邊坡分析中，隨機有限元素法 (Random Finite Element Method, RFEM) 能結合隨機場模型與有限元素分析，模擬土壤性質之空間變異性，進而評估破壞機率。然而，透過 RFEM 執行大量蒙地卡羅模擬以獲得可靠的破壞機率，計算成本極為龐大。為提升效率，Hsiao et al. (2022) 結合機器學習方法與 RFEM 模擬資料，建立可快速預測安全係數 (Factor of Safety, FOS) 與邊坡破壞滑動面之模型。該

doi:10.6342/NTU202502436

研究比較人工神經網路 (ANN) 與卷積神經網路 (Convolution Neural Network, CNN) 於處理隨機場資料的效能，結果顯示 CNN 模型能更有效捕捉輸入資料的空間關係，於此複雜問題中表現優於 ANN 模型。此外，CNN 在訓練樣本數量增加時，其預測表現顯著提升，顯示 CNN 對高維度與空間相關性強的資料具有更佳的學習潛力。

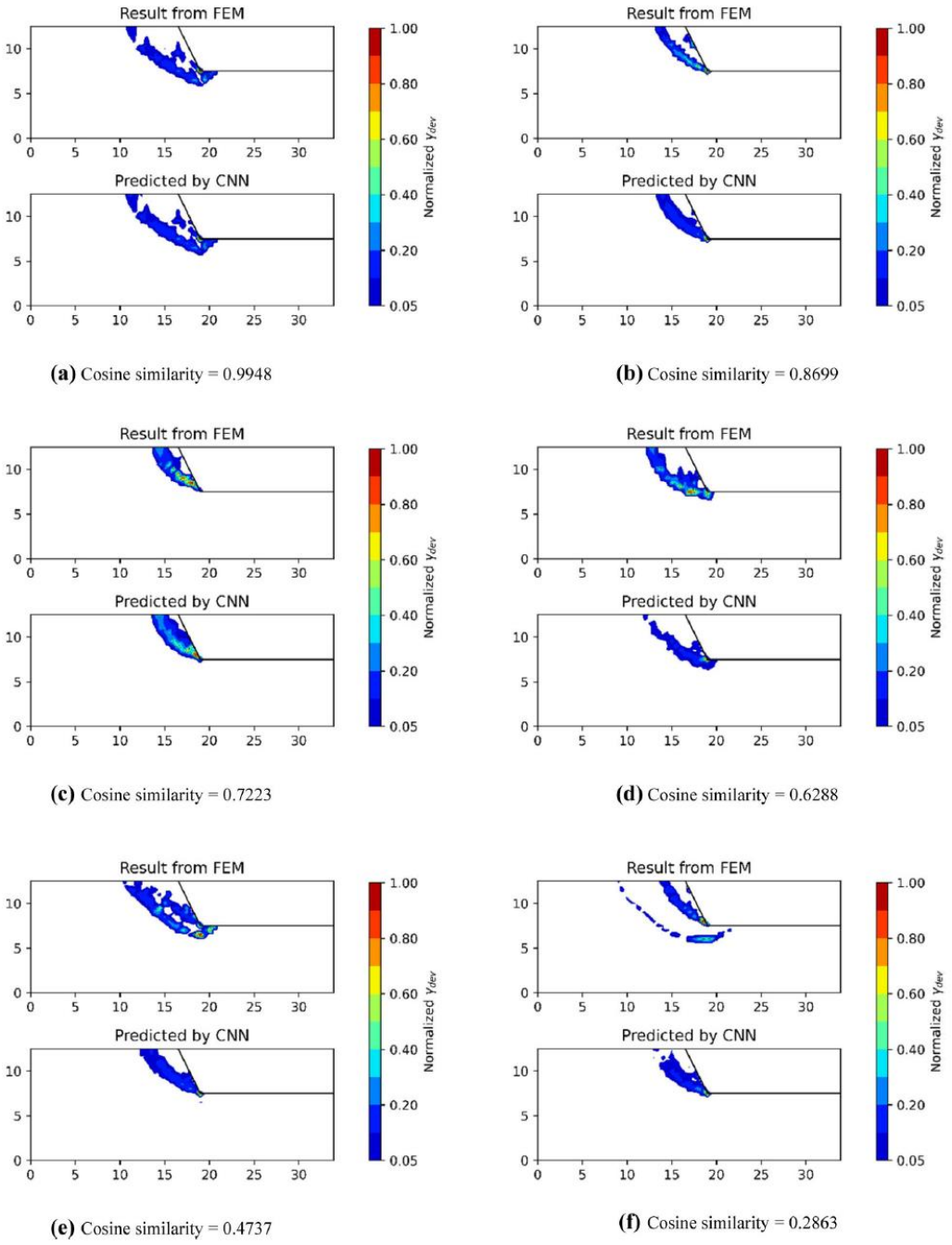


圖 2.7 CNN 模型邊坡破壞面預測 (Hsiao et al., 2022)

doi:10.6342/NTU202502436

Arif et al. (2025) 回顧中種機器學習術數在岩體邊坡穩定預測中的應用，常見輸入特徵包括坡度、坡高、言行、節裡面特徵與地下水位等，預測目標則為安全係數 (FS) 或是否穩定之分類結果。 研究指出 Random Forest 與 ANN 在處理非線性資料上具有優勢，並展現穩定的預測能力與泛化效果。部分研究近一步整合敏感度分析與特徵重要性排序以提升模型可解釋性，亦有研究使用混合型模型進行參數優化。整體而言，機器學習模型能有效捕捉多變數間的非線性關係，適用快速、初步的穩定性判斷。

地層辨識與分類

在地層辨識與建模方面，機器學習亦展現良好潛力。Wu et al. (2021) 針對紐西蘭南島粉土區的地層進行分類與三維建模，提出一套以 CPTu 參數為基礎的機器學習地層分類模型與分層邊界辨識方法。研究使用來自紐西蘭地工資料庫 (New Zealand Geotechnical Daatabase, NZGD) 的現地試驗資料，包含 33 比 CPTu 測點總計 3529 比樣本。輸入特徵包括錐尖阻抗 (qc)、摩擦比 (Rf)、孔隙水壓 (u_2) 等 CPTu 參數。研究首先針對土層分類進行監督式機器學習模型訓練與比較，使用包含隨機森林 (Random Forest, RF)、梯度提升機 (Gradient Boosting Machine, GBM)、支持向量機 (Support Vector Machine, SVM)、K-最近臨法 (K-Nearest Neighbors, KNN) 與多項式邏輯回歸 (Multinomial Logistic Regression, MLR)。在模型廷估方面，透過交叉驗證並採用正確率與 F1-score 作為模型表現指標，結果顯示 RF 模型整體表現最佳平均 F1-score 可達 0.90，較其他模型高出 5~12%，且在各土層類別中的分類一致性最高。SVM 與 GBM 亦展現出良好效能，但在模型參數調整與泛化表現上較差，而 KNN 與 MLR 則因對邊界養本判斷力較弱而表現不如預期。在三維建模部分，作者採用廣義回歸神經網路 (General Regression Neural Network, GRNN) 進行三維插值 (3D interpolation)，利用分類模型預測出的土層標籤作為訓練目標，將空間中未量測點之土壤類型進行預測與分

doi:10.6342/NTU202502436

部呈現。實證結果顯示，GRNN 所生成的三維地層模型與人工分層圖具有高度一致性，驗證該方法具備自動化、可擴展與解釋性兼具的應用潛力。研究結論指出，透過 CPTu 資料與機器學習模型結合，可有效實現土壤分類與地層邊界判釋之標準化與效率化流程，進一步應用於地工建模、場址風險分級及地質數位孿生系統等工程應用情境。

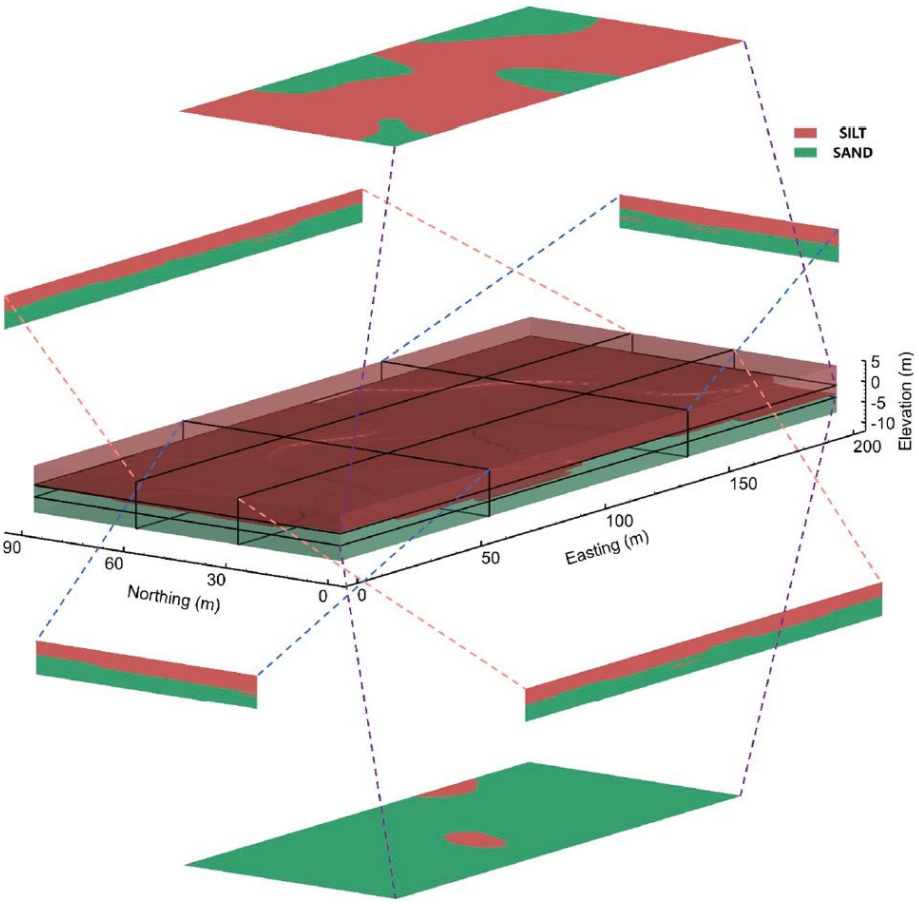


圖 2.8 場址 1 的 3D 建模分層結果 (Wu et al., 2021)

數值模型

近年有研究開始嘗試將機器學習與傳統數值模擬做結合，以提升其對土壤性質與行為預測的效率與準確性。Mitelman et al. (2023) 提出一套結合有限元素分析 (Finite Element Analysis, FEA) 與機器學習模型的數值-數據耦合架構用以建立可回歸土壤參數的預測模型。研究以三軸試驗數據為基礎，建構代表性數值模

doi:10.6342/NTU202502436

型，並將模擬結果作為訓練資料，以機器學習模型預測目標土壤參數，達到反算 (inverse analysis) 之目的。此研究資料來源為 1000 比數值模擬結果，每筆包含不同土壤初始狀態與材料參數，輸出則為對應之應立-應變行為。研究共測式多種迴歸模型包括 Random Forest、XGBoost、支持向量回歸 (Support Vector Regression) 與 ANN，並比較個模型在不同參數組合下的預測能力。結果顯示，ANN 與 XGBoost 於多數情境下具備最佳預測表現，平均決定係數 R^2 達 0.96 以上，顯示其可有效近似有限元素模擬結果與土壤材料性質之非線性對應關係。研究室提出一套基於模型誤差分析的參數重要性評估機制，辨識出對整體應力-應變行為影響最顯著的幾個材料參數，例如初始孔隙比、摩擦角與剪脹角等。該方法不僅可用於參數敏感度分析，亦具備高效率的預測潛力，可做為地工數值反算與場址校正分析的輔助工具。

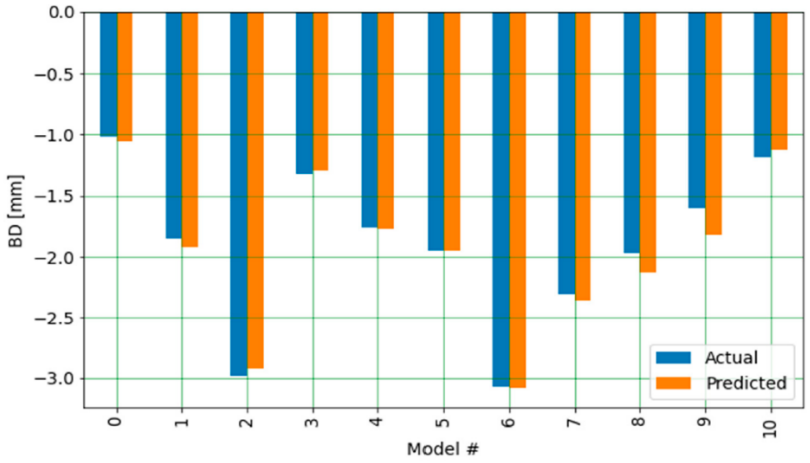


圖 2.9 機器學習於 10 個 FE 模型中預測值與真實質之比較圖 (Mitelman et al., 2023)

可靠度分析

近年來，地工工程對不確定性與風險控制的關注日益增加，機器學習與可靠度分析的跨領域研究也逐漸興起。Chwała et al. (2023) 提出一個具前瞻性的概念架構，目的是將過去、現在與未來的場址資訊整合入智慧化風險預測架構中，結合資料庫建立、機器學習模型、可靠度指標與可更新 (update) 技術，強調透過跨

doi:10.6342/NTU202502436

世代資料的累積與重複評估，提升場址判斷的長期準確性與韌性。該研究強調資料驅動模型 (data-driven models) 在處理地工不確定性時的優勢，並探討機器學習如何輔助傳統可靠度評估方法。研究指出，地工風險具有「時間變異性」，例如地下水位變化等因素會逐年改變場址條件與風險結構，因此傳統一次性分析方法已無法應對長期設計需求，呼籲未來應建立動態新式風險模型，並結合人工智慧工具與地工專業知識，提升工程決策的前瞻性與穩健性。

針對機器學習應用於地工可靠度分析的整體發展，Zhang et al. (2023) 提出一篇全面性的文獻回顧，系統性歸納各類機器學習方法於可靠度評估中的應用範疇、建模策略與未來挑戰。文章指出，傳統可靠度分析方法 (如一次性二階矩法、蒙地卡羅模擬等) 計算成本高且難以處理高維與非線性問題，機器學習模型因具備資料驅動與非參數建模能力，逐漸被用以建立替代模型 (surrogate models) 或極限狀態近似函數 (limit state function approximation)，已提效率並強化風險預測表現。該研究中涵蓋多種演算法，包括 SVM、RF、ANN、高斯過程迴歸 (Gaussian process regression, GPR) 與貝葉斯建模等，並指出這些模型可整合在可靠度分析流程中執行樣本生成、分類預測、表現評估與失效推估機率 (failure probability estimation) 等分析。討論數值模擬資料與實驗數據混合建模的重要性，並強調不確定性量化 (uncertainty quantification) 與模型解釋性在風險管理中的必要性。研究指出，機器學習方法提升了可靠度分析的效率與靈活度，但仍面臨如訓練資料不足、高維空間下的樣本稀疏、黑箱模型不度解釋等挑戰。

地盤沉陷預測

在大地工程領域中，盾構隧道 (shield tunneling) 施工引起的地表沉陷問題，對鄰近結構物安全與施工控制具有關鍵影響。在地盤沉陷預測方法方面，Chen et al. (2019) 以四個盾構區間的現地資料為基礎，包含土層性質，盾構操作參數與隧道幾何資訊，建立六種傳統機器學習模型進行比較，包括反向傳播神經網路

doi:10.6342/NTU202502436

(Back Propagation Neural Network, BPNN)、小波神經網路 (Wavelet Neural Network, WNN)、一般迴歸神經網路 (General Regression Neural Network, GRNN)、極限學習機 (Extreme Learning Machine, ELM)、支持向量機 (Support Vector Machine, SVM) 與隨機森林 (RF)。模型效能評估採用平均絕對誤差 (Mean Absolute Error, MAE)、均方根誤差 (Root Mean Squared Error, RMSE) 與決定係數 (Coefficient of Determination, R^2) 三項指標。其中，GRNN 與 RF 模型表現最佳，特別是 RF 在訓練集與測試集皆展現低誤差與高擬合度，顯示其擁有優良的泛化能力 (Generalization ability)。研究亦指出，部分參數 (如注漿量) 與沉陷量的關聯性可能不如預期，此結果突顯傳統經驗判斷在處理非線性問題時的侷限，而機器學習模型則具備彌補此一不足之潛力。

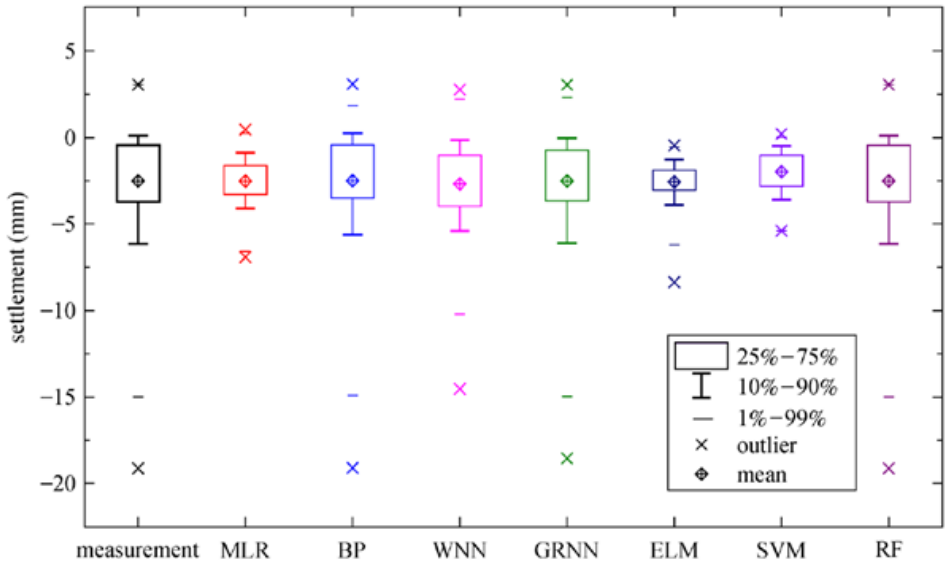


圖 2.10 機器學習模型於訓練集中預測值與量測值之箱型圖 (box plot) (Chen et al., 2019)

除施工引起之地表沉陷量外，液化所導致的建築物亦為一項重要議題。Liu and Macedo (2024) 提出一套以機器學習模型預測液化引起之建築物沉陷量方法。研究以大型震後實測資料庫為基礎，涵蓋土層性質、震動參數與建築物幾何資訊 (如層數、地基型式與建物寬度等)，建立三種迴歸模型，包含 RF、SVR 與 XGBoost，用以預測最大建築物沉陷量。

doi:10.6342/NTU202502436

在特徵選擇方面，研究使用皮爾森相關係數 (Pearson Correlation Coefficient) 與遺傳演算法 (Genetic Algorithm) 進行變數篩選，並進行五折交叉驗證 (5-Fold validation) 以提升模行之泛化能力。結果顯示，XGBoost 模型在各項評估指標 (MAE、RMSE 與 R^2) 中階表現最優，能更穩定得擬合不同地震與建築條件下的沉陷行為。RF 與 SVR 雖表現次之，但在特定地層條件中亦具良好解釋能力。此外，研究指出建物寬度、地震規模、液化深度與震後超額孔隙水壓比 (r_u) 為影響沉陷量的重要因子。

研究顯示，機器學習模型不僅可應用於地表沉陷預測，也具備估算建築物層級沉陷量的能力，對震後風險評估與城市韌性 (resilience) 提升具有實務價值。

液化潛能評估

在眾多應用機器學習進行液化潛能預測的研究中，Demir and Sahin (2022) 針對液化潛能預測任務，比較三種樹狀類機器學習演算法，包括典型相關森林 (Canonical Correlation Forest, CCF)、旋轉森林 (Rotation Forest, RotFor) 與隨機森林 (RF)。研究約以兩組約 480 筆的 CPT 實地案例資料進行模擬訓練，資料涵蓋震源特徵 (如地震規模 M_w 與地表加速度 a_{max})、CPT 指標 (如錐尖阻抗 q_c 與摩擦比 R_f)、地層條件、細粒料含量與深度資訊等，以二元分類方式預測目標為液化發生與否。

為確保模型評估具代表性，研究採用分層隨機抽樣 (Stratified Random Sampling) 劃分訓練與測試集，並測試不同抽樣比例 (50:50、60:40、70:30) 下模型之準確性。結果顯示，三種演算方式皆具穩定預測能力，其中 CCF 在多數資料組合下表現最佳，其除了正確度較高，也具超參數設定簡單的優勢。RotFor 和 RF 則在部分場景下表現接近，顯示資料結構與模型選擇間存在交互關係，如圖 2.11 所示。

doi:10.6342/NTU202502436

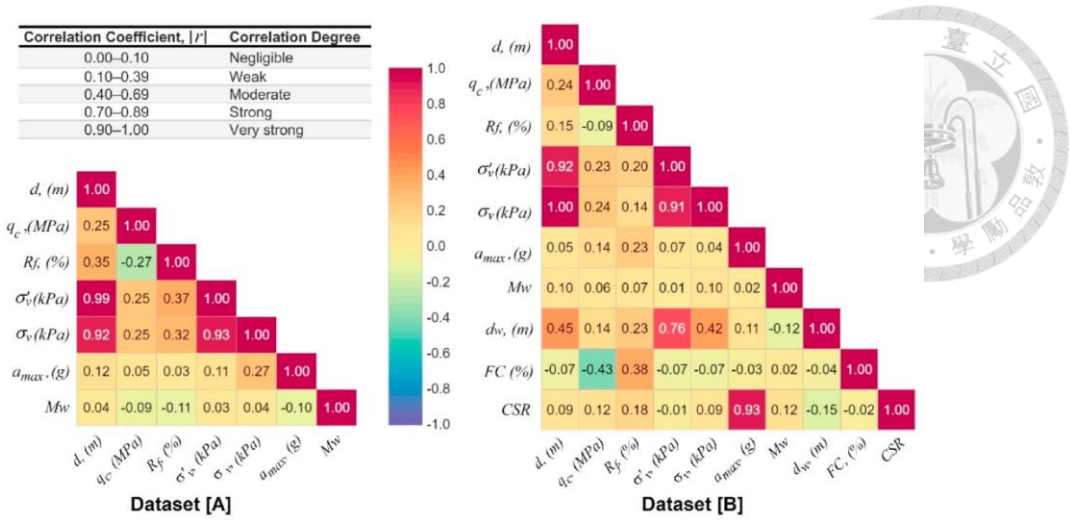


圖 2.11 每對變數的相關係數矩陣熱圖 (Demir & Sahin, 2022)

Jas and Dodagoudar (2023a) 提出一套結合可解釋機器學習技術的液化潛能預測流程，旨在提升模型準確性與預測結果的透明性。該研究以震後場址調查資料為基礎，使用 XGBoost 建構預測模型，並進一步結合 Shapley Additive exPlanations (SHAP) 分析以探討特徵對預測結果的貢獻，提升模型解釋性。

研究共蒐集 253 比經震後調查之 CPT 資料，輸入參數涵蓋震源特性(地震規模 M_w)、地表最大加速度 (a_{max})、循環剪應力比 (CSR)、土層深度 (Z)、細料含量 (FC)、錐尖阻抗 (q_c)、摩擦比 (R_f) 等常見指標，預測目標為液化是否發生，同樣視為二元分類分析問題。該研究將資料分為 70% 訓練集與 30% 測試集，並使用多項性能指標進行評估，包括正確率 (Accuracy)、精準率 (Precision)、召回率 (Recall)、F1 分數 (F1 Score) 與 ROC 曲線下面積 (Receiver Operating Characteristic-Area Under Curve, ROC-AUC)。

結果顯示，XGBoost 模型於測試集上表現穩定，正確率達 90%、AUC 值達 0.95，具高預測能力。透過 SHAP 值分解模型機制後發現，CSR、q_c 與 R_f 為影響模型判斷液化與否的三個主要特徵。SHAP 總結圖 (SHAP summary plot) 與個別貢獻圖清楚表示特徵值變化如何導致預測結果偏向液化或非液化，如圖 2.12 ，對於理解模型判斷依據與未來工程應用具有實質幫助。

doi:10.6342/NTU202502436

該研究不僅證實 XGBoost 可作為液化潛能預測的有效模型，亦展現 SHAP 解釋技術在土壤工程領域的應用潛力，為實務工程導入機器學習模型提供良好示範。

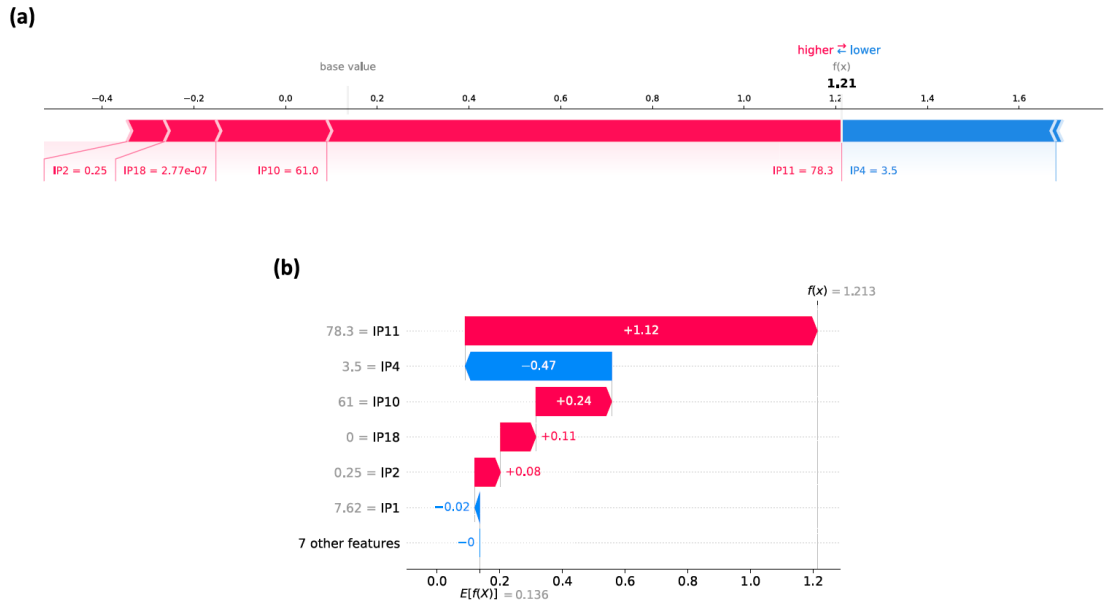


圖 2.12 各特徵之 SHAP 對模型的預測結果示意圖 (Jas & Dodagoudar, 2023a)

Jas and Dodagoudar (2023b) 針對近年機器學習於液化潛能評估領域的應用進行綜述，整理多篇以實驗數據或震後場址資料為基礎之研究，探討各類模型架構、輸入特徵與預測目標之差異。文獻以分類與迴歸任務為主軸，分析近 20 年來機器學習方法在液化潛能預測中之發展脈絡與挑戰。

該研究指出，液化潛能評估常用之輸入參數包括標準灌入試驗 (SPT-N 值)、錐尖阻抗 (CPT-q_c)、摩擦比 (R_f)、有效覆土應力 (σ'_{vo})、震度規模 (M_w)、最大地表加速度 (a_{max})、初始孔隙比 (e_0)、細粒料含量 (FC)、液性限度與塑性指數 (LL、PI) 等。在預測目標方面，可區分為是否液化之二元分類與連續變數如 CSR、FS 或液化指標 (如 Liquefaction Potential Index, LPI) 之迴歸預測分析。

在模型方面，該研究比較了機器學習方法 (如 K-最近臨法，K-Nearest Neighbors, KNN、SVM、DT、RF、XGBoost) 與深度學習架構 (如多層前饋神經網路，Multilayer Perceptron, MLP、CNN)，並總結各模型之優缺點。以分類任務

doi:10.6342/NTU202502436

為例，多數研究指出 RF 與 XGBoost 正確率穩定、對特徵尺度不敏感，且具備一定程度的可解釋性；深度學習模型如 MLP 與 CNN 雖具備處理大規模資料與特徵自動學習之優勢，但需大量資料與較長的訓練時間，且解釋性相較不足。

整體而言，該研究顯示多數模型在液化潛能預測分析中可達 85%以上之正確率，但資料來源與處理方式對結果影響顯著。

大地工程資料預測

土壤性質為大地工程設計與分析中最關鍵的參數之一，近年機器學習技術以廣泛應用於其預測任務中。Zhao et al. (2024) 整理 48 篇近年研究，深入比較各種機器學習在不同地工資料集上的預測效能，並探討特徵選擇策略對模型表現與解釋力的影響。研究指出，常用模型包括 Random Forest、XGBoost、SVM、K-NN 與 ANN。其中，Random Forest 與 XGBoost 在多數實驗中表現出穩定性高，適用於大樣本且高維度的資料集，但對訓練資料品質與參數設定較為敏感。該研究特別強調特徵選取的重要性。廣泛應用如遺傳演算法 (Genetic Algorithm, GA)、遞迴式特徵消除 (Recursive Feature Elimination, RFE)、相關性分析 (Correlation-based feature selection) 與樹模型內建重要性排序等方法。結果顯示，透過適當的特徵選取策略，不僅能提升模型的預測正確率，亦可降低過擬合風險與提升模型可解釋性。此外，該研究亦探討資料來源的多樣性與資料品質對模型泛化能力之影響。部分文獻採用實驗室的試驗結果，亦有結合現地監測數據與數值模擬輸出的混合資料集。總和而言，此研究凸顯了機器學習在地工性質預測中所具備的潛力，並指出選擇適當的特徵組合與模型架構可提升預測準確性與工程解釋力。

doi:10.6342/NTU202502436

第三章 研究方法

本研究流程如圖 3.1 所示，包含資料準備和模型建立、訓練與測試兩大部分，第一部分包含資料蒐集、資料前處理和特徵分類；第二部分包含各模型的建置、超參數調整、訓練與測試。首先透過文獻收集相關試驗數據，進行特徵編碼，同時採用多種方法處理缺失值，補足缺漏資料。資料經處理後，依比例劃分為訓練集與測試集，並以不同模型進行訓練，最後評估各模型預測結果，最後探討其預測表現與適用性。

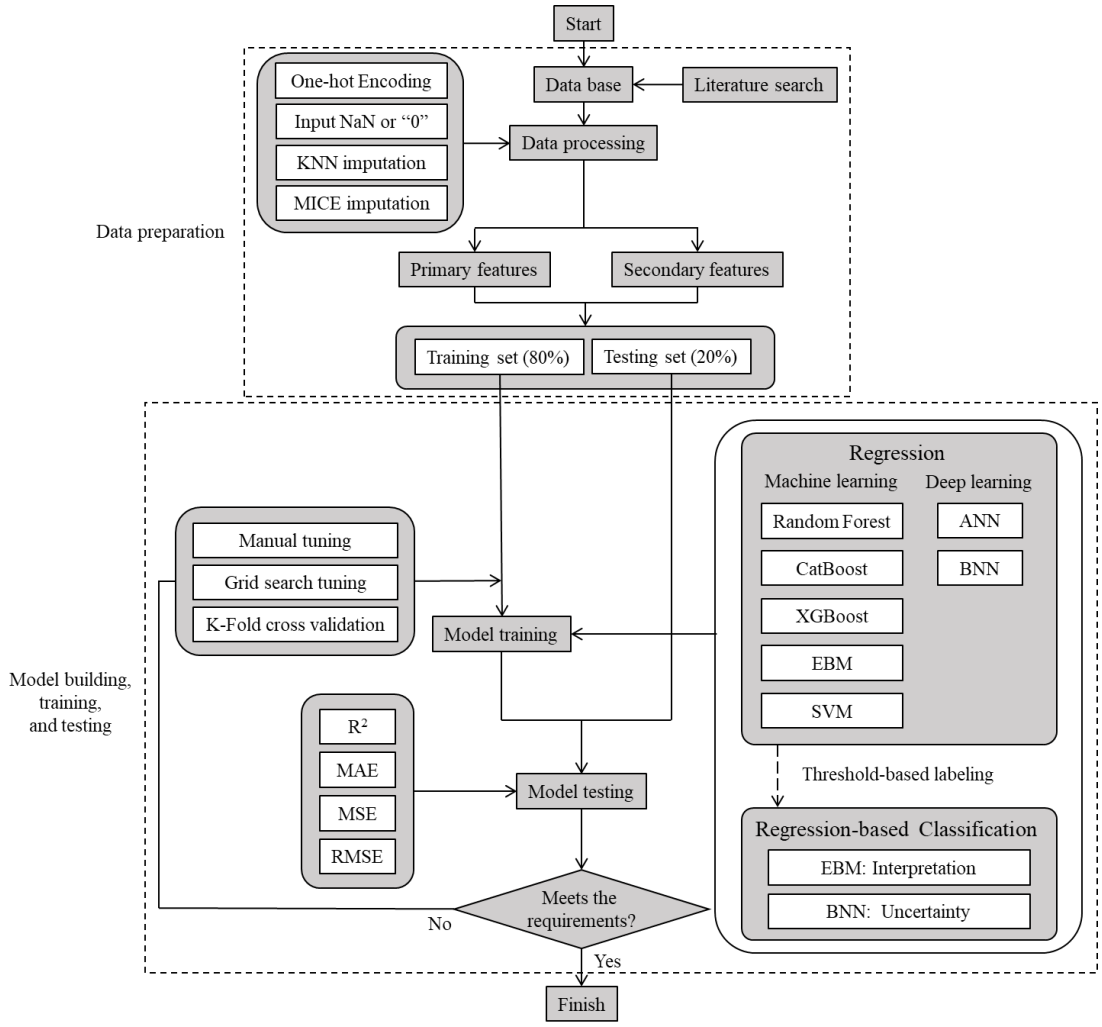


圖 3.1 本研究流程圖

doi:10.6342/NTU202502436

本研究以機器學習與深度學習技術建立迴歸 (Regression) 與分類 (Classification) 模型，進行動態三軸試驗數據之預測與判別分析。模型實作環境為 Python 版本 3.12.9，主要透過 Visual Studio Code (VS Code) 撰寫程式，並使用 `.venv` 虛擬環境以確保執行環境的一致性與重現性。深度學習模型採用 PyTorch 框架建構，部分模型訓練階段亦結合 GPU 提升運算效率。

3.1 數據來源與描述

本研究所使用的資料來源為整理自 23 篇已發表之學術期刊論文與博士學位論文中之動態三軸試驗資料，共整理出 997 筆試驗數據，皆為實驗室之試驗結果。為後續模型訓練與變數分析，本研究根據資料特性，將各項試驗變數依資料完整度及資料類別分為「主要特徵 (Primary features) 」與「次要特徵 (Secondary features) 」兩類。其中，主要特徵包含細粒料含量 (f_c, %)、初始孔隙比 (e_o)、塑性指數 (PI, %)、試驗圍壓 (p'_o)、循環剪切次數 (N)、剪切頻率 (Frequency, Hz)、液化判斷標準 (Liquefaction criteria)、以及破壞模式 (Failure pattern)。以上 8 個特徵幾乎於所有進行動態三軸試驗之文獻中均有完整紀錄，涵蓋試體的物理性質及試驗的控制與判斷參數，除了其缺失值比例極低、資料完整性高之外，更重要的是這些參數在動態三軸試驗領域中被廣泛視為影響試驗結果之關鍵因素，因此被歸類為主要特徵。

本研究的「液化判斷標準 (Liquefaction criteria)」為蒐集文獻裡常見之動態三軸試驗中停機或判定液化發生的依據，共整理出六種主要判斷標準，分別為:

- 雙振幅軸向應變達 5% ($\varepsilon_{\mathrm{DA}} = 5\%$)
- 初始有效應力為零 ($\sigma'_o = 0$)
- 超額隙水壓比R_u大於 0.95 或雙振幅軸向應變達 5% (超額隙水壓比R_u大於 0.95 或雙振幅軸向應變達 5% ($R_u > 0.95$ or $\varepsilon_{DA} = 5\%$)

doi:10.6342/NTU202502436

- 循環剪切次數大於等於 20 或雙振幅軸向應變達 5% 循環剪切次數大於等於 20 或雙振幅軸向應變達 5% ($N \geq 20$ or $\varepsilon_{DA} = 5\%$)
- $\varepsilon_{\text{axial}} = 20\%$ or Seriously deformed

「破壞模式 (Failure pattern)」是針對試驗中觀察到的土體變形行為進行分類，其於第 2.1.1 節之破壞機制已有說明，包含 4 類:

- 反覆軟化 (Cyclic Softening)
- 反覆流動 (Cyclic Mobility)
- 殘餘變形累積 (Residual Deformation Accumulation)
- 流動液化 (Flow Liquefaction)

由於不同文獻對於液化的定義與判斷標準並不一致，且不同的破壞模式亦可能影響其所對應之 CSR 值。因此，本研究將液化判斷標準作為一項關鍵資訊納入主要特徵當中，期望模型能在訓練過程中考量此特徵針對 CSR 預測結果之影響。

次要特徵則包含相對密度 (D_r, %)、比重 (G_s) 、均勻係數 (C_u) 、粒徑相關指標 (D_{10}、D_{30}、D_{50}、D_{60}, mm)、最大與最小孔隙比 ($e_{\max}$、$e_{\min}$)、塑性限度 (PL, %)、液性限度 (LL, %)、粉土含量 (Silt content, %)、黏土含量 (Clay content, %)、B 值 (Skempton's B-coefficient) 與反水壓 (Back pressure, kPa)。此類共 15 個特徵並非於所有文獻中均有紀錄，具有不同程度的缺失率，但考慮其在描述土壤試體與試驗條件的重要性，故納入資料庫中供後續模型訓練使用。

為掌握整體資料結構與欄位缺失情形，表 3.1 與表 3.2 則分別統計主要與次要特徵之欄位名稱、資料型態和缺失率，作為後續補值策略與特徵選擇的依據。此外，圖 3.2 呈現各特徵之缺失分佈熱區 (Heat map)，其中橫軸為特徵名稱，縱軸為資料編號，深藍色區塊代表缺失資料位置，可直觀呈現資料集在不同變數上的完整性狀況。

doi:10.6342/NTU202502436

表 3.1 主要特徵之資料類別與缺失率表

Column Name	Data Type	Missing Percentage (%)
fc	float64	0.00
e0	float64	0.00
PI	object	0.00
p'0	float64	0.00
N	float64	0.00
Frequency	float64	0.00
Liquefaction criteria	object	0.00
Failure pattern	object	79.50

表 3.2 次要特徵之資料類別與缺失率表

Column Name	Data Type	Missing Percentage (%)
Dr	float64	58.99
Gs	float64	27.44
Cu	float64	32.06
D60	float64	60.00
D50	float64	20.90
D30	float64	60.00
D10	float64	60.00
emax	float64	29.65
emin	float64	29.65
LL	float64	8.54
PL	float64	9.65
silt content	float64	4.02
clay content	float64	4.02
B value	float64	8.04
Back pressure	float64	52.46

doi:10.6342/NTU202502436

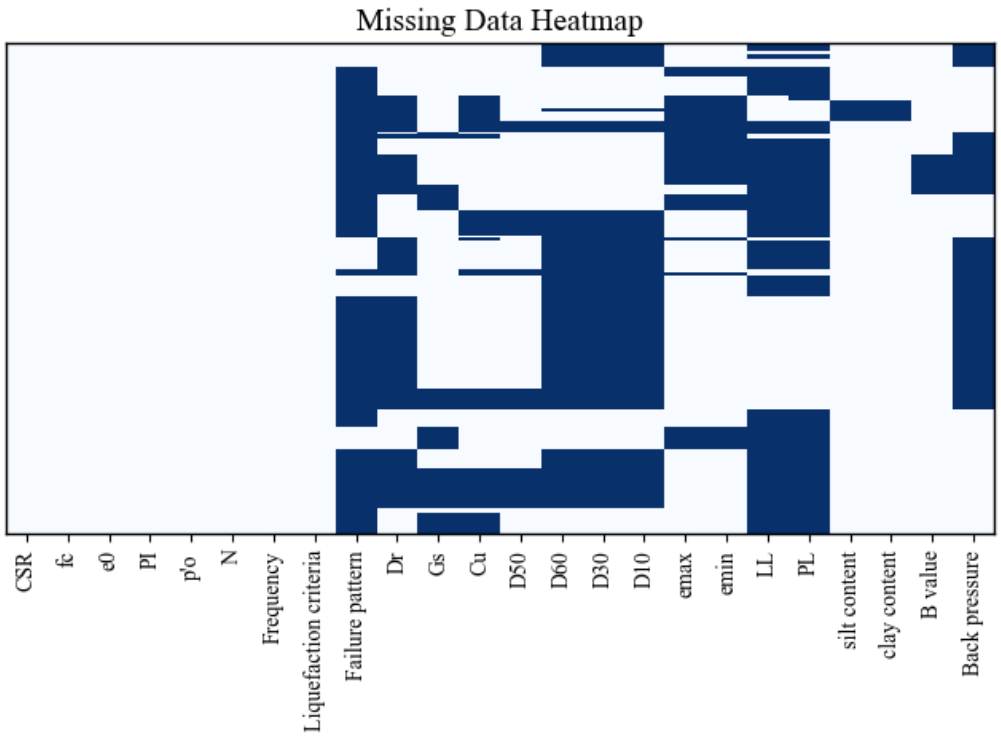
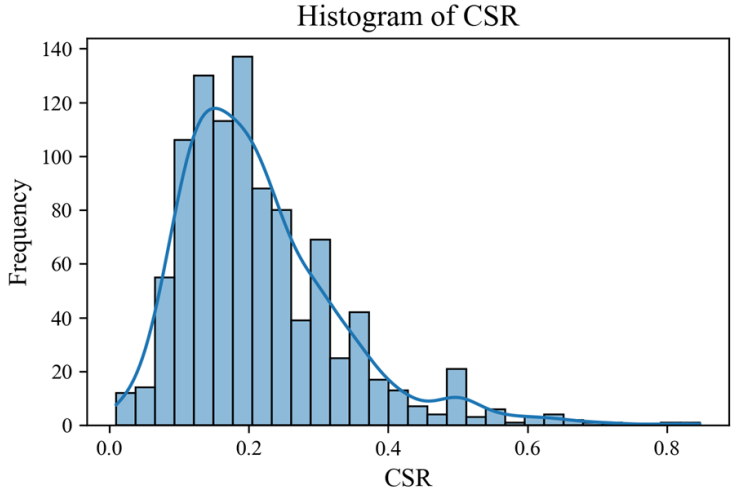


圖 3.2 缺失熱區分佈圖

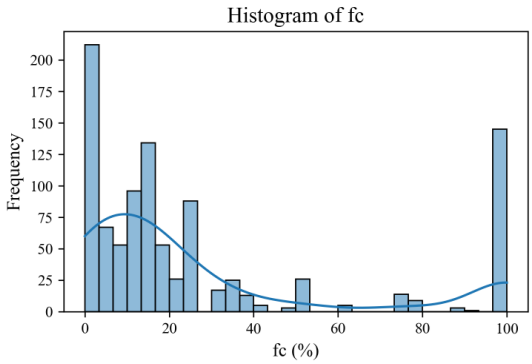
而為了解各項數值型特徵的資料分布情形，針對所有 float 型欄位繪製直方圖 (histogram)，如圖 3.3 所示。圖中展示每一變數的數值分布與密度變化，可作為後續特徵工程與模型建構的參考。

最後，所有資料皆包含循環剪應力比 (Cyclic Stress Ratio, CSR)，作為本研究之預測目標變數 (Target)，以探討各特徵在不同試體及試驗條件下對 CSR 的影響與預測能力。

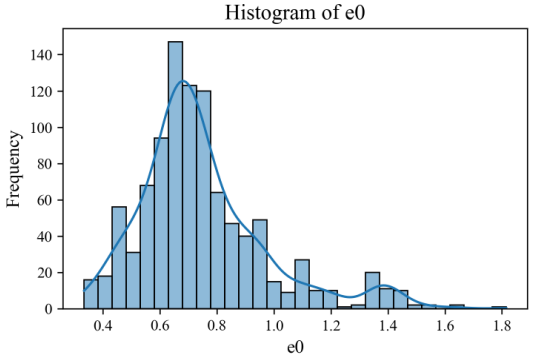


(a)

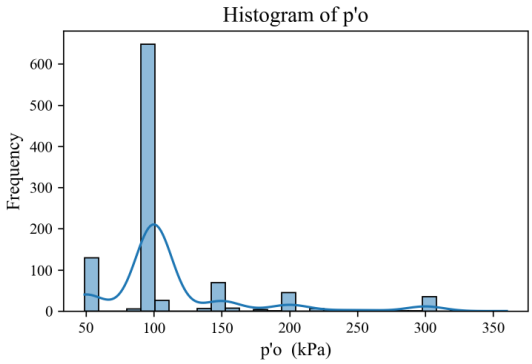
doi:10.6342/NTU202502436



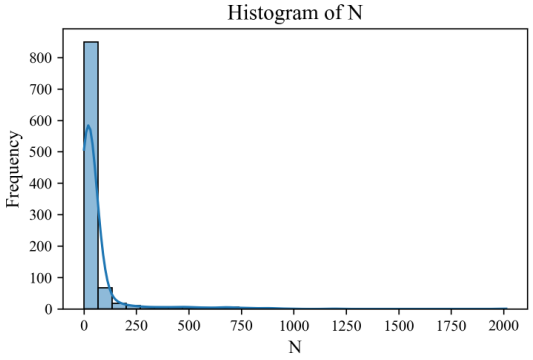
(b)



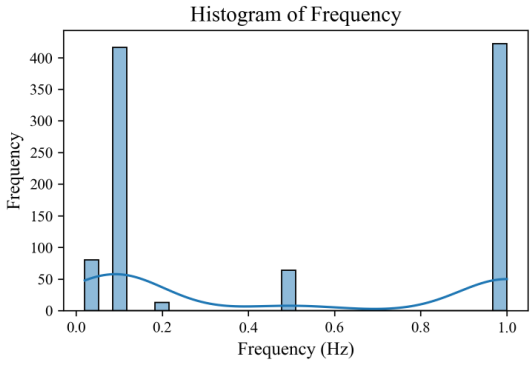
(c)



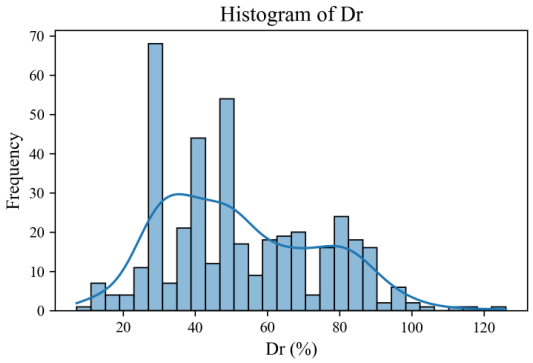
(d)



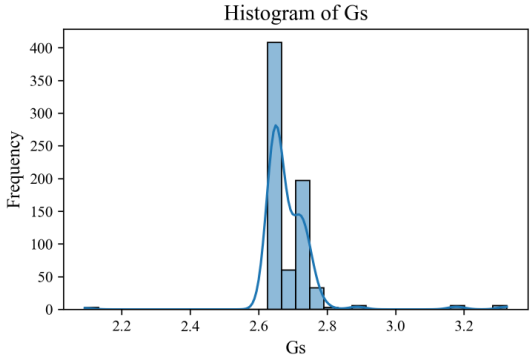
(e)



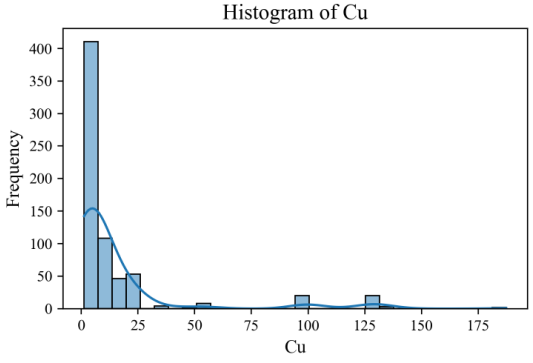
(f)



(g)

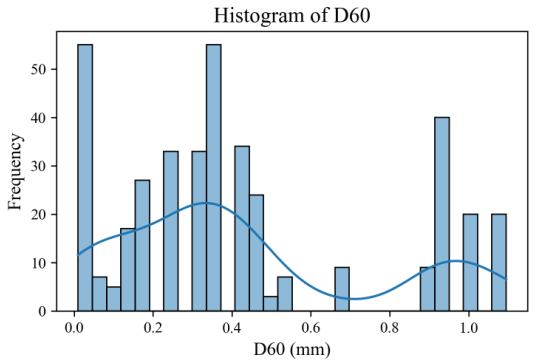


(h)

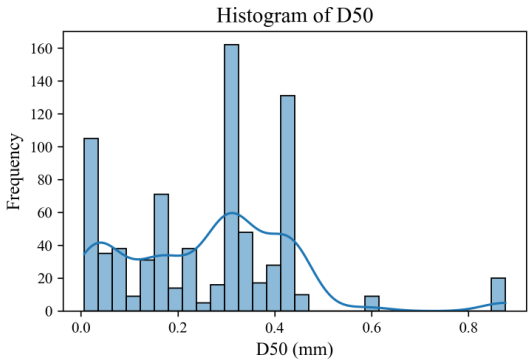


(i)

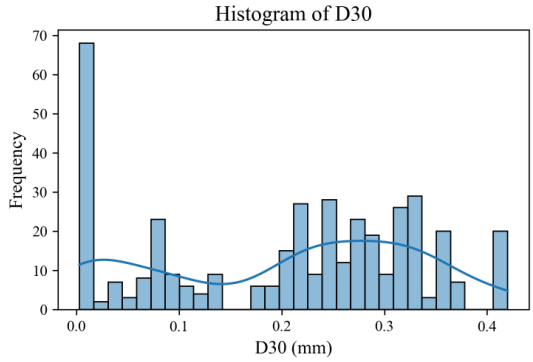
doi:10.6342/NTU202502436



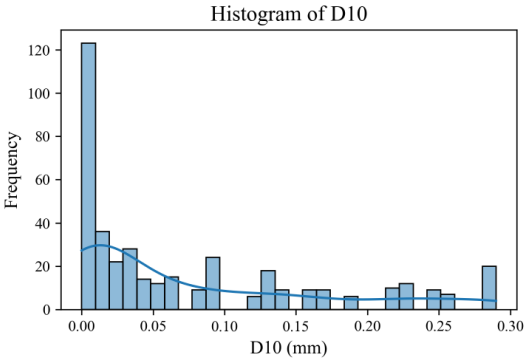
(j)



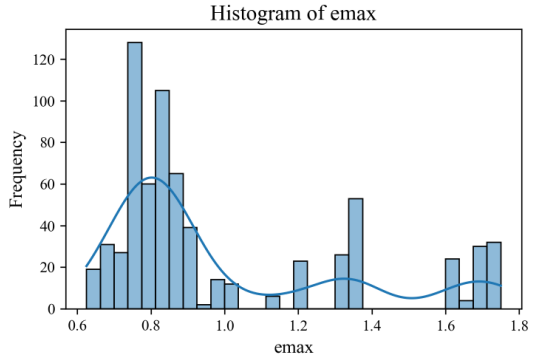
(k)



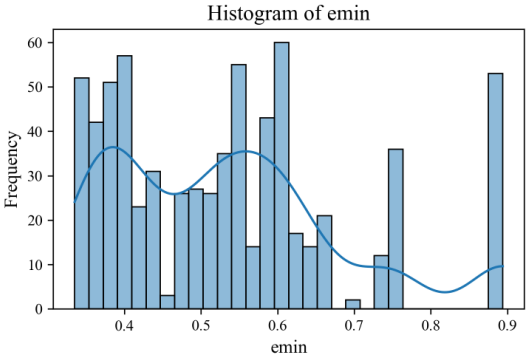
(l)



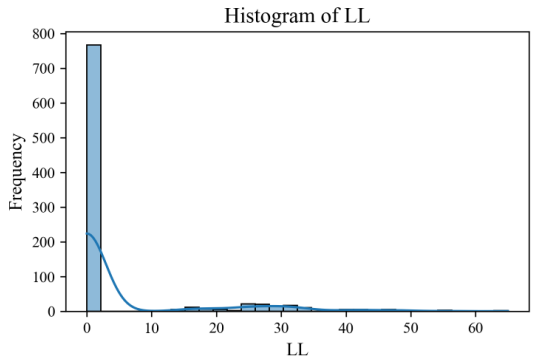
(m)



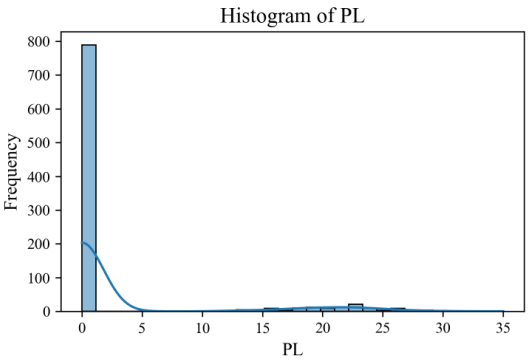
(n)



(o)



(p)



(q)

doi:10.6342/NTU202502436

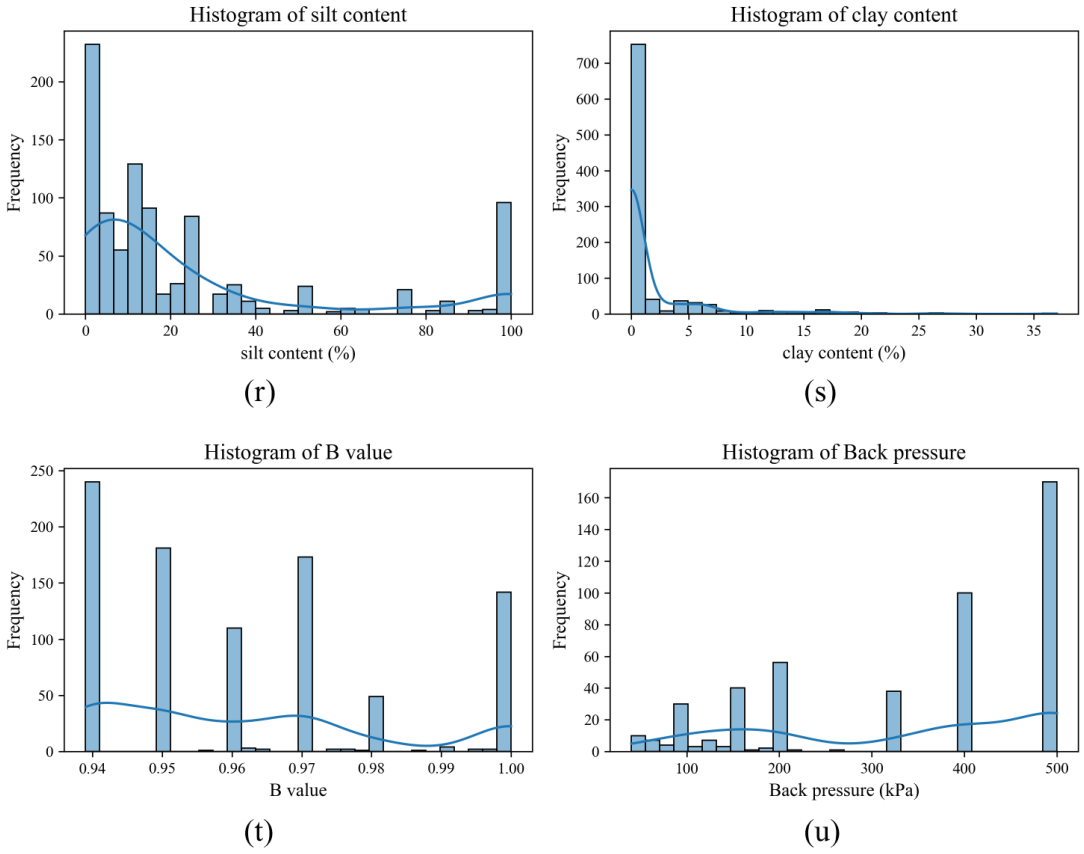


圖 3.3 各數值型特徵之長方圖 (a) CSR；(b) f_c；(c) e_0；(d) p'_0；(e) N；(f) Frequency；(g) D_r；(h) G_s；(i) C_u；(j) D_{60}；(k) D_{50}；(l) D_{30}；(m) D_{10}；(n) e_{max}；(o) e_{min}；(p) LL；(r) PL；(s) silt content；(t) B value；(u) Back pressure

由於本研究旨在探討含細粒料土壤於動態三軸試驗中的液化行為，因此在資料蒐集階段專注於含細粒料之土壤試體。然而，相較於純砂試驗資料的普遍性，針對含細粒料之試驗資料相對較少，蒐集過程具有一定困難。多數文獻傾向分別記錄所使用之砂與細粒料的物性試驗結果 (如砂主要是以相對密度做描述，細粒料則有液限或塑限等指標)，再直接按比例混合進行動態三軸試驗，並未進一步提供混合後試體的物理性質。為確保資料品質與一致性，本研究僅選取有明確記錄混合後土體性質之文獻作為分析依據，然而此一資料選擇策略造成部分常見參數 (如相對密度或粒徑相關參數) 出現缺漏，最終自瀏覽之 165 篇文獻中篩選出 23 篇符合條件之文獻作為資料蒐集來源。

doi:10.6342/NTU202502436

3.2 數據前處理與特徵工程

在建立機器學習模型之前，通常資料的前處理與特徵工程是不可或缺的步驟，資料本身的品質將直接影響模型的學習效果，以及其預測與判別能力。尤其在工程領域中，完整性、品質與代表性兼具的資料取得不易，本研究蒐集之資料亦不例外，受到來源差異與紀錄限制的影響，資料有缺失、標準不一致、資料分佈不均與含異常值 (Outlier) 等情況均為常見，若未經適當處理，將可能降低模型準確性與穩定性，甚至導致學習偏差或過擬合 (Overfitting)。

此外，特徵工程亦有助於強化資料與目標變數之間的關聯性。透過合理的變數轉換、特徵篩選與類別編碼，不僅可以提升模型的學習效率和預測與判別的表現，亦有助於提升模型對未知資料的泛化能力 (Generalization ability)。

因此，本章節將依序說明本研究中所採用的缺失值處理、類別變數編碼與數值特徵標準化等策略，作為後續建立機器學習模型之資料基礎。

3.2.1 缺失值處理方式

本研究針對不同資料型態與缺失特性使用四種缺失值處理方式，依序說明如下:

填補 NaN (Not a Number)

在初步資料處理階段，將原始資料中以空白字串、"-"、"NA"、"Na"、"null"、"N/A"等標示缺失的欄位，統一轉換為 NumPy 系統中的 NaN 形式，使後續模型能正確辨識並處理。

將缺失欄位填入 NaN 雖非補值行為，但具有其實務與數據合理性上的考量。首先，保留欄位為空白狀態，可忠實反映文獻中未記載或未量測該變數的事實，使模型有機會辨識並學習「空值」本身所代表的資訊。其次，若直接採用統計補值 (如平均值或中位數) ，可能因資料分佈偏態或含有異常值，導致填入偏離常

doi:10.6342/NTU202502436

理的數值；或若採用其他補值策略，亦有可能補入不合理的值 (如負的平均粒徑 D_{50} 或不具物理意義的極端值) ，進而扭曲資料真實分佈。相較之下，標記為 NaN 可視為一種相對保守的非補值處理方式，可避免過度假設性補值所造成的偏差。

然而，此方法亦有缺點。若變數存在過多的 NaN，將造成其數據分佈不連續，進而降低模型對該變數的辨識與學習能力，可能導致模型無法有效捕捉其潛在規律。此外，模型亦可能需額外學習「缺失」本身所代表的隱含資訊或潛在規則，進一步增加模型的參數負擔與訓練資源耗費，影響整體訓練效率與預測與判斷效能。

以 0 補值

由於有些機器學習與深度學習演算法 (包括支持向量機 Support Vector Machine, SVM、人工神經網路 Artificial Neural Network, ANN 與貝葉斯神經網路 Bayesian Neural Network, BNN) 皆無法處理 NaN 值，若不事先進行填補，將導致模型無法執行訓練、預測或判斷等流程。若不採用其他適合的補值方法，將缺失值以 0 補上，可確保資料格式完整並能順利輸入模型。

此外，部分變數的缺失並不單純代表缺漏，而可能是具有特定物理意義。例如，在塑性指數 (PI) 欄位中常見的"NP" (non-plastic) 即表示該試體不具塑性，此時以 0 作為補值，實為一種明確且合乎工程邏輯的處理方式。

然而，以 0 補值將失去「空值」所代表的資訊特性，並可能嚴重影響該變數原有的數值分佈結構。再者，補入的 0 值若與實際資料中具有實際意義的 0 混淆，將導致模型錯誤解讀該變數的工程意涵，影響特徵權重評估與模型準確性。因此，補 0 雖能提升模型可讀性與運算效率，但採用此法前仍須審慎判斷其合理性及適用性。

doi:10.6342/NTU202502436

KNN 補值 (K-Nearest Neighbors Imputation)

Zhang (2012) 指出，KNN 插補法的設計目的是針對資料集中具有遺漏值的數據，自所有完整數據中尋找與其最相似的 k 筆鄰近樣本，並根據這些鄰近樣本的資訊進行補值。若欲填補之值為類別型資料，則以鄰近樣本中最常出現的類別進行補值，稱多數法則 (majority rule)；若為數值型資料，則以鄰近樣本的平均值進行補值，稱為平均法則 (mean rule)。

KNN 補值常採用歐式距離 (Euclidean Distance, d) 作為樣本相似性衡量指標，其計算公式如式(3.1):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.1}$$

其中，x_i 與 y_i 分別為兩筆樣本在第i個變數空間上的值，總計 n 個變數。

KNN 補值適用於具區域分佈特性與結構相似性之資料，透過搜尋特徵空間中最相近的 k 筆樣本進行加權平均 (或中位數) 補值，能有助於保留原始資料的分佈型態與局部特徵。其優點為填補過程不須先建立數學模型、運算邏輯簡單，且適合處理非線性與多變數間之資料補齊。然而，其缺點為對異常值敏感、在高維資料下計算成本高、補值準確度受限於距離計算方式與 k 值的設定，此外，若所有變數於距離計算中權重相同，則在變數尺度差異顯著的情況下，容易產生不精確的補值結果。

本研究採用 Scikit-learn 提供之 `KNNImputer` 模組進行補值，設定參數如下: `n_neighbors = 5`，代表每筆缺失值將參考其於特徵空間中距離最近的五筆完整樣本進行補值；`weights = "uniform"`表示所有鄰居的權重相等；`metric = "nan_euclidean"` 則允許樣本中部分欄位缺失時仍可進行距離計算，以提升模組彈性與適用性。

doi:10.6342/NTU202502436

MICE 補值 (Multiple Imputation by Chained Equations)

Zhang (2016) 提到，多重插補法 (Multiple Imputation, MI) 是一種處理缺漏值的進階方法。相較於傳統的單一插補 (Single Imputation)，MI 會針對每筆缺失產生 *m* 個合理的插補植，進而建立 *m* 個完整資料集。這些插補值在產生過程中會補值的不確定性，避免對遺漏資料的估計過度簡化。在後續統計分析時，將針對每一組補值資料集分別進行統計量估計，最後在將這些估計結果整合為一組綜合性的推論結果，如圖 3.4。

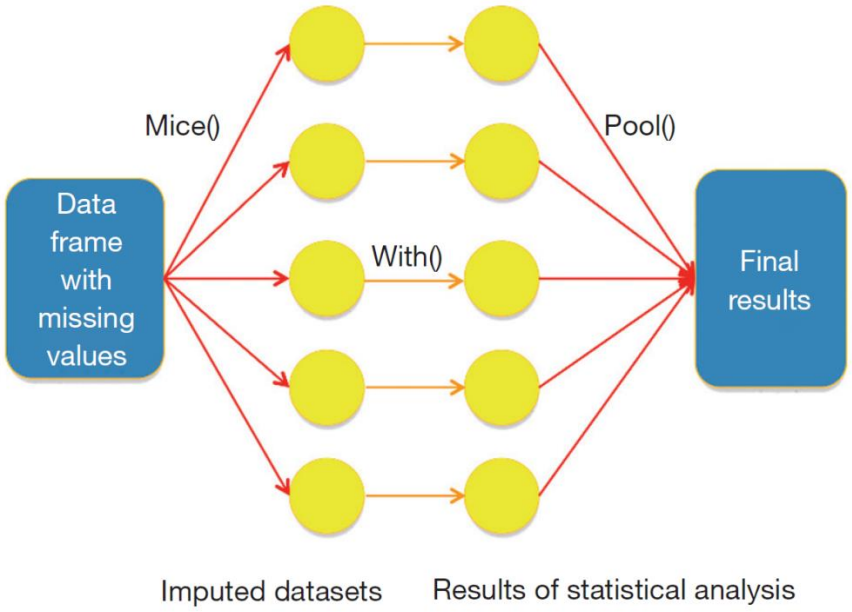


圖 3.4 MICE 套件處理含缺失值的資料框架示意圖 (Zhang, 2016)

MICE 為多重插補中常用的方法，其核心特徵是透過串聯迴歸模型 (Chained Equations) ，逐一針對每個含缺失值之變數建立預測模型，進行遞迴式補值。每一輪補值迴圈 (chaining cycle) 皆會將前一輪的補值結果納入下一輪模型 ，重複數輪後收斂至穩定結果 (Kose et al., 2020)。其流程如圖 3.5 所示。

doi:10.6342/NTU202502436

```
A simple imputation such as “mean” is performed for each missing value in the dataset
For each iteration
    For each variable v with missing values
        Train a regression model with
          (1) v as the dependent variable
          (2) the others as independent variables
        Impute the missing values of v using the regression model
    End
End
```

圖 3.5 MICE 建模流程示意圖

此方法可處理連續型、類別型、有序型等混合變數，具實用性與應用彈性。然而，此 MICE 補值法亦存在缺點，若可用變數不足或變數相關性低，補值準確性將會受影響；若資料維度高或缺失比例大，運算成本高，此外，多輪補值可能產生樣本間的依賴性，若後續統計推論未妥善考慮這些不確定性，能可能導致結果產生偏誤。

本研究亦採用 Scikit-learn 套件中所提供之 `IterativeImputer` 模組進行多重插補，設定參數如下: `random_state = 42` 表示固定隨機種子以確保重現性；`max_iter = 10` 則代表做多進行 10 次完整的變數補值迴圈 (cycle)。此設定有助於提升補值結果之收斂性與穩定性，為後續機器學習模型建構提供較完整且一致的資料基礎。

3.2.2 類別型欄位轉換

在進行機器學習模型建構之前，資料需轉換為模型可接受的數值格式。針對類別型 (Categorical) 變數之欄位進行轉換處理，常採用標籤編碼 (Label Encoding) 與獨熱編碼 (One-Hot Encoding) 等兩種方式，以提升模型可讀性，兩方法說明如下:

doi:10.6342/NTU202502436

標籤編碼 (Label Encoding)

標籤編碼 (Label Encoding) 是透過為每個類別分配唯一的整數數值 (如 0、1、2…) 來進行表示，如圖 3.6 所示。此法適用於具明確順序性之變數，能保留類別間的相對大小或階層關係，協助模型理解其排序關係。

Cat	Fruit
Persian	Blueberry
Persian	Blueberry
Persian	Kiwi
Persian	Kiwi
Ragdoll	Kiwi
Ragdoll	
Exotic Shorthair	
Exotic Shorthair	Peach
Persian	Peach
Persian	
Persian	

Cat
0: Persian
1: Ragdoll
2: Exotic Shorthair

Fruit
0: Blueberry
1: Kiwi
2: " "
3: Peach

Cat	Fruit
0	0
0	0
0	1
0	1
1	1
1	2
2	2
2	3
2	3
0	2
0	2

圖 3.6 標籤編碼示意圖

因本研究之類別型數據皆為獨立且無順序性的離散類別，故選擇不採用標籤編碼作為類別型欄為轉換方式，以避免模型誤解類別型資料間具備數值上的關聯性。

獨熱編碼 (One-Hot Encoding)

獨熱編碼 (One-Hot Encoding) 是將每個類別型資料轉換為獨立欄位 (dummy variable) 的方式，透過建立與類別數相同的虛擬欄位，分別以 1 與 0 表示該樣本是否屬於該類別。此方法適用於無順序性之離散類別型變數，能有效避免模型誤判類別間的數值關係，並保留每個類別的獨立性，如圖 3.7 所示。

doi:10.6342/NTU202502436

Cat	Fruit
Persian	Blueberry
Persian	Blueberry
Persian	Kiwi
Persian	Kiwi
Ragdoll	Kiwi
Ragdoll	
Exotic Shorthair	
Exotic Shorthair	Peach
Persian	Peach
Persian	
Persian	

⇨

Cat_Persian	Cat_Ragdoll	Cat_Exotic Shorthair	Fruit_Blueberry	Fruit_Kiwi	Fruit_Peach
1	0	0	1	0	0
1	0	0	1	0	0
1	0	0	0	1	0
1	0	0	0	1	0
0	1	0	0	1	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	1
1	0	0	0	0	1
1	0	0	0	0	0
1	0	0	0	0	0

圖 3.7 獨熱編碼示意圖

本研究中，類別型變數如「液化判斷標準 (Liquefaction criteria) 」與「破壞模式 (Failure pattern) 」，皆為不具順序關係之類別資料，因此採用獨熱編碼進行轉換。於程式實作上，使用 Pandas 中 `pd.get_dummies()`函數對類別欄位轉換，並設定 `drop_first = True` 以避免產生完全共線性 (dummy variable trap) 問題。完全共線性是指某一欄位可由其他一個或多個欄位以線性方式完全表示，使得該欄位成為冗餘資訊，這會導致模型在運算過程中產生矩陣不可逆等問題，而設定 `drop_first = True` 可使每個類別變數移除一個基準欄位，藉此消除冗於欄位同時避免完全共線性問題。

另外，本研究所採用的 Categorical Boosting (CatBoost) 模型具備原生處理類別型變數的能力，因此無須再訓練前進行標籤編碼或讀熱編碼。

3.2.3 數值型欄位標準化

在機器學習模型建構前，數值型欄位常具有不同的尺度 (scale) 與單位 (unit)，若未進行適當的縮放處理，可能導致模型在參數學習過程中受到變數尺度影響產生偏誤，特別是在需進行距離運算 (如 SVM) 或梯度優化 (如 ANN 與

doi:10.6342/NTU202502436

BNN) 的演算法中影響特別顯著。因此，對數值型特徵進行標準化處理有助於提升模型之穩定性與效能。

Sujon et al. (2024) 提到兩種常見特徵縮放方式: 正規化 (Normalization) 與標準化 (Standardization)，分別說明如下:

正規化 (Normalization)

正規化是一種將數值縮放至固定範圍 (通常為 0 至 1) 之方式，其公式如式(3.2):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3.2}$$

其中，X 為原始特徵值，$X_{\min}$ 與 $X_{\max}$ 分別為該特徵的最小與最大值。此法適用於需維持資料比例關係的模型，缺點則是對於資料中的極端值或離群值相當敏感，當資料分佈不均或含有極端值時，可能會壓縮大部分數據，導致資訊損失與模型辨識能力下降。

標準化 (Standardization)

標準化則是將資料轉換為平均值為 0、標準差為 1 的常態分佈型式，其公式如式(3.3):

$$Z = \frac{X - \mu}{\sigma} \tag{3.3}$$

其中 Z 為標準化後之數值，X 為原始特徵值，μ為平均值，σ為標準差。此方法不受原始資料尺度與分佈之影響，適用於多數機器學習演算法，其缺點為標準化後的數值較不直觀，且對於筆數龐大或變數數量多的資料集，在計算平均值與標準差時，運算負擔較大。

由於本研究的資料來自不同文獻與實驗室，部分特徵 (如循環剪切次數 N) 具有極端值或偏態分佈，若採用正規化方法 (Normalization) 可能導致模型學習

doi:10.6342/NTU202502436

偏誤，進而影響模型的辨識能力。因此，本研究採用標準化 (Standardization) 做為數值型特徵的縮放方法，並以 Scikit-learn 套件中提供之 `StandardScaler` 模組進行實作。

3.3 模型介紹

本研究旨在預測不同土壤試體與試驗條件下之循環剪應力比 (Cyclic Stress Ratio, CSR) ，屬於連續型數值預測問題，故本研究主要問題性質為迴歸型 (Regression) 分析。由於資料中每筆樣本皆對應一組輸入特徵與目標變數，因此本研究採用監督式學習架構進行模型訓練與預測。

監督式學習能有效建構輸入變數與目標變數之間的數學關係，適用於含有標籤資料 (label，即 CSR) 之資料集。本研究採用多種監督式學習模型進行建模與比較，涵蓋機器學習方法 (Machine Learning, ML) 與深度學習方法 (Deep Learning, DL)，如下列所示，並進行系統性的調參與效能評估:

- 機器學習方法
    1. 隨機森林 (Random Forest, RF)
    2. 分類增強模型 (Categorical Boosting, CatBoost)
    3. 極端梯度提升機 (Extreme Gradient Boosting, XGBoost)
    4. 可解釋增強模型 (Explainable Boosting Machine, EBM)
    5. 支持向量機 (Support Vector Machine, SVM)
- 深度學習方法
    1. 人工神經網路 (Artificial Neural Network, ANN)
    2. 貝葉斯神經網路 (Bayesian Neural Network, BNN)

doi:10.6342/NTU202502436

3.3.1 模型架構與特點

本小節將逐一介紹各模行之架構邏輯與特性差異，作為後續模型訓練結果之比較依據。

隨機森林 (Random Forest, RF)

1. 基本架構與核心邏輯

隨機森林為由多棵決策樹所組成的集成式學習方法。其訓練過程透過袋外抽樣法 (Bootstrap Aggregation, Bagging) 對訓練資料與特徵進行隨機抽樣，並在各樹中構建獨立模型，最終預測結果透過多棵樹的平均值 (迴歸問題) 或多數投票 (分類問題) 進行結果輸出，如圖 3.8。

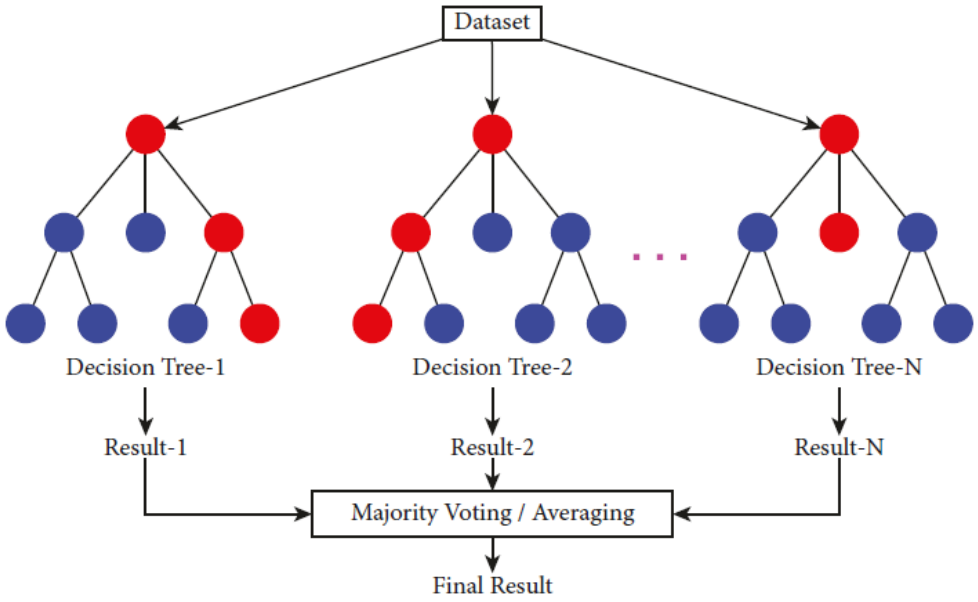


圖 3.8 Random Forest 架構示意圖 (Khan et al., 2021)

2. 優勢與限制

- 優點包括:

(1) 不須特徵標準化處理

(2) 可處理高維資料與多變數

doi:10.6342/NTU202502436

(3) 對異常值與部分缺失值具容忍性

(4) 可估計特徵重要性以輔助解釋模型

- 其限制為:

(1) 較難解釋單顆樹的決策過程，整體為黑箱模型

(2) 在處理極度不平衡資料或過多噪音時可能表現不佳

(3) 訓練時間與資源需求隨樹數增加而上升

3. 適用性

本研究資料來源異性值高，包含多筆具缺失值與非線性特徵的土壤試驗數據。隨機森林可有效捕捉複雜的變數關係，並具備一定容錯能力，適合作為本研究之基礎預測模型，應用於預測循環剪應力比 (CSR)。

分類增強模型 (Categorical Boosting, CatBoost)

1. 基本架構與核心邏輯

CatBoost (Categorical Boosting) 為一種基於梯度提升決策樹 (Gradient Boosted Decision Trees, GBDT) 所發展之改良演算法，核心理念為透過逐步建構一系列弱機器學習器 (通常為決策樹)，使每一步的預測誤差 (Residual) 能由下一棵樹所補償，如圖 3.9。其演算法流程與傳統 GBDT 類似，透過梯度下降法 (Gradient Descent) 在函數空間中尋找最小損失函數 (Loss Function)。

相較於傳統 GBDT，CatBoost 引入兩項重要步驟以解決實務應用中常見的問題 (Prokhorenkova, 2017; Hancock & Khoshgoftaar, 2020):

(1) 有序增強 (Ordered Boosting)

傳統 GBDT 在每棵樹的訓練過程中使用整體訓練資料計算梯度，這種做法在包含類別變數的情境下，容易造成目標洩漏 (target leakage)，即模型在尚未正式訓練完成前即已接觸到未來資訊，導致預測偏移與過擬合。

CatBoost 提出排序提升 (Ordered Boosting) 以解決此問題，將對訓練資料依序切分為多個資料子集，在每次構建弱學習器時，僅使用當前樣本「之前」的資

doi:10.6342/NTU202502436

料來估計梯度與統計量。此策略能模擬模型實際預測未知資料時的反應，有效降低資訊洩漏並提升模型泛化能力，尤其在樣本數有限或類別變數比例偏高的資料集中表現尤為穩定。

(2) 類別特徵編碼策略 (Categorical Feature Encoding Strategy)

傳統 GBDT 模型在處理類別變數時，須先將其轉換為數值格式，常見方法如 One-Hot Encoding。然而，當類別數量眾多時，One-Hot Encoding 容易導致維度災難 (curse of dimensionality)，進而影響模型效能。

CatBoost 採用了具備順序限制 (ordered constraint) 的編碼策略，於每筆樣本中僅使用先前樣本的統計資訊 (例如該類別在前面樣本中出現的次數) 來計算編碼值，避免引入當前樣本的標籤資訊。此技術在保留類別資訊辨識力的同時，有效降低過擬合風險，尤其適用於高基數 (high cardinality，一個類別型特徵中包含了許多不同類別) 或樣本分佈不均的資料。

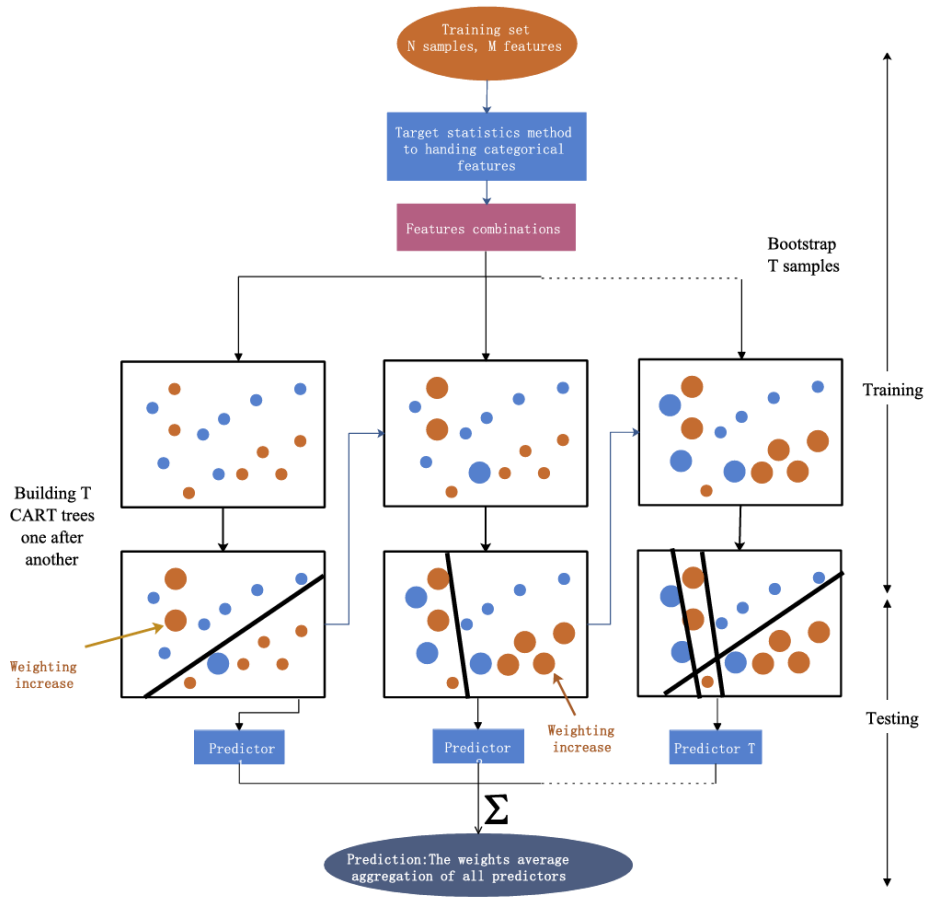


圖 3.9 CatBoost 架構示意圖 (Huang et al., 2019)

doi:10.6342/NTU202502436

2. 優勢與限制

- 優點包括:

(1) 支援原生類別型變數

(2) 可處理高基數類別特徵，避免維度災難

(3) 採用 Ordered Boosting，可降低目標洩漏與過擬合風險

(4) 對於非線性特徵關係建模效果良好，預測效能高

(5) 不須進行特徵標準化處理

- 其限制為:

(1) 模型結構相對複雜，邏輯較不易完全解析

(2) 參數眾多，調整需耗費時間與資源

(3) 在樣本極度不平衡或資料高度稀疏 (highly sparse，大多數特徵值都是空值或零) 情況下，仍需額外處理

(4) 若資料集龐大，訓練與推論時的記憶體需求相對較高

3. 適用性

本研究所使用之土壤實驗資料具備非線性和有部分缺失值等特性，且數據來源異質性 (heterogeneity) 高，資料分佈不均。CatBoost 原生支援類別型欄位處理並具備 Ordered Boosting 架構，能有效避免 One-Hot Encoding 造成的維度災難及過擬合問題，對於非均勻分佈與小樣本類別亦具穩定性，故適合作為本研究之預測模型之一。

極端梯度提升機 (Extreme Gradient Boosting, XGBoost)

1. 基本架構與核心邏輯

XGBoost (Extreme Gradient Boosting) 為建立於傳統 GBDT 基礎上的強化版演算法，由 Chen and Guestrin (2016) 所提出，具備高效能與強大彈性。其核心概念為以串連方式疊加多個弱學習器，透過每次對前一輪預測誤差的學習進行修正，同樣利用梯度下降法在函數空間中持續優化損失函數。模型最終輸出為所有

doi:10.6342/NTU202502436

樹預測值的加權總和，而非多數表決 (如 Random Forest)，其架構如圖 3.10 (Wang et al., 2020; Faska et al., 2023; Öztornaci et al., 2024)。

XGBoost 相較於傳統 GBDT 在效能與泛化能力方面皆有顯著提升，其主要技術特色包括:

- 正則化項 (Regularization): 透過懲罰複雜模型結構 (例如過多分裂或過深的樹)，以避免過擬合，使預測結果更加平滑且穩健。
- 二階梯度資訊 (Hessian): 每次訓練時不僅計算一階導數 (預測誤差的方向)，還同時利用二階導數 (損失函數的曲率) 來建構分裂點與葉節點 (leaf node) 的最佳化。這使得每一步的模型更新更精準，收斂速度更快，也有助於穩定性提升。
- 內建缺失值處理: 自動學習樣本缺失值的最佳分裂方向，無須事前補值處理，能針對高維度資料與不完整數據進行高效率運算。
- 序列是串聯結構: 並行計算與快取機制，優化訓練時間與記憶體效率，適用於大規模資料集。

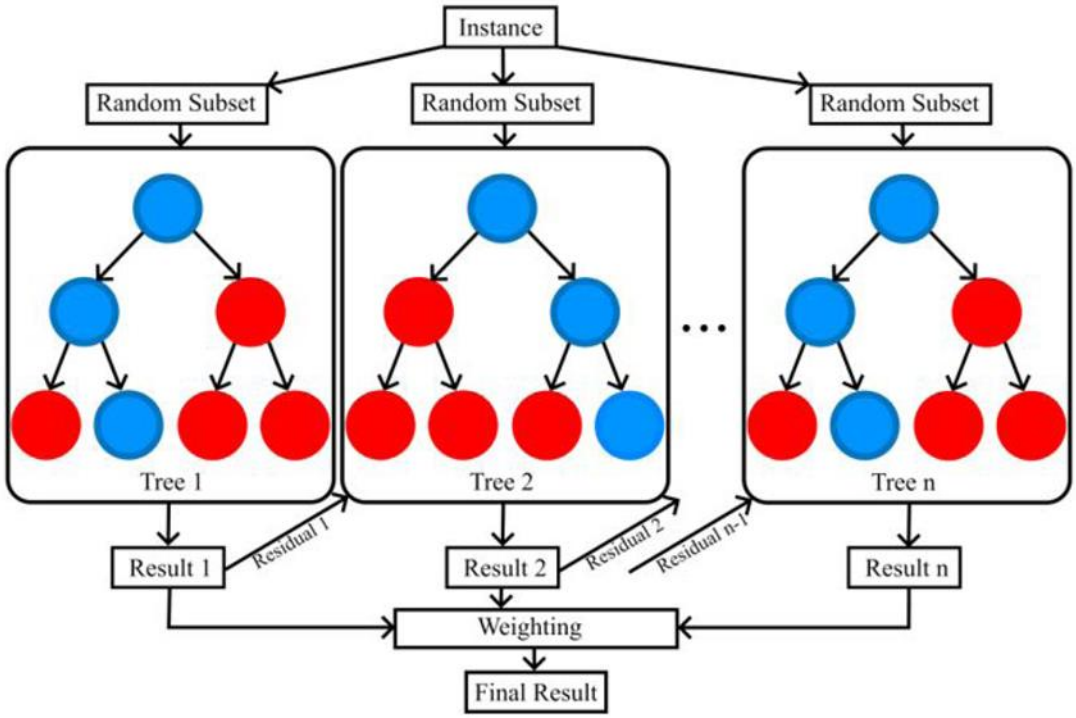


圖 3.10 XGBoost 架構示意圖 (Öztornaci et al., 2024)

doi:10.6342/NTU202502436

2. 優勢與限制

- 優點包括:
  (1) 具高預測準確性與穩定性
  (2) 模型泛化能力強，能處理稀疏、高維與部分缺值的資料，適用於多種複雜資料情境
  (3) 模型架構中具多項參數，可精細的對應不同分析需求
  (4) 具良好的可解釋性，能輸出個特徵的重要性，有助於後續的特徵選擇與分析
- 其限制為:
  (1) 模型參數繁多，需花時間進行調參與驗證，對初學者而言較具挑戰性
  (2) 在資料量有限或雜訊較多的情況下仍有過擬合風險
  (3) 雖適用於非線性建模，對連續且平滑變化的資料擬合效果仍不及神經網路

3. 適用性

本研究資料來自不同文獻，具有一定的異質性與不均勻性，部分欄位也存在缺失值，同時特徵之間可能具有非線性與交互關係。XGBoost 在這類結構化資料中表現穩定，具備良好的非線性擬合能力，能有效捕捉特徵間的潛在關聯，因此適合用在本研究的預測分析中。

可解釋增強模型 (Explainable Boosting Machine, EBM)

1. 基本架構與核心邏輯

(Explainable Boosting Machine, EBM) 是近年發展的可解釋性機器學習，屬於白箱 (glass-box) 模型的一種，能兼顧預測正確率與模型可理解性。其核心為廣義加法模型 (Generalized Additive Model, GAM) 的擴充形式，透過逐一學習各個特徵對目標變數的影響，並以低學習率 (learning rate) 進行反覆訓練，同時融合梯度提升 (Boosting) 技術以降低特稱共線性的影響並提升模型穩定性。EBM 的

doi:10.6342/NTU202502436

預測結果由各變數的貢獻值相加而成，因此每個特徵對預測結果的影響都可以被單獨觀察與解釋，如圖 3.11。此外，EBM 不同於傳統的 GAM 架構，其訓練過程中可自動偵測具有預測價值的特徵交互作用 (pairwise interactions)，如圖 3.12，進一步提升模型的預測準確性與實用性 (Hernández et al., 2023; Wahab et al., 2024)。

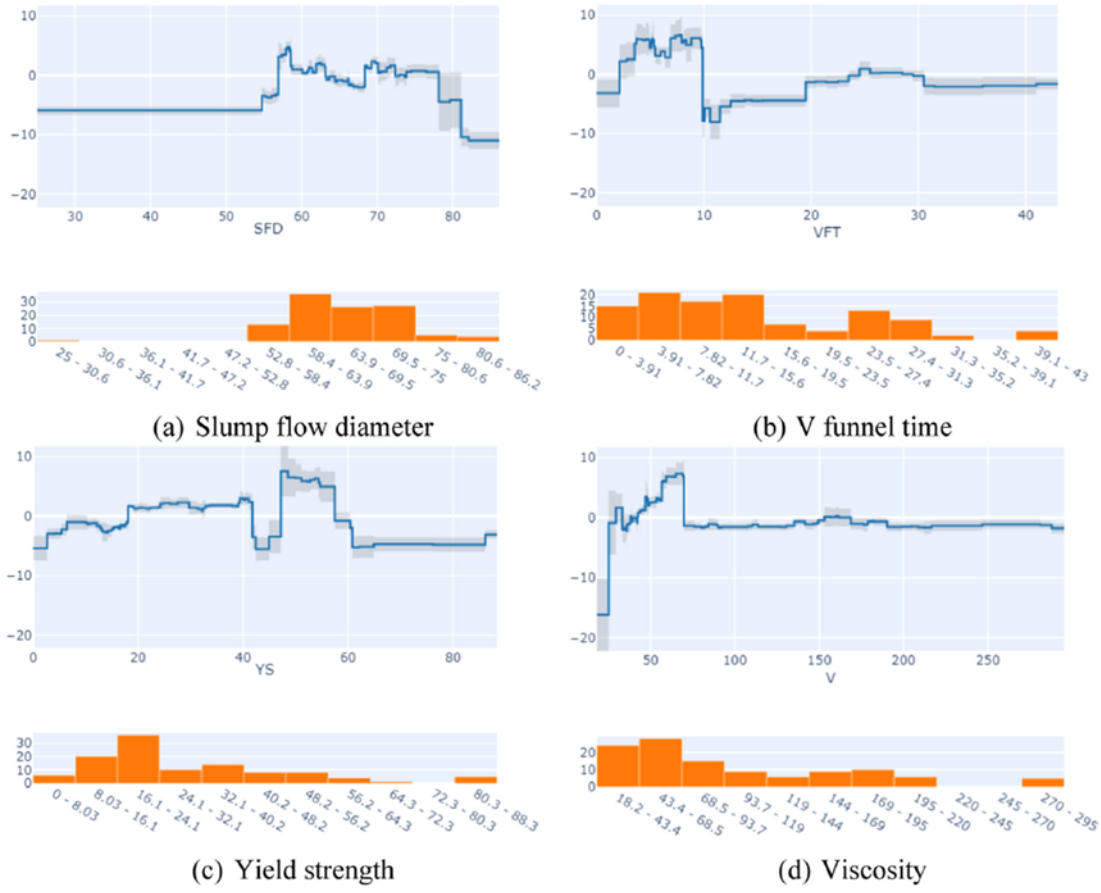


圖 3.11 EBM 中貢獻值的全域解釋 (global explanation) 示意圖 (Wahab et al., 2024)

doi:10.6342/NTU202502436

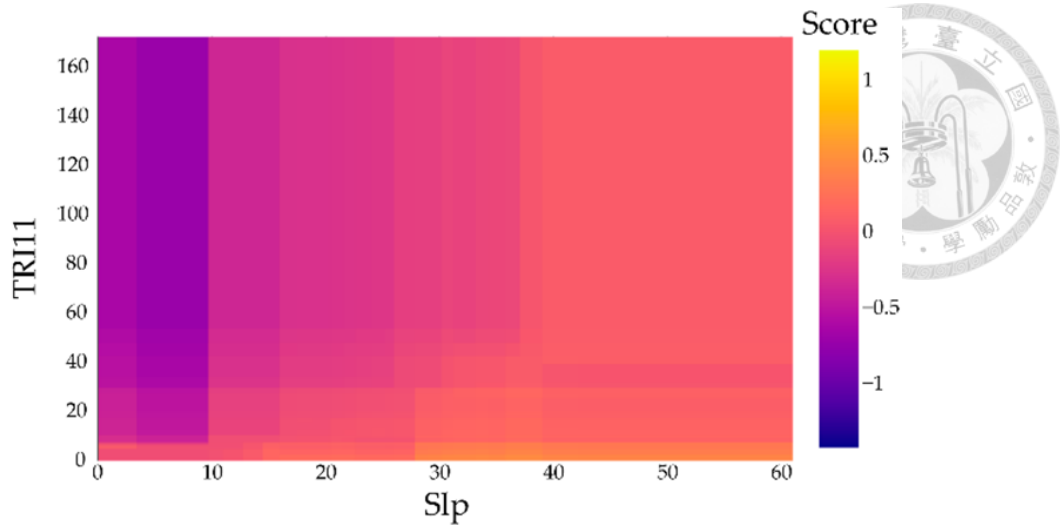


圖 3.12 EBM 特徵交互關係示意圖 (Maxwell et al., 2021)

2. 優勢與限制

- 優點包括:

(1) 模型本身具高度可解釋性，可視化每個特徵對預測結果的貢獻

(2) 採用逐特徵訓練與低學習率，有助於降低特徵共線性帶來的影響

(3) 可自動偵測具預測價值的特徵交互作用

(4) 適用於結構化資料，對非線性關係與多變數影響具良好處理能力

- 其限制為:

(1) 相較於其他樹模型 (如 XGBoost、CatBoost)，訓練速度較慢

(2) 整體模型結構仍偏簡單，面對高度複雜資料時表現可能受限

(3) 目前主要支援迴歸與二元分類分析，對多類別問題與非標準目標函數的支援性較低

3. 適用性

本研究使用之資料具有一定程度的異質性與非線性特徵，變數之間亦可能存在交互作用。EBM 模型透過逐一學習各變數的影響，並可自動偵測重要的特徵交互組合，使其在處理此類複雜關係時表現穩定且具可解釋性。此外，EBM 可視化每個變數對預測結果的貢獻，有助於釐清影響 CSR 的關鍵因子，對本研究在模型解釋與工程應用層面皆具實質幫助，因此被納入本研究之預測模型之一。

doi:10.6342/NTU202502436

支持向量機 (Support Vector Machine, SVM)

1. 基本架構與核心邏輯

SVM (Support Vector Machine) 是一種常用於分類與迴歸問題的監督式學習方式，其核心概念為在特徵空間中尋找一個最適合分隔資料的超平面(hyperplane)，使資料點在空間中具有最大的邊界間隔，如圖 3.13。在迴歸分析中，SVM 則以支持向量迴歸 (Support Vector Regression, SVR) 的形式運作，目標是找到一個容許誤差範圍內最佳的擬合函數。在進行非線性問題建模時，SVM 可透過核函數 (kernel，如 RBF、polynomial 等) 將原始資料映射到高維空間，以達到非線性可分的效果。SVM 的學習過程中，會自動選出少數幾個靠近邊界 (margin) 的資料點 (即 support vectors) 會被用來決定最終的決策邊界，使模型具備良好的泛化能力與數學穩定性 (Manjrekar & Duduković, 2019)。

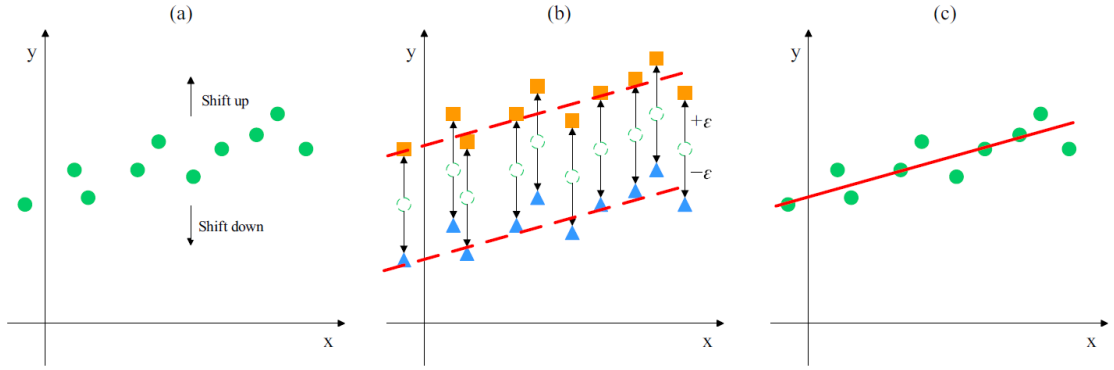


圖 3.13 SVM 線性迴歸問題 (a) 原始資料集其移動方向 (b) 移動資料與ε−區間 (ε-tube) (c) 最終迴歸平面 (Zhang et al., 2021)

2. 優勢與限制

- 優點包括:

(1) 對高維資料具良好處理能力

(2) 可透過不同核函數對應線性與非線性問題，彈性高、適用性廣

(3) 模型結構簡潔，僅依賴 support vectors 建構決策邊界，有助於避免過擬合問題

(4) 支援正規化與誤差容忍參數設定，使用者可依需求控制模型複雜度與泛化能力

doi:10.6342/NTU202502436

- 其限制為:

  (1) 面對大型資料集時，訓練時間與計算成本相對較高

  (2) 對雜訊資料較為敏感，前處理與特徵篩選的品質會影響模型穩定性

  (3) 須先進行類別型特徵的編碼處理

3. 適用性

本研究所使用之資料以連續型數值特徵為主，特徵間可能存在非線性關係，且樣本數屬中等規模。SVM 可透過不同核函數處理非線性資料分佈，具備良好的泛化能力與預測穩定性，特別適合應用於此類高維度但樣本量有限的情境中。由於 SVM 僅依賴少數 support vectors 建構模型，能有效抑制過擬合風險，對於本研究中 CSR 預測分析具高度適用性，故納入作為本研究之預測模型之一。

人工神經網路 (Artificial Neural Network, ANN)

1. 基本架構與核心邏輯

ANN (Artificail Neural Network) 是一種模仿人類神經元運作方式所建構的機器學習模型，其基本結構包含輸入層 (Input layer) 、隱藏層 (Hidden layer) 與輸入層 (Output layer)，每層由多個神經元 (neuron) 組成。每個神經元會接收前一層的輸出，並透過加權係數與偏差項 (bias) 進行線性加總後，輸入激活函數 (activation function)，產生非線性輸出訊號再傳遞至下一層，如圖 3.14 (a)。若輸入層有 n 個變數，則任依隱藏層中的神經元 j 會接收所有輸入，計算方式如下:

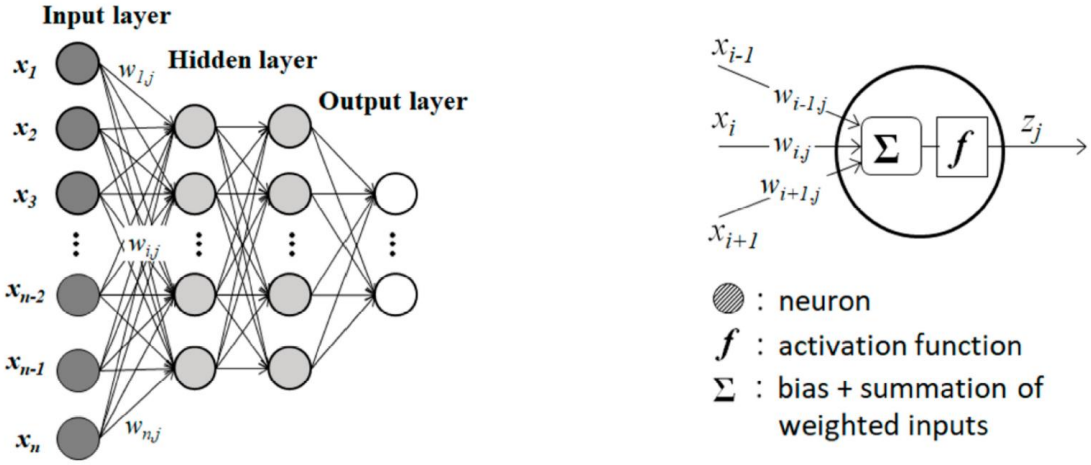
$$y_j = b_j + \sum_{i=1}^{n} w_{i,j} x_i \tag{3.4}$$

$$z_j = f(y_j) \tag{3.5}$$

其中，x_i為第 i 個輸入變數，$w_{i,j}$為對應權重 (weight)，b_j為偏差項，y_j為該神經元尚未激活的輸入總和，f為激活函數 (常見如 ReLU、sigmoid 或 tanh)，z_j為第 j 個神經元的輸出值，如圖 3.14 (b)。

doi:10.6342/NTU202502436

ANN 透過反向傳播 (Backpropagation) 與梯度下降法調整參數，以最小化損失函數，進而提升預測正確度。ANN 屬於非參數模型，不需事先假設資料分佈或數學關係，能有效建構複雜的非線性模型，廣泛應用於分類與迴歸等分析中 (Kimura et al., 2019)。



(a) 深度學習中的資料流程　　(b) 神經元的內部結構

圖 3.14 ANN 模型結構示意圖 (Kimura et al., 2019)

2. 優勢與限制

- 優點包括:
  - (1) 模型架構彈性高，可依資料特性調整隱藏層層數與神經元數量
  - (2) 可處理複雜的非線性關係，對於資料間隱含結構具有良好的辨識能力
  - (3) 不需事先假設資料分佈或變數間的數學關係，適用於多元且結構不明確的問題
  - (4) 可結合不同激活函數與正則化技術，增強模型穩定性與泛化能力
  - (5) 廣泛應用於迴歸、分類與時間序列預測等各類問題，具備高度通用性
- 其限制為:
  - (1) 需要大量資料與充足訓練次數，否則容易出現過擬合或欠擬合問題
  - (2) 超參數設定繁多，需經多次調整與驗證

doi:10.6342/NTU202502436

(3) 模型為黑盒結構，難以清楚解釋各特徵對預測結果的貢獻

(4) 訓練過程對硬體資源需求較高

(5) 對資料尺度與標準化敏感，需進行適當的標準化處理

3. 適用性

本研究資料涵蓋多個連續型工程參數，變數間可能存在高度非線性與複雜交互關係，且資料來源異質性高、分佈不一。ANN 為非參數模型，具備處理非線性資料與多元特徵組合的能力、架構具彈性，且能依據資料規模與特性進行調整，並透過激活函數與正則化策略抑制過擬合風險。綜合而言，ANN 能夠因應本研究資料的多樣性與非線性特性，故納入作為本研究之預測模型之一。

貝葉斯神經網路 (Bayesian Neural Network, BNN)

1. 基本架構與核心邏輯

BNN (Bayesian Neural Network) 為人工神經網路的一種擴充形式，結合了貝葉斯統計概念，將模型中的權重與偏差視為機率變數，透過機率分佈進行推論。與傳統 ANN 僅學習一組固定參數不同，BNN 會為每一個權重與偏差建立後驗分佈 (posterior distribution)，透過多次從該分佈中取樣進行向前傳遞 (forward pass)，再將結果平均以得出預測值，如圖 3.15 (Häse et al., 2019; Mahajan et al., 2024)。

本研究採用變分推論 (Variational Inference) 方法，透過設計一個具特定形式 (如高斯分佈) 的候選分佈，不斷調整其參數，以逼近真實後驗分佈，並進行近似推導。訓練過程中以最小化 KL 散度 (Kullback-Leibler Divergence) 為目標，優化代表參數分佈的平均與標準差。其數學定義如式(3.6)所示 (Belov & Armstrong, 2011) :

$$D_{KL}(g \parallel h) = \int_{-\infty}^{+\infty} g(\theta) ln \frac{g(\theta)}{h(\theta)} d\theta \tag{3.6}$$

doi:10.6342/NTU202502436

其中，

- $g(\theta)$: 為變分分佈，用來近似真實的後驗分佈
- $h(\theta)$: 為在觀測資料下的真實後驗分佈
- θ: 代表模型參數 (如權重與偏差)
- D_{KL}: 為 KL 散度，用以衡量兩個分佈之間的差異

在變分推論中，透過梯度下降法持續最小化D_{KL}，使模型得以逼近真實後驗分佈，進而實現有效的不確定性估計。

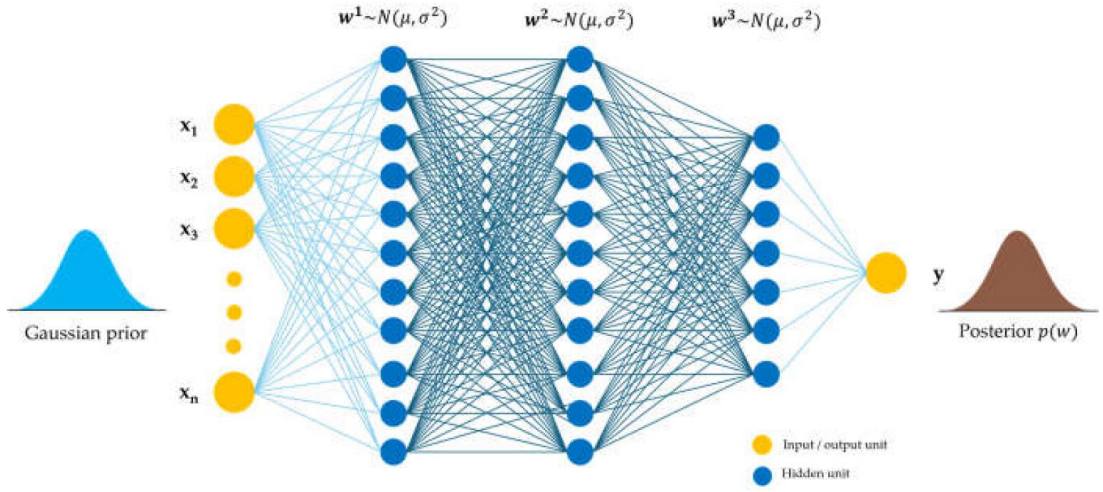


圖 3.15 BNN 架構示意圖 (Mahajan et al., 2024)

2. 優勢與限制

- 優點包括:
  - (1) 除了輸出預測值外，能同時提供不確定性
  - (2) 保有傳統神經網路處理非線性與多維輸入的強大建模能力
  - (3) 模型輸出為機率分佈，可推估可信區間 (confidence interval)，適合用於資料異質或樣本數有限的問題
  - (4) 在多次前向傳遞下具備更強的樣本表現穩定性與泛化能力
- 其限制為:
  - (1) 模型架構與實作較為複雜

doi:10.6342/NTU202502436

(2) 訓練過程需額外估計參數分佈，相較於傳統 ANN 計算負擔較高

(3) 需多次向前傳遞 (forward pass) 才能產生穩定預測，對運算資源要求較高

3. 適用性

本研究的資料特性包含變數多、關係複雜，以及來源豐富，且樣本數並不算多。BNN 能夠處理這類非線性與資料不均的情況，同時在預測時提供結果的範圍與不確定性，讓模型不只給出數值，還能反映預測的信心水準。整體而言，BNN 適合應用於本研究這類資料結構多樣、且希望同時了解預測可信度的情境，因此被納入作為本研究的預測模型之一。

3.4 模型訓練與評估方式

在機器學習麼型的建立與訓練過程中，超參數的設定與模型效能的評估為兩項核心要素。超參數 (hyperparameter) 指的是在模型訓練前由使用者指定，非自動學習而得的參數，不同的超參數組合將導致模型在學習過程中出現不同的收斂速度、泛化能力與預測表現，因此透過適當的調整找出需針對本研究資料的最佳超參數組合，以提升模型性能。

另一方面，模型訓練完成後，需透過適當的評估指標 (evaluation metrics) 來衡量模型的表現。評估指標能客觀反映模型對預測與測試資料的預測準確程度，並協助辨識是否存在過擬合、欠擬合或模型不穩定等問題。不同分析類型所適用之評估指標意有所不同，因此選擇適合的指標將有助於更全面地了解模型的好壞。

為此，本章節將依序說明本研究所涉及之超參數與模型評估指標。

doi:10.6342/NTU202502436

3.4.1 超參數

本研究所採用之機器學習模型涵蓋多種演算法，其訓練過程皆需預先設定多項超參數以進行訓練與調整。表 3.3 至表 3.9 分別整理各模型所考慮之超參數，包含其預設值、定義與詳細說明，作為本研究後續調參與模型比較之依據。

表 3.3 Random Forest 之超參數

代碼名稱	預設值	定義	說明
n_estimators	100	決策樹的數量	森林中樹的數量。數量太多計算成本高，且容易過擬合；數量太少可能會欠擬合。
max_depth	None	決策樹的最大深度	限制每棵決策樹的最大深度。深度太大會過擬合；深度太小會欠擬合。
min_sample_split	2	一個節點需要至少多少筆樣本，才能進行下一次分裂	增加可避免生成許多小樹枝，降低過擬合機率，太大會使模型欠擬合。
min_samples_leaf	1	葉節點上保留多少筆樣本	葉節點有一定的樣本數，可使模型更平滑避免太極端的小節點，在訓練不平衡的資料集時會很有幫助。
max_features	auto	每次節點分裂時所考慮的特徵數量	減少每次考慮的特徵數可增加樹的多樣性，同時減少過擬合的發生。
n_jobs	None	設定可平行運算的CPU核心數	設定為“-1”時將使用所有可用的核心，可大幅加速訓練速度。
verbose	0	控制訓練時輸出過程資詳細程度	用於觀察模型訓練進度。

doi:10.6342/NTU202502436

表 3.4 CatBoost 之超參數

代碼名稱	預設值	定義	說明
iterations	1000	Boosting的總回合數	Boosting流程中逐步加入的樹的數量。每棵樹是根據上一步的殘差或梯度來訓練，樹越多模型越複雜，越容易過擬合；樹越小，模型越簡單，越容易欠擬合。
depth	6	決策樹的最大深度	(同Random field的max_depth)
learning_rate	0.03	每次更新時，數對整體模型影響的縮放因子	小的值需要更多iterations，模型收斂得更穩、更準；大的值模型收斂快，但容易過擬合。
l2_leaf_reg	3	L2正則化係數	用於懲罰葉節點的權重。提高可防止模型過擬合，使葉節點上的預測值更保守。
random_strength	1	隨機性程度	控制在建樹時特徵選擇的隨機強度。增加隨機性可助於避免過擬合發生。
border_count	254	分箱數	將連續型特徵分成多少段，以便後續建樹時做二元切分。與EBM的max_bins不同的是使用特定演算法選出最佳分界點。值越大，分得越細，模型能學到的細節越多，越容易過擬合；值越小，分得越粗，模型越簡單，較能避免過擬合。
cat_featrues	(需手動設定)	指定類別型資料	CatBoost內建特有的類別特徵編碼，需指定正確，否則將會被當成數值處理。
tack_type	CPU	指定執行裝置	可選CPU或GPU。使用GPU可大幅增加訓練速度，但某些功能有限制或表現不同。

doi:10.6342/NTU202502436

表 3.5 XGBoost 之超參數

代碼名稱	預設值	定義	說明
objective	reg:squarederror / binary:logistic	損失函數的類型	回歸分析用reg:squarederror；二分類分析用 binary:logistic；多類別問題用multi:softpro。
n_estimators	100	決策樹的數量	Boosting的總回合數。樹越多模型越複雜，越容易過擬合；樹越小，模型越簡單，越容易欠擬合。
learning_rate	0.3	每次更新時，數對整體模型影響的縮放因子	值越小，模型越穩定，但須越多棵樹，n_estimators越多；值越大，訓練快，但可能過擬合。
max_depth	6	決策樹的最大深度	決定每棵樹的複雜度。深度大模型複雜，容易過擬合；深度小模型簡單，容易欠擬合。
min_child_weight	1	節點上樣本總權重的最小值	低於此值則節點不再分裂。提高此值可使模型更保守，避免學習雜訊。
gamma	0	節點分裂時所需的最小損失函數減少值	值越大節點越不容易分裂，即增加此值可控制模型複雜程度，以降低過擬合風險。
reg_alpha	0	L1正則化 (悉疏正則化) 項的權重	可強制讓部分特徵權重趨近於0，產生稀疏模型，達到特徵選擇的效果，並可提高模型的可解釋性，減少對雜訊特徵的依賴。
reg_lambda	1	L2正則化 (平滑正則化) 項的權重	可平滑所有特徵權重，使模型避免對單一特徵的過度依賴，可有效降低過擬合風險，但不會造成特徵淘汰。
subample	1	每顆樹訓練時，隨機選取樣本的比例	值小於1將會引入隨機性，降低資料間的相依性，可幫助防止過擬合。
colsample_bytree	1	每顆樹訓練時，隨機選取特徵的比例	增加模型的多樣性，與subsample搭配使用能提升模型泛化能力。
enable_catergorical	FALSE	是否啟用原生的類別型特徵處理功能	若資料中有純類別欄位，可設為True。自動學習分類特徵與目標變數的關係。
tree_method	auto	以建構每棵樹的演算法策略	可針對訓練資料選擇不同演算法方式，不同方法在訓練速度、資源消耗和適用資料大小上會有所不同。

doi:10.6342/NTU202502436

表 3.6 EBM 之超參數

代碼名稱	預設值	定義	說明
max_bins	255	分箱數	將連續型特徵分成多少個區間，然後以這些區間為單位進行學習。與CatBoost的border_count不同的是使用等寬分箱或基於梯度統計做最佳分割。值越大，分得越細，模型能學到的細節越多，越容易過擬合且成本越高；值越小，分得越粗，模型越簡單，較能避免過擬合。
max_interaction_bins	32	建構特徵交互作用時，使用的分箱數	針對互動特徵做分箱。分箱多容易過擬合；分箱少，模型穩定但捕捉互動關係能力下降。
interactions	10	是否自動探索並納入二階 (或高階) 特徵交互作用	模型會選出表現最好的interactions個交互作用特徵並加入模型中，可提升模型準確率，但也會降低解釋性，若希望保持高度解釋性可設為'0'。
learning_rate	0.01	每次更新時，數對整體模型影響的縮放因子	控制每次迭代中，對特徵影響力更新的幅度。值越小，需要迭代越多輪，但訓練越穩、收斂越穩定。
max_rounds	5000	最大的boosting回合數	決定最大訓練次數的上限。可配合early_stopping_rounds使用。
early_stopping_rounds	50	超過設定回合數訓練誤差沒改善，則提前停止訓練	可加速訓練並防止過擬合。若資料雜訊大，可設大值；若資料乾淨，則可設小一點。
min_samples_leaf	2	每個分箱 (bin) 最少需要的樣本數	若分箱後每個區間樣本太少，這個分箱會被合併或是忽略，可以防止過擬合，特別是當資料中有異常值時。
max_leaves	3	每個特徵函數 (特徵貢獻曲線) 中允許的最大節點數	控制特徵函數的複雜度 (曲線可以有多少轉折點)，節點數越多，模型能擬合更複雜的模式，但容易過擬合。

doi:10.6342/NTU202502436

表 3.7 SVM 之超參數

代碼名稱	預設值	定義	說明
C	1	誤差懲罰項 (regularization parameter)	控制模型對錯誤分類或預測誤差的容忍度。值越大對錯誤越不能容忍，傾向擬合訓練資料，容易過擬合；值越小越可以容忍一定錯誤，模型簡單，泛化能力好。
kernel	rbf	核函數	決定將資料投影到什麼樣的高維空間進行分割。rbf: 高斯核 (高斯逕向基底函數) 最常用，適合非線性資料；poly: 多項式核；linear: 線性核，效果像線性回歸或線性分類；sigmoid: 類似神經網路的激活函數。
epsilon	0.1	在ε-容許區間 (epsilon-tube)內的預測誤差不會被計入損失函數 (loss) 中	設定一容忍區間，只對超出這個範圍的誤差進行懲罰。值越大，預測結果更平滑，但可能漏掉小變化；值越小，可更精確擬合，但容易受雜訊影響。
gamma	scale	控制單個訓練樣本影響範圍	在rbf、poly、sigmoid核中使用，可輸入定值如0.01、0.1、1等。值越大，單個樣本的影響範圍越小，模型越複雜，越容易過擬合；值越小，單個樣本的影響範圍越大，模型越平滑。
degree	3	多項式核函數poly的次數	決定多項式核的多項式階數。次數高。可以擬合更複雜的邊界，但可能會導致過擬合。
coef0	0	核函數中的常數項	在poly和sigmoid核中使用。在poly中控制高階多項式與低階多項式的影響平衡；在sigmoid中控制核的曲線偏移。值越大，越重視高次項，決策邊界越複雜；值越小，越接近線性邊界。

doi:10.6342/NTU202502436

表 3.8 ANN 之超參數

代碼名稱	預設值	定義	說明
hidden_layers	(需自行定義)	隱藏層的數量與每層神經元的數目	通常以代碼中的list呈現。增加層數與神經元能提高模型容量，以學習更複雜的關係。層數太少會欠擬合；層數太多，容易過擬合且訓練困難。
activations	linear	每層隱藏層使用的激活函數	引入非線性能力，讓模型能學習複雜的線性關係。隱藏層建議使用ReLU，輸出層若為回歸問題，建議使用linear；若為二元分類問題，建議使用sigmoid；若為多元分類問題，建議使用softmax。
dropout_rates	(需自行定義)	每層的dropout機率	用於隨機停用部分神經元。為正則化手段，有效防止過擬合。建議在隱藏層後加，輸出層通常不加入dropout。
use_batch_norm	FALSE	是否使用批次正規化(Batch Normalization)	讓每層輸出維持穩定的分佈，加速訓練、穩定收斂。在較深的模型或是資料特徵範圍變化大時效果較為顯著。Batch Normalization是一種在神經網路訓練過程中，對每層輸入進行標準化(使均值為0，變異數為1) 的方式，能夠加速模型的收斂和提高穩定性，同時具有一定的正則化效果。
optimizer	rmsprop	優化器種類	用來更新權重參數以最小化損失。為一種演算法，根據損失函數的梯度調整模型參數，以達到最小化損失的目標。不同優器對收斂速度、穩定性與最終模型效果有顯著影響。
loss_fn	(依分析類型自動推論)	損失函數	訓練模型時追求的最小化目標函數。若為回歸問題，常使用mse、mae、huber；若為二元分類問題，常使用binary_crossentropy；若為多元分類問題，常使用categorical_crossentropy或sparse_categorical_crossentropy。
metrics	(需自行定義)	除損失函數外，用來監控訓練與驗證效果的評估指標	若為回歸問題，常使用mse、mae、自訂之r2等；若為分類問題，常使用accuracy、precision、recall、AUC等。
batch_size	32	每次訓練所取用的資料筆數	值越小，模型學習得越穩定，訓練時間越長；值越大，訓練速度越快，但可能學不到細節。
epochs	(需自行定義)	模型訓練完整資料集的次數	每次epoch結束後，模型的參數（如權重與偏+A1:D10差）會更新過一次，接著進入下一個epoch。多次epoch可讓模型逐步收斂至更加的參數組合，但過多的epoch會導致過擬合。

doi:10.6342/NTU202502436

表 3.9 BNN 之超參數

代碼名稱	預設值	定義	說明
hidden_layers	(需自行定義)	隱藏層的數量與每層神經元的數目	(同ANN的hidden_layers)
use_dropout	FALSE	是否啟用Dropout	控制是否在每層加入dropout機制，以降低過擬合風險。設定為True時，每層將以指定機率隨機停用部分神經元，提升模型泛化能力。
activation_fn	ReLU	每層的激活函數	引入非線性能力，讓模型能學習複雜的線性關係。實務上通常使用ReLU，因其具稀疏性與梯度穩定性。
learning_rate	0.001	每次更新時，數對整體模型影響的縮放因子	決定變分後驗分佈 (如均值與標準差) 更新的幅度。值越小收斂越穩定，但訓練時間較長；過大可能導致不穩定或發散。通常需與 KL 項權重共同調整以達穩定學習。
epochs	(需自行定義)	模型訓練完整資料集的次數	(同ANN的epochs)
optimizer_type	Adam	優化器種類	用來更新權重參數以最小化損失。推薦使用Adam，因模型中參數為分佈 (有均值和標準差)，收斂難度更高，Adam的自適應特性有助於收斂。
dropout_rate	(需自行定義)	每層的dropout機率	控制每層dropout的比例，用來降低過擬合風險。設定為0.0表示不使用dropout，常見值為0.1~0.5。
loss_fn	(需自行定義)	損失函數	包括誤差項與KL散度。在BNN中，Loss通常為「預測誤差+KL散度」。

另外補充節點 (node) 與葉節點 (leaf node) 的差異。在決策樹中，節點為整體樹結構中的分裂點，每個節點是根據某一特徵的切分條件將資料做分流；葉節點則為不再進行分裂的終端節點，用於儲存最終的預測值。

透過對各模型常用超參數的整理與說明，將有助於理解其對模型效能的潛在影響，並作為後續參數調整策略與模型效能分析的基礎。

doi:10.6342/NTU202502436

3.4.2 參數調整與交叉驗證

在建立機器學習模型時，參數設定對模型表現具有顯著影響。若未妥善調整參數，可能導致模型過度簡化或過度擬合，降低預測正確度與泛化能力。為提升模型效能與穩定性，機器學習領域中常見的參數調整方式包括手動調參 (manual tuning)、隨機搜尋 (Random search) 與網格搜尋 (Grid search)，而模型效能驗證則多透過交叉驗證 (如 K-折交叉驗證) 進行，具體說明依序如下:

手動調參 (Manual tuning)

在模型初期建構階段，本研究以手動方式進行超參數設定，根據資料特性與模型架構，調整如學習率、隱藏層數、神經元數量、樹深等關鍵超參數，並觀察分析結果的變化趨勢。此方式雖仰賴經驗與實驗回饋，但可作為後續系統化搜尋的依據，同時亦有助於加深對各模型特性的理解與掌握。

隨機搜尋 (Random search)

隨機搜尋 (Random search) 為最早且實作簡易的超參數優化 (hyperparameter optimisation) 方法之一，其核心概念為在參數空間中隨機抽樣若干組參數組合進行測試，不需針對每個參數離散化或遍歷所有組合，如圖 3.16，因此在高維空間中能更有效率得找尋潛在解。與網格搜尋 (Grid search) 相比，隨機搜尋能避免將過多資源分配於對模型影響較小的參數上，進而提升搜尋效率。

儘管 Random search 效率高且具備一定隨機性，但其結果不具完整性，不保證涵蓋所有潛在的優化區域，因此在實務上常需搭配多次試驗與交叉驗證，以提升結果穩定性與可重現性 (Iqbal et al., 2023)。

doi:10.6342/NTU202502436

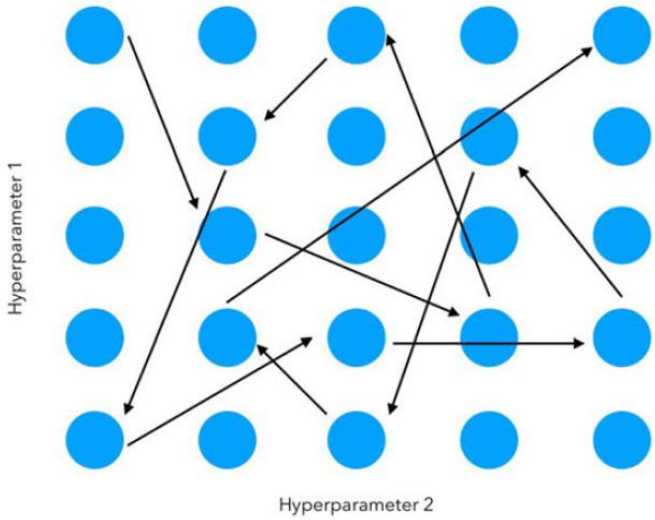


圖 3.16 Random Search 隨機挑選超參數組合示意圖 (Iqbal et al., 2023)

網格搜尋 (Grid search)

網格搜尋 (Grid search) 為最常見的超參數搜尋方式之一，透過窮舉法將所有由使用者定義的參數組合加以測試，如圖 3.17，進而尋找最佳超參數設定。實作簡單且值觀，通常可作為初期模型調參的基礎工具，尤其在超參數維度不高時，能有效找出最佳參數組合。然而，Grid search 難以處理高維度的參數空間，因所需計算量會隨參數數量呈指數成長，造成維度災難 (Belete & D H, 2021)。

本研究搭配 Scikit-learn 中之 `GridSearchCV` 函式進行網格搜尋與交叉驗證，提升調參效率與結果可信度。

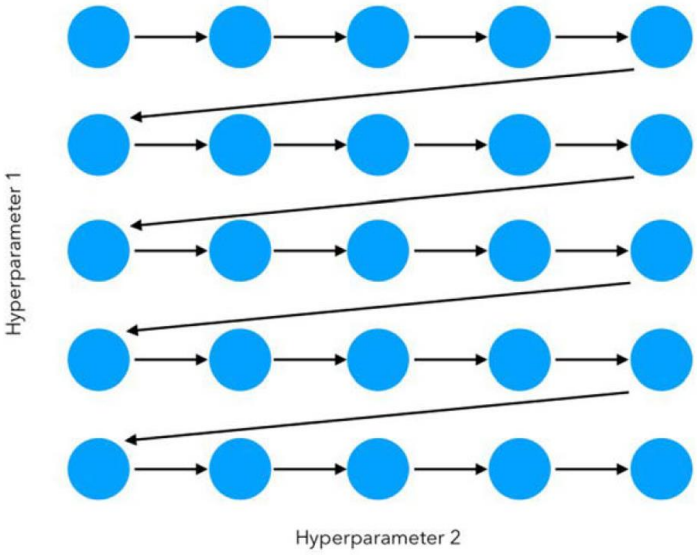


圖 3.17 GridSearch 歷遍所有超參數組合示意圖 (Iqbal et al., 2023)

doi:10.6342/NTU202502436

K-Fold cross-validation K 折交叉驗證

K-Fold 折交叉驗證為評估機器學習模型效能與泛化能力的常用方法，特別適用於資料量有限的情境。其作法為將資料平均分為 K 個子集 (folds)，每次將其中一個子集作為驗證集，其餘 K−1 個子集作為訓練集，重複 K 次並取平均驗證結果作為模型整體表現指標，如圖 3.18 所示。此方法不僅能有效評估模型預測誤差，也可觀察模型預測結果的變異 (variance) 程度，進一步判斷是否有過擬合現象。

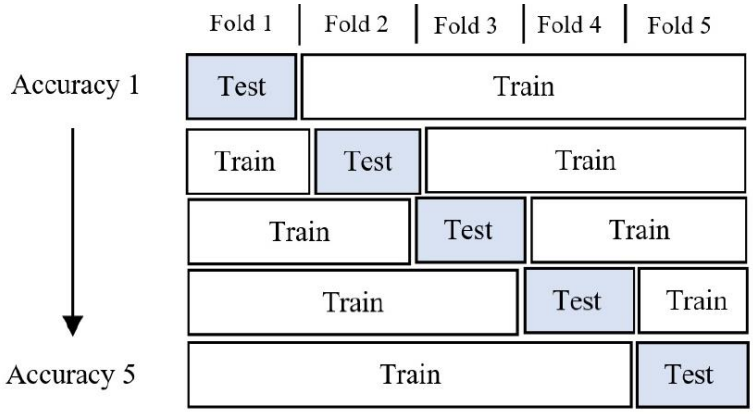


圖 3.18 K 折交叉驗證示意圖 (K = 5) (Ragb et al., 2021)

K 的選擇會影響模型評估結果，K 值較小時偏差 (bias) 可能較高，K 值較大則可能導致變異增加。K 折驗證亦常與超參數搜尋方法結合使用，協助選出同時具備低誤差與穩定性的最佳模型組合 (Ashfaque & Iqbal, 2019)。

本研究採用 K 折交叉驗證 (K = 5 或 10) 作為模型訓練與調參過程之標準驗證機制，並應用於各模型類型，以提升模型結果的穩定性與可信度。

doi:10.6342/NTU202502436

3.4.3 模型評估指標

為了能客觀衡量機器學習模型的表現，須採用適當的評估指標來進行分析結果之比較，而不同類型的分析問題對模型的評估標準亦有所差異。本節將根據迴歸與分類分析的不同需求，分別說明常用之評估指標之定義與意義。

迴歸分析 (Regression analysis)

在迴歸問題中，本研究採用四項常見指標來評估模型預測值與實際值之間的差異，即為模型中的損失函數，接下來將依序說明。令y_i為第 i 筆樣本的實際值，$\hat{y}_i$為模型預測值，$\bar{y}_i$為y_i的平均值，n 為樣本數。

1. 平均絕對誤差 (Mean Absolute Error, MAE)

用來衡量預測值與實際值之間的平均絕對差距，計算如式(3.7)，單位與原始資料一致，具良好解釋性，且相較於其他誤差指標，較不易受極端值影響。

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.7}$$

2. 均方誤差 (Mean Squared Error, MSE)

為預測誤差的平方平均，計算方式如式(3.8)，將會放大較大的誤差，對異常值較敏感，適用於強調大誤差懲罰的情境。

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.8}$$

3. 均方根誤差 (Root Mean Squared Error, RMSE)

為 MSE 的平方根，計算如式(3.9)，能保持單位與原始資料一致，同時兼具對大誤差的懲罰效果。

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(y_i - \hat{y}_i)^2} \tag{3.9}$$

doi:10.6342/NTU202502436

4. 決定係數 (Coefficient of Determination, R^2)

用來衡量模型預測值與實際值之間的擬合程度，計算如式(3.10)，數值介於 0 至 1 之間，越接近 1 代表模型的預測能力越好。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3.10}$$

本研究採用決定係數 R^2 作為迴歸模型之主要績效指標，其餘評估指標作為判斷模型好與壞的輔助，Chicco et al. (2021) 指出，相較於 MAE、MSE 即 RMSE 等常用指標，R^2 能更有效且誠實的反映模型在整體樣本的預測準確性，且能避免其他指標因值域不固定所導致的可解釋性問題。此外，R^2 對於錯誤模型會給予負分，具有良好的警示較果，故被建議做為評估迴歸模型之標準度量方式。

在實務應用與產業標準上，模型整體 R^2 分數介於 70%至 90%之間皆屬合理範圍，其中 70%通常被視為可接受的下限門檻，達此水準表模型輸出的預測結果具有實現可用性 (realistic usability)。

在模型評估上，本研究不僅觀察測試資料之 R^2 分數 (training score)，也同步列出訓練資料的預測表現 (testing score)。由於模型在訓練階段直接利用訓練資料做學習，因此訓練分數通常會明顯高於測試分數，為可預期之結果。

然而，若兩者差距過大，可能代表模型過擬合 (overfitting)，即模型過度學習訓練資料中的細節，無法有效應用於新資料上。反之若訓練與測試分數皆偏低，則可能出現欠擬合 (underfitting) 的情況，顯示模型學習能力不足，無法學習資料間的有效關聯。相對地，若訓練與測試分數皆偏高且相差不大，則可視為模型學習效果良好，且具有高度穩定性與應用潛力。透過觀察兩者的表現差距，可輔助判斷模型是否存在過度複雜或學習偏差等問題。

此外，本研究資料總筆數約為 1000 筆，屬中等規模，為保留足夠樣本進行預測分析，僅將資料分為訓練與測試集兩組，若再劃出第三組驗證集 (validation set)，將壓縮訓練與測試資料的規模，影響模型穩定性與結果之代表性。

doi:10.6342/NTU202502436

分類分析 (Classification analysis)

在分類問題中，模型的預測結果為離散類別，評估重點不僅在於整體準確性，也包括對不同類別的辨識能力。為全面了解分類模型的效能，需透過混淆矩陣 (Confusion matrix) 進行預測結果的分類分析，並以此為基礎計算多項分類指標。

1. 混淆矩陣 (Confusion matrix)

混淆矩陣是一種將實際標籤與模型預測結果進行交叉比對的表格，能同時顯示正確與錯誤的分類情形。以一簡易二元分類為例，假設共 100 個樣本需分類為 True 或 False，其預測結果可整理為一 2×2 的混淆矩陣，如圖 3.19 所示，包含下列四種分類結果：

- True Positive (TP)：實際為正樣本，預測也為正樣本。於此範例中其值為 35。
- True Negative (TN)：實際為負樣本，預測也為負樣本。於此範例中其值為 38。
- False Positive (FP)：實際為負樣本，預測錯誤為正樣本。於此範例中其值為 11。
- False Negative (FN)：實際為正樣本，預測錯誤為負樣本。於此範例中其值為 16。

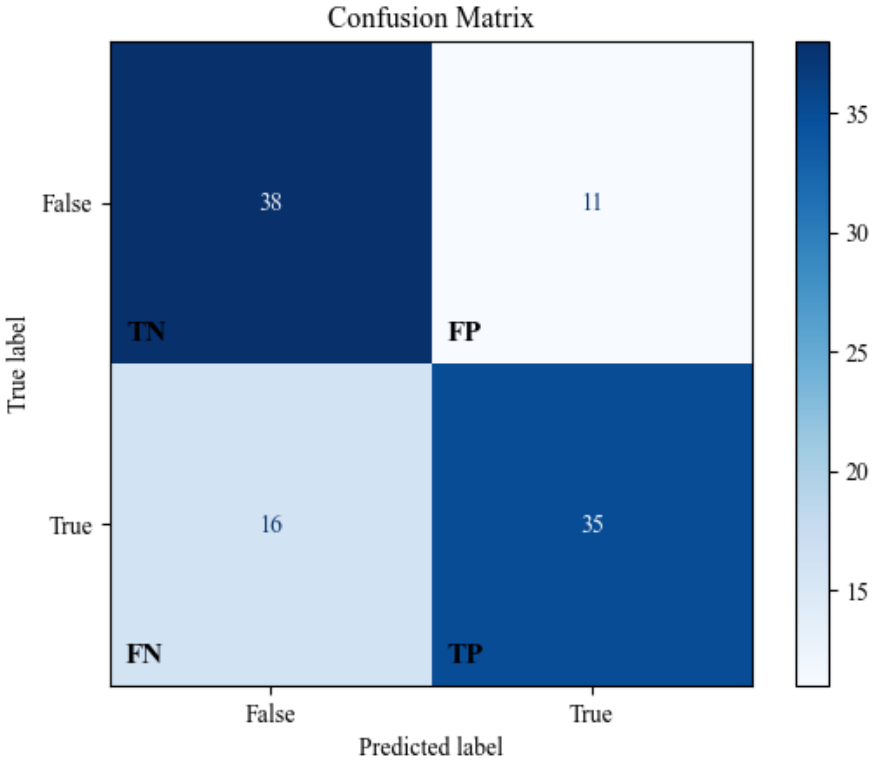


圖 3.19 範例之混淆矩陣

doi:10.6342/NTU202502436

2. 正確率 (Accuracy)

正確率為最值觀且常見的分類評估指標，用以衡量模型預測正確的整體比例。其定義為模型正確預測的樣本數 (即 TP 與 TN 之和) 佔全部樣本的比例，計算方式如式(3.11)，於前述範例中其值為 0.730。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.11}$$

此評估標準適用於正負樣本數量相對平衡的分類問題，反之，若在樣本不均的情況下，正確率可能因偏重多數類別而產生誤導，需搭配其他指標進行綜合評斷。

3. 精準率 (Precision)

精準率為衡量模型在所有預測為正樣本中，有多少比例實際為正樣本的指標，計算方式如(3.12)，於前述範例中其值為 0.761。

$$Presicion = \frac{TP}{TP + FP} \tag{3.12}$$

特別適用於需降低誤報率 (False Positive) 的情境 (如醫學檢測中) ，缺點是無法反映模型對正樣本的辨識能力，仍需搭配其他指標進行綜合評斷。

4. 召回率 (Recall)

召回率為衡量模型在所有實際為正樣本中，有多少比例成功被預測為正樣本的指標，計算方式如式(3.13)，於前述範例中其值為 0.686。

$$Recall = \frac{TP}{TP + FN} \tag{3.13}$$

適用於重視漏判 (False Negative) 風險的情境，特別在錯過正樣本可能造成嚴重後果的應用領域 (如疾病篩檢或災害預警系統中)，惟其無法單獨反映預測為正樣本的正確性，通常需與精準率共同參考。

doi:10.6342/NTU202502436

5. F1 分數 (F1-score)

F1 分數為精準率與召回率的調和平均，綜合反映模型在正樣本預測上的正確性與覆蓋率，計算方式如式(3.14)，於前述範例中其值為 0.722。

$$F1-score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \tag{3.14}$$

特別適用於正負樣本不平均的情境，當需在誤報與漏報之間取得平衡時，F1 分數可作為一較全面的指標，協助評估模型對正樣本的綜合辨識能力。

doi:10.6342/NTU202502436

第四章 模型表現與預測結果比較

為驗證本研究所建立之機器學習模型在預測液化試驗資料中的效能，本章將針對模型的整體表現進行分析，並比較不同條件下模型預測結果之差異。透過量化指標與視覺化圖表，評估各模型於不同特徵組合與資料填補方法下的預測能力與穩定性，並進一步探討資料特徵的重要性。

4.1 模型表現分析

本節旨在分析本研究所使用之機器學習模型的預測表現，並探討其在不同超參數設定下的結果。由於模型表現可能受到多項因素影響，包含資料前處理方式、特徵選擇策略、以及模型本身結構與參數設定等，因此在進行模型比較前，先就模型的超參數對預測結果的影響進行初步探討。

4.1.1 超參數設定與其影響

本研究在此步驟中進行多組超參數之手動調整，以了解不同設定對模型表現之影響。採用固定其他參數不變的條件下，針對單一超參數進行調整，觀察其對預測結果的變化趨勢。此步驟分析雖不具系統性搜尋的完整性，但對於辨識影響模型性能的關鍵超參數仍具有參考價值。

以一組 EBM 模型為範例，在資料僅包含主要特徵的情況下，選定「每個分箱最少需要的樣本數 (min_samples_leaf)」做為變因，觀察其變化對模型 R^2 的影響。其餘固定參數之設定如表 4.1 所示:

doi:10.6342/NTU202502436

表 4.1 EBM 調參範例中各超參數設定

代碼名稱	值
max_bins	256
max_interaction_bins	32
interactions	10
learning_rate	0.01
max_rounds	5000
early_stopping_rounds	50
max_leaves	3

分析結果如圖 4.1，降低 min_samples_leaf，訓練與預測之 R^2 分數有上升的趨勢，顯示當每個分箱所需的最小樣本數越少時，模型能更細緻的擬合資料特性，提升預測準確度。

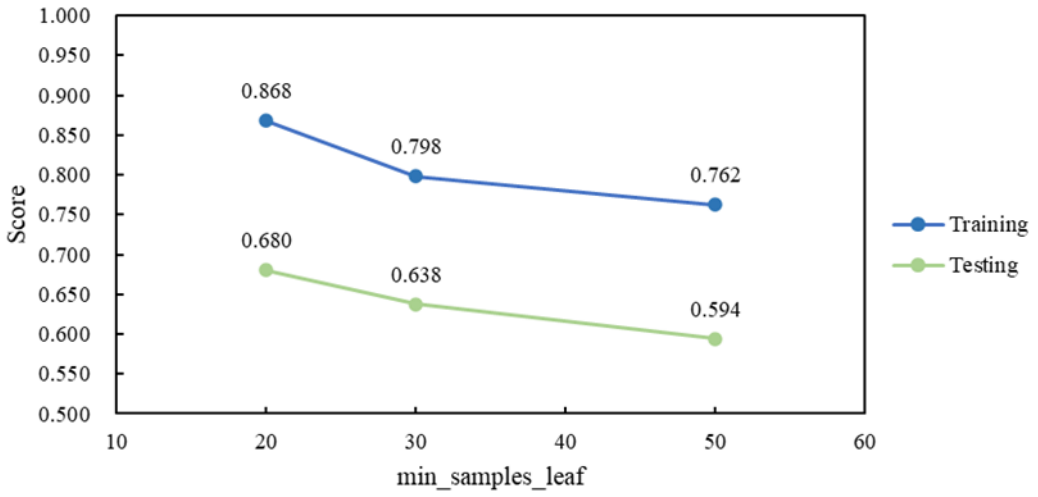
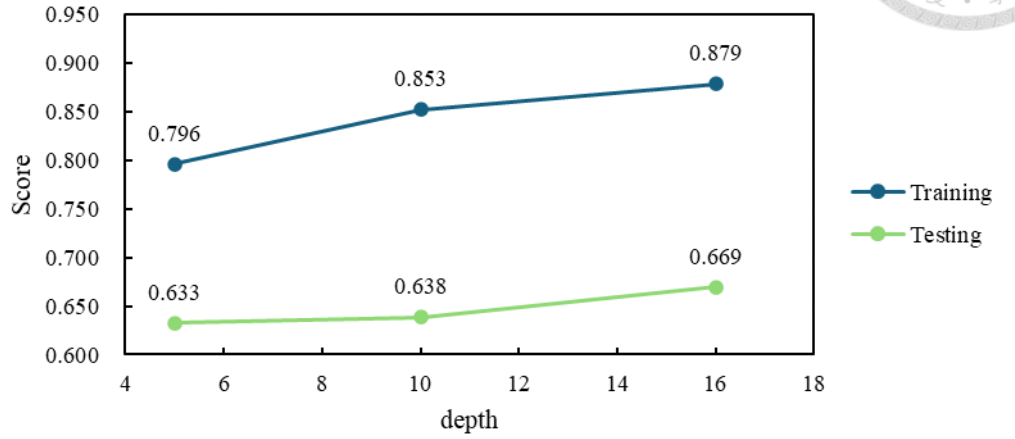


圖 4.1 EBM 範例之 R^2 分數比較圖

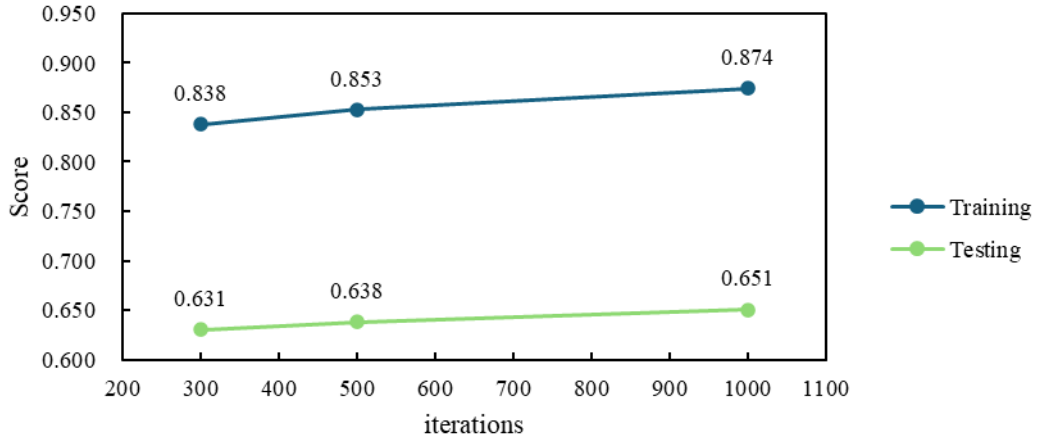
除上述範例外，本研究針對多組模型進行相似分析。於折線圖中，各組模型以變因為橫軸，訓練與測試之 R^2 分數為縱軸進行繪圖；而在橫向長條圖中，變因則置於縱軸，R^2 分數置於橫軸，以呈現模型表現之變化趨勢。由於各模型之固定參數組合眾多，本節僅彙整圖表結果以供說明，詳細設定請參見附錄 A。

doi:10.6342/NTU202502436

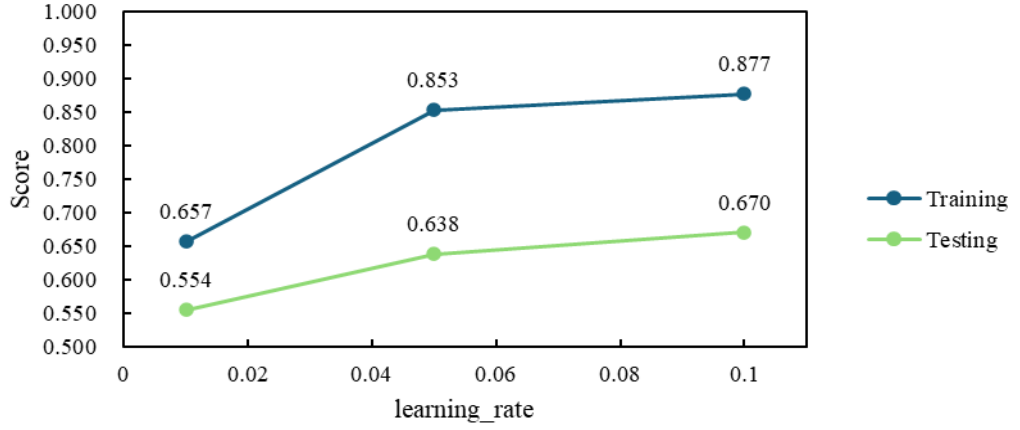
圖 4.2 至圖 4.7 為各機器學習模型之手動調參分析圖，本階段所得之最適超參數組合，將於後續章節中加以應用與比較。



(a)

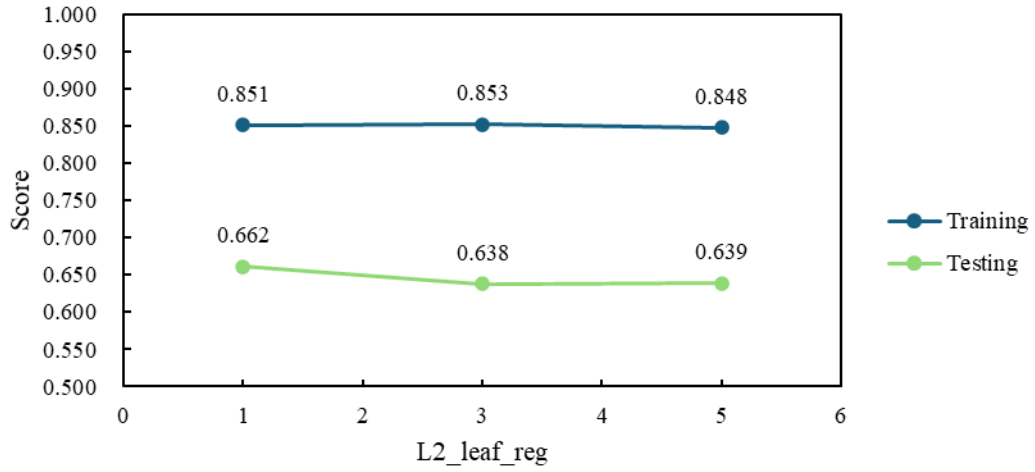


(b)

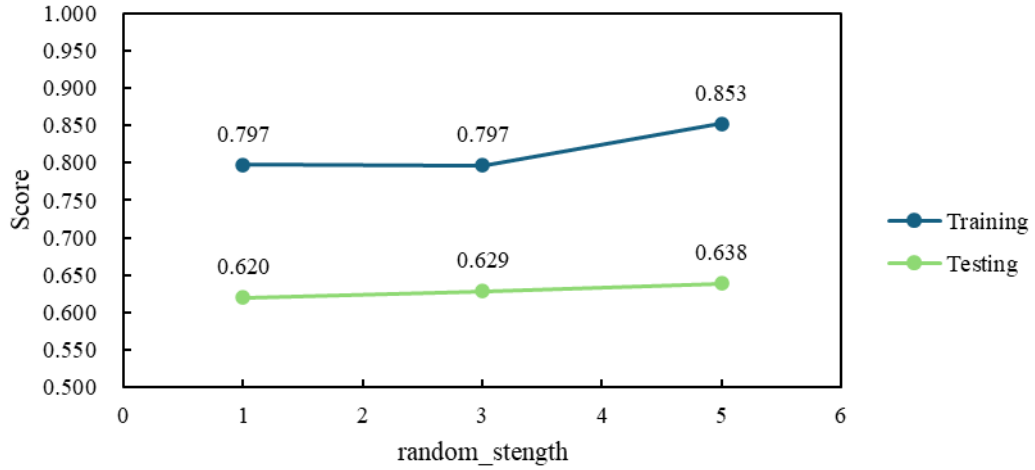


(c)

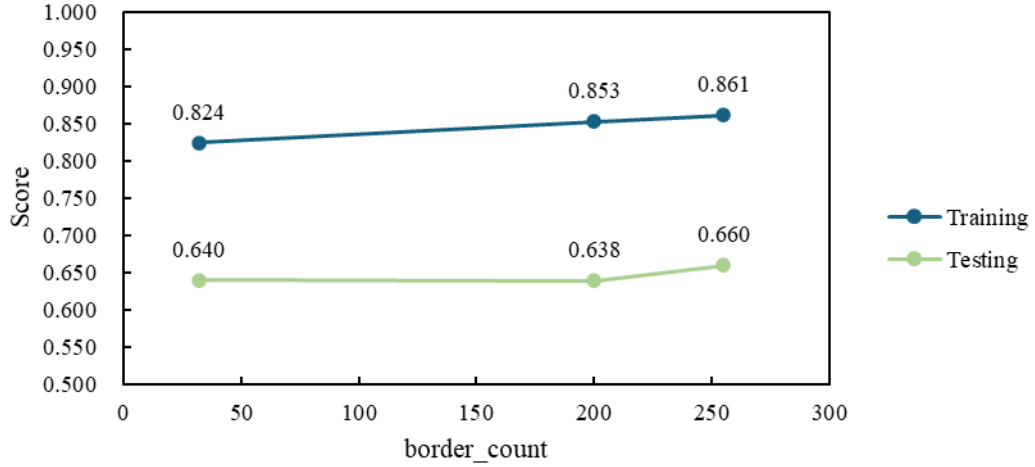
doi:10.6342/NTU202502436



(d)



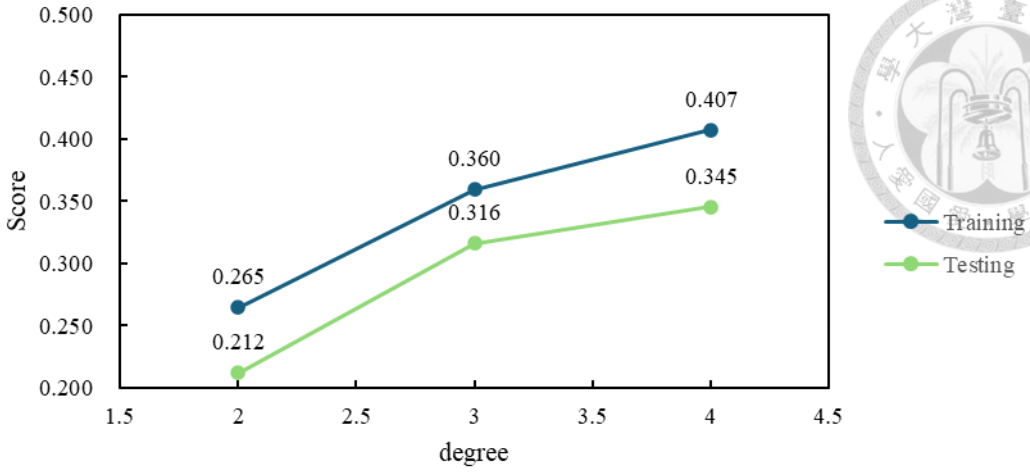
(e)



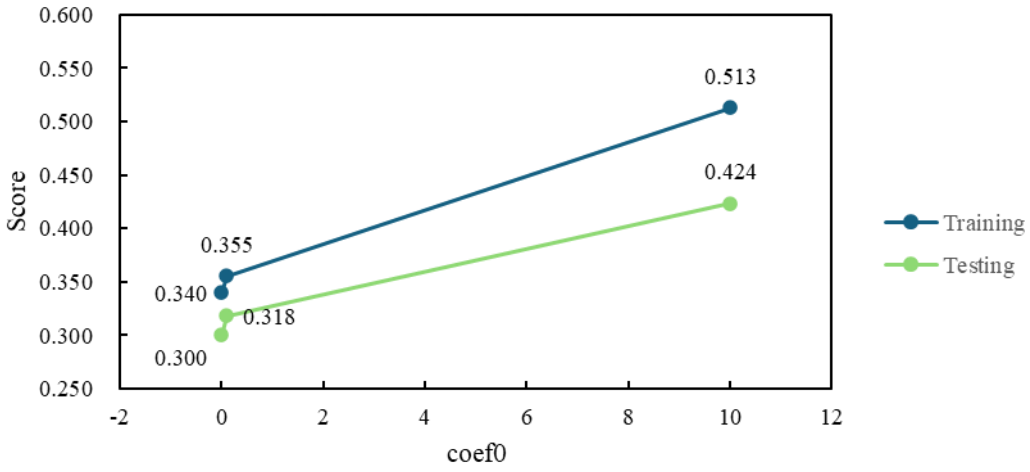
(f)

圖 4.2 CatBoost 不同超參數之 R^2 分數比較圖 (a) depth；(b) iterations；(c) learning_rate；(d) L2_leaf_reg；(e) random_stength；(f) border_count

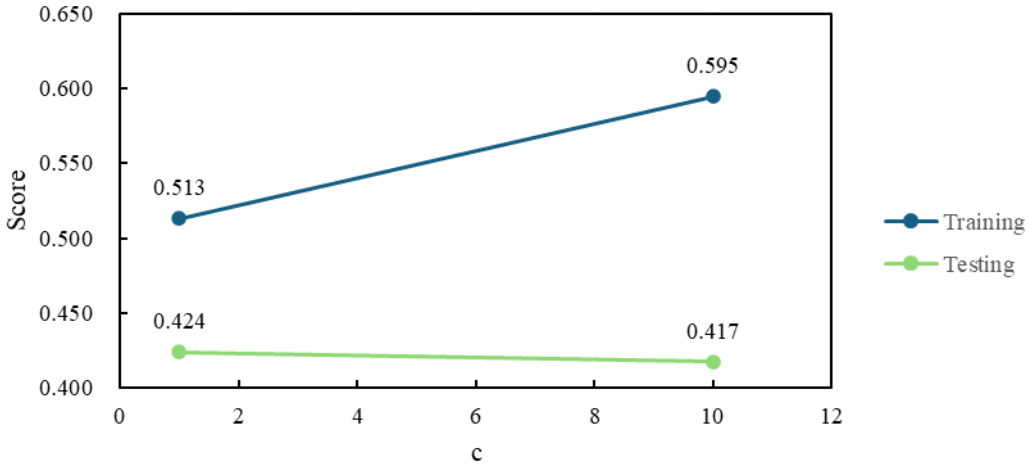
doi:10.6342/NTU202502436



(a)

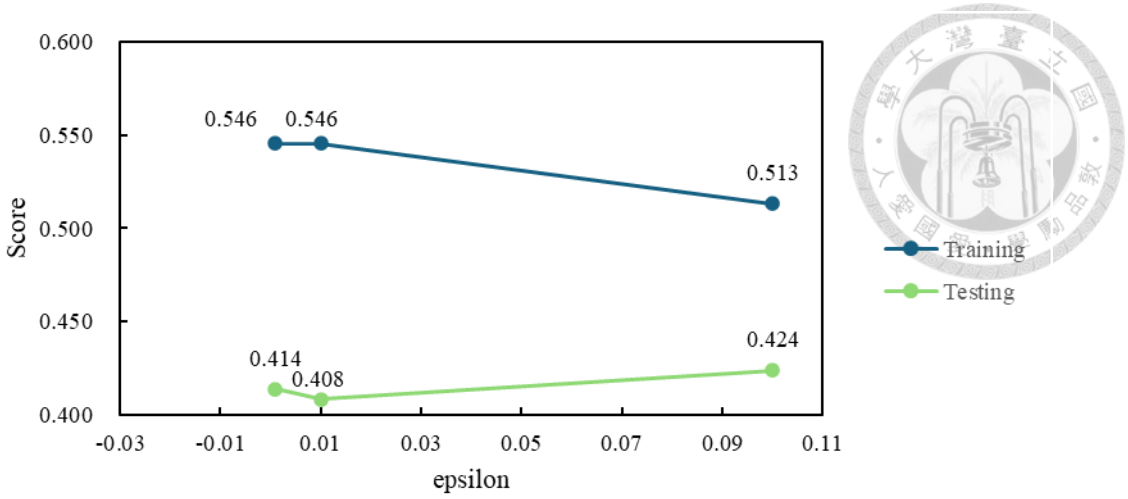


(b)

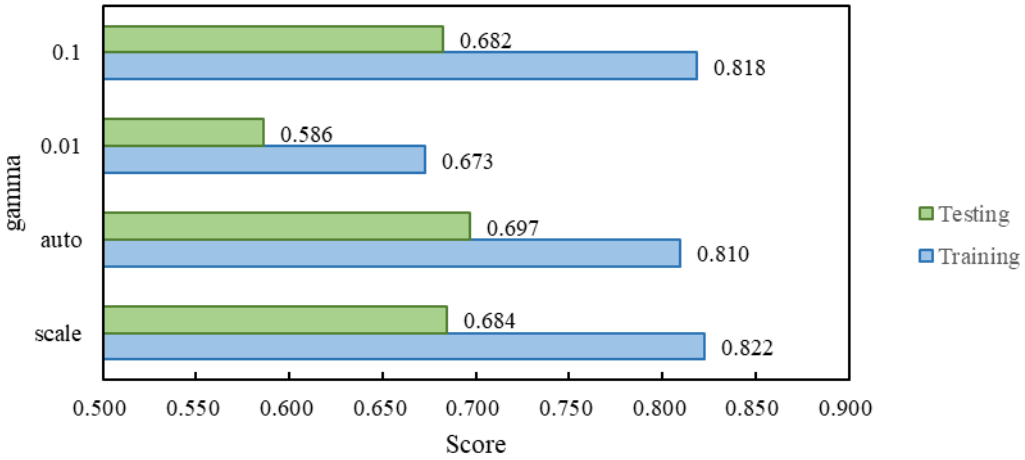


(c)

doi:10.6342/NTU202502436



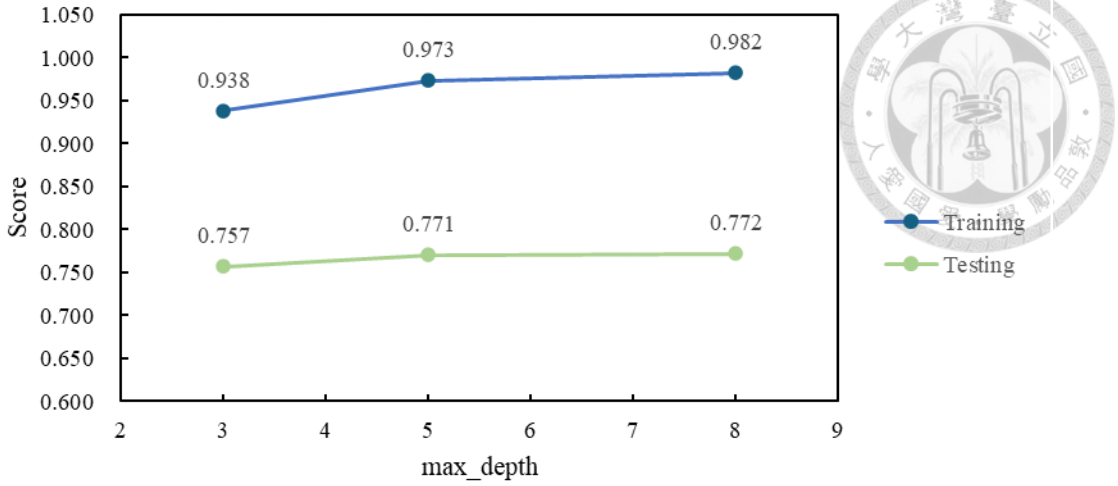
(d)



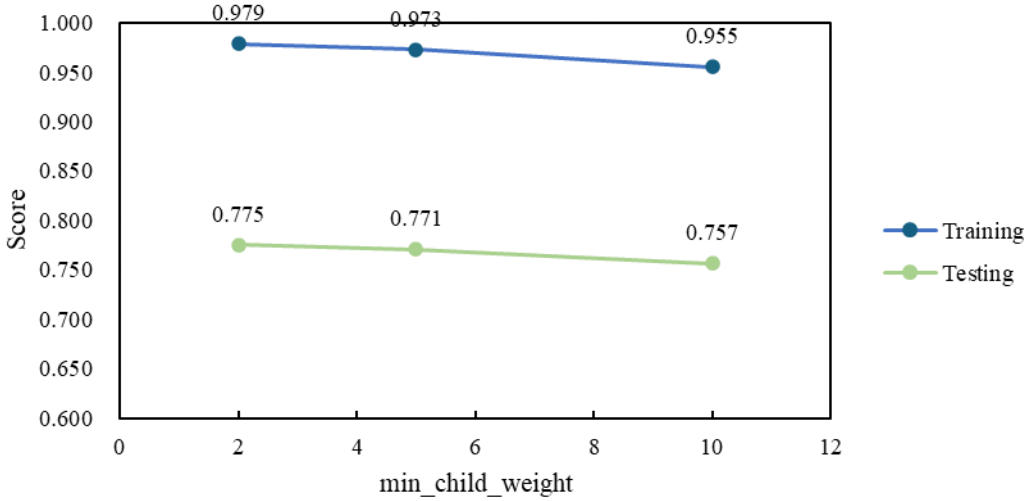
(e)

圖 4.3 SVM 不同超參數之 R^2 分數比較圖 (a) degree；(b) coef0；(c) c；(d) epsilon；(e) gamma

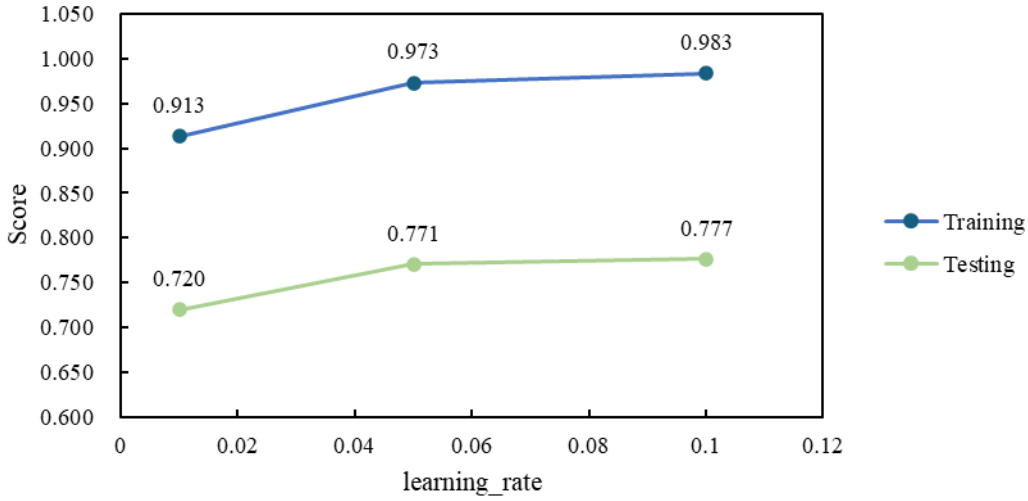
doi:10.6342/NTU202502436



(a)

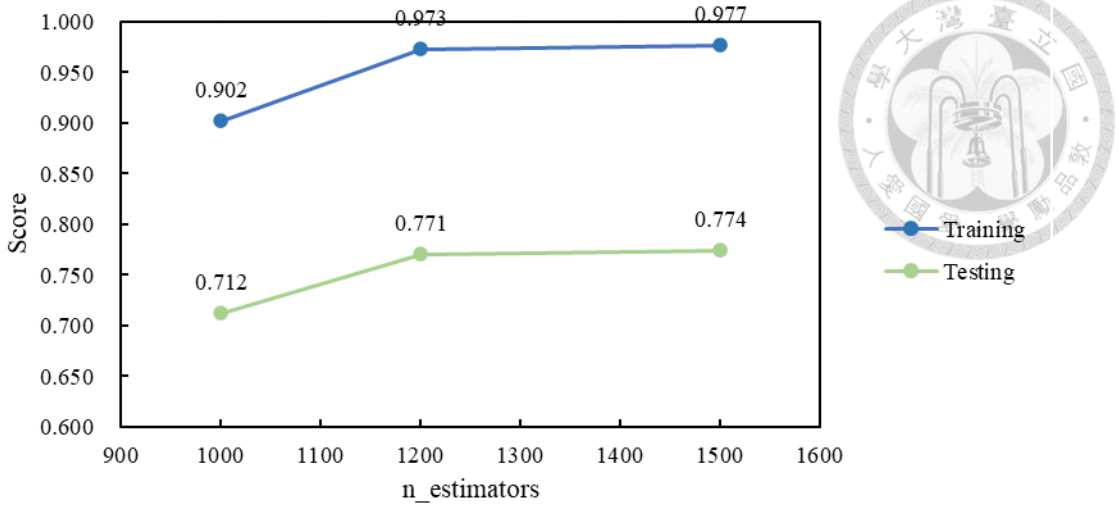


(b)

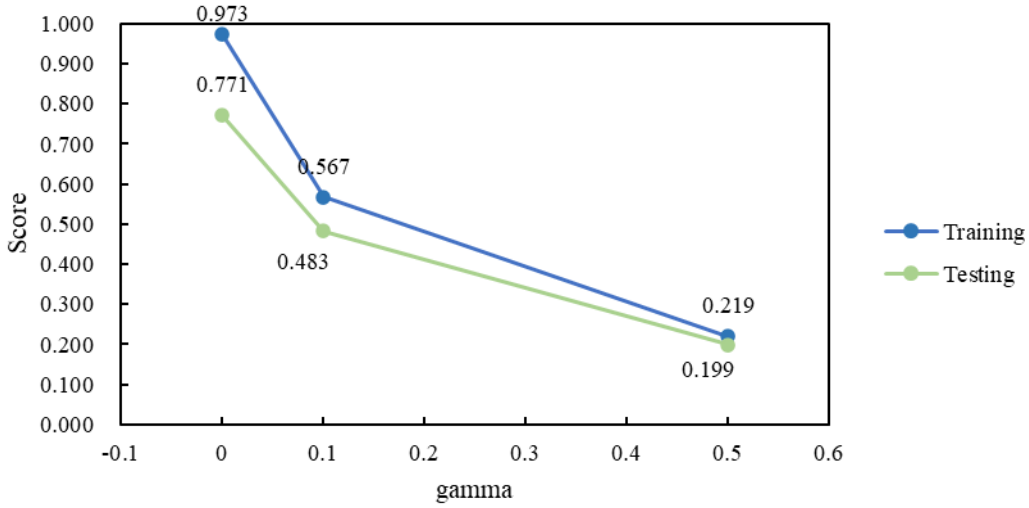


(c)

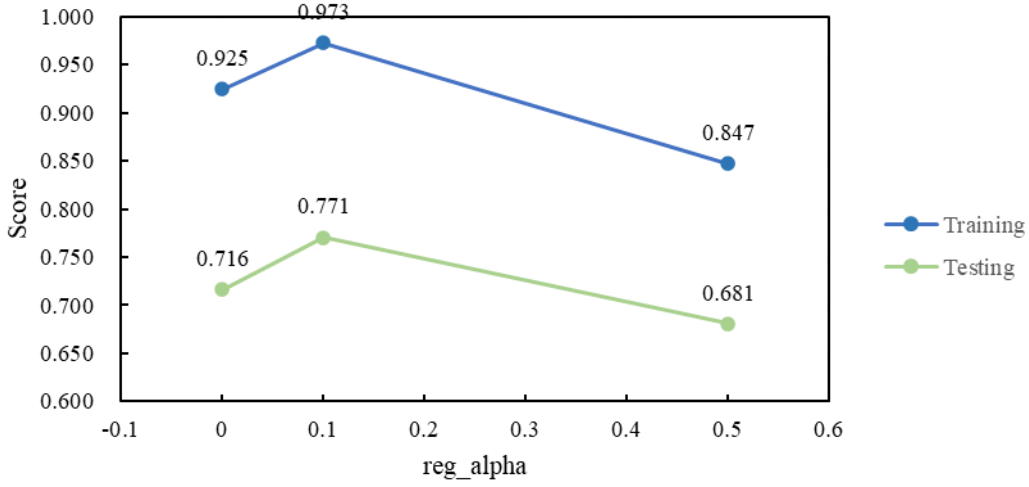
doi:10.6342/NTU202502436



(d)

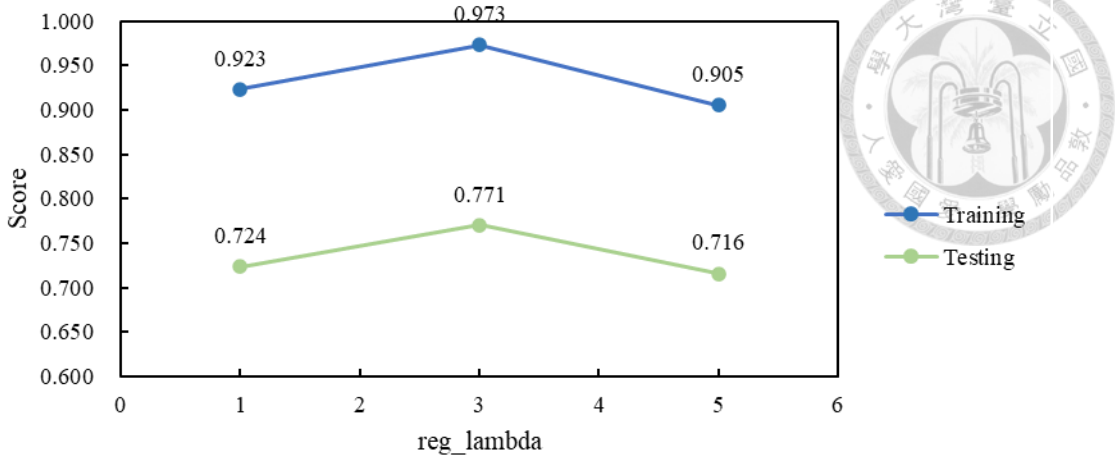


(e)

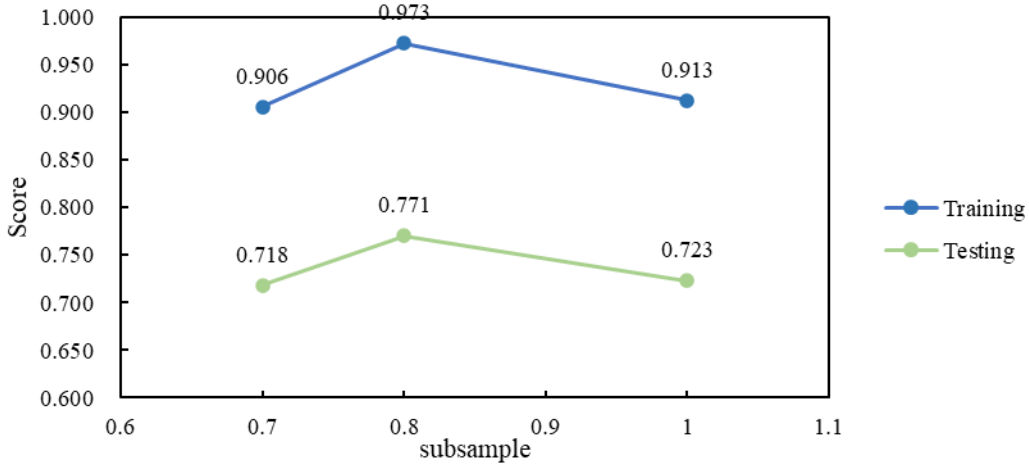


(f)

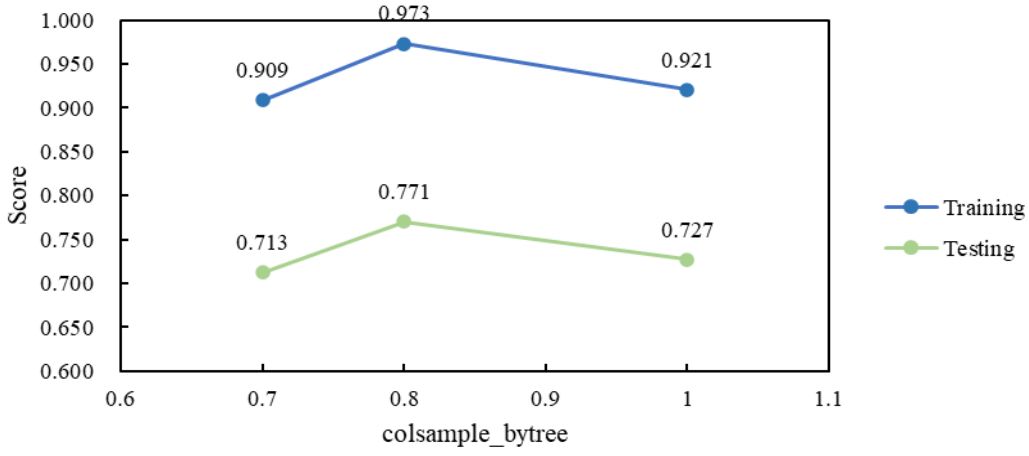
doi:10.6342/NTU202502436



(g)



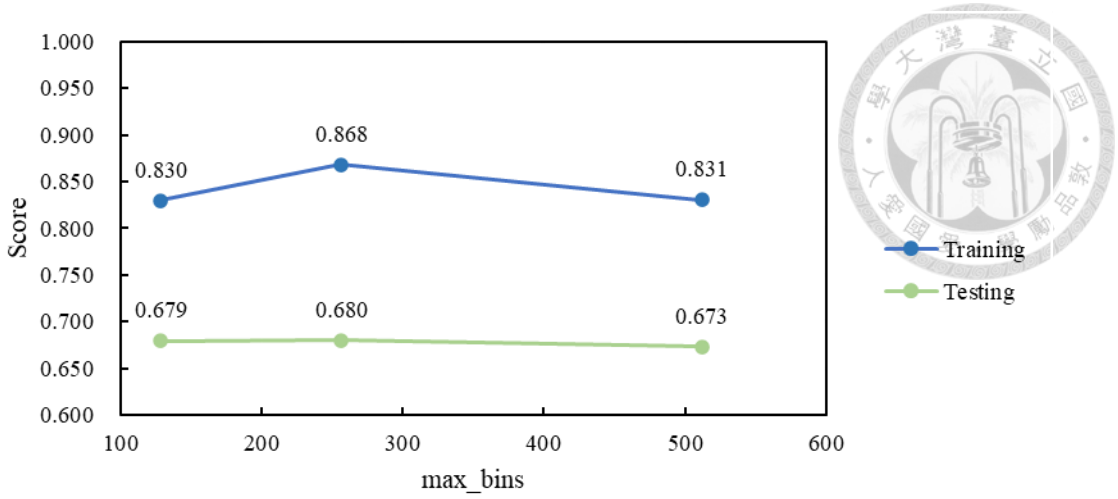
(h)



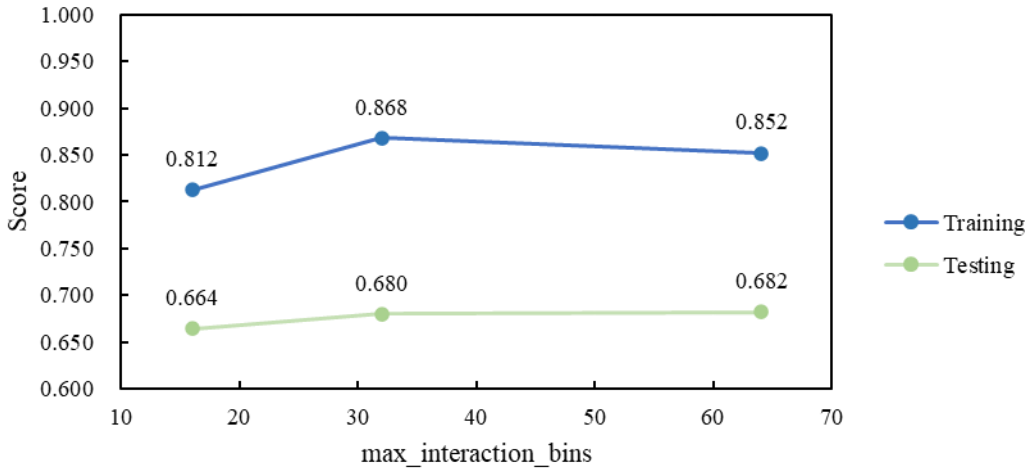
(i)

圖 4.4 XGBoost 不同超參數之 R^2 分數比較圖 (a) max_depth；(b) min_child_weight；(c) learning_rate；(d) n_estimators；(e) gamma；(f) reg_alpha；(g) reg_lambda；(h) subsample；(i) colsample_bytree

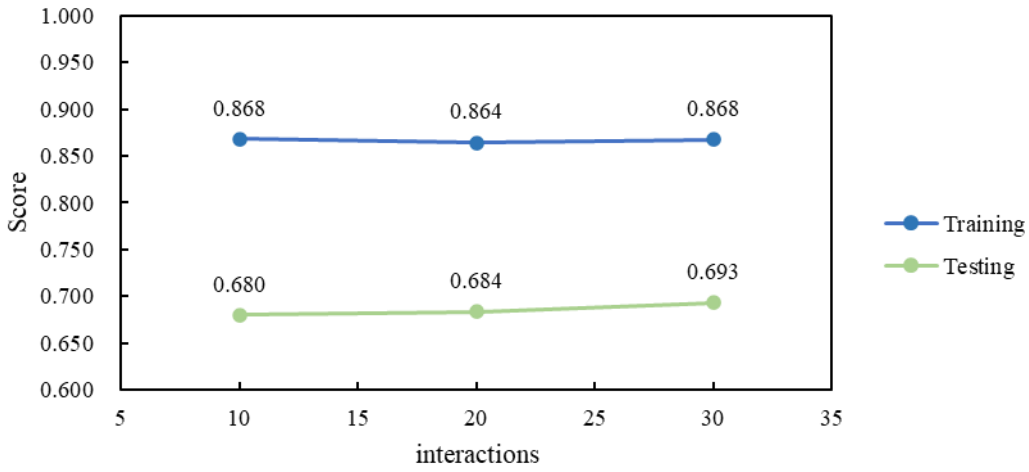
doi:10.6342/NTU202502436



(a)

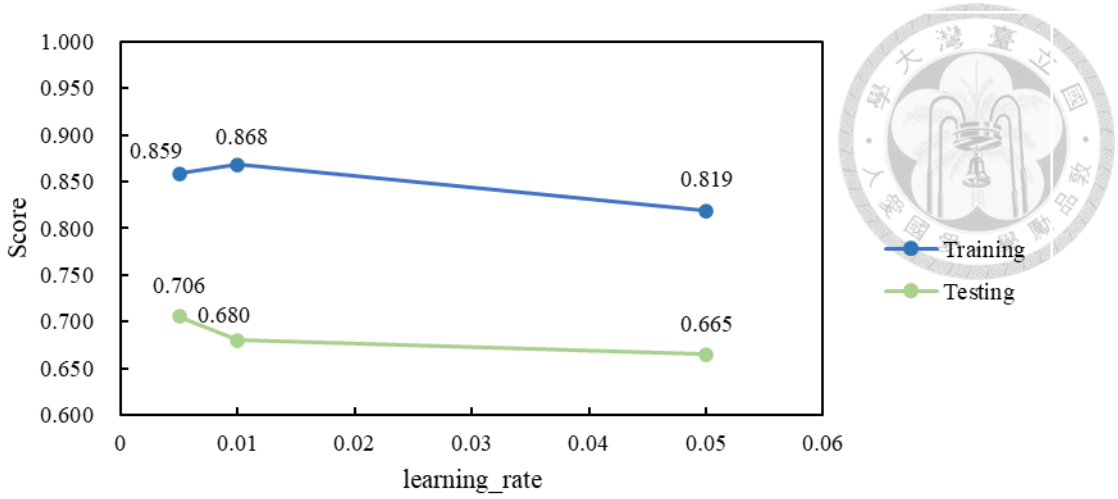


(b)

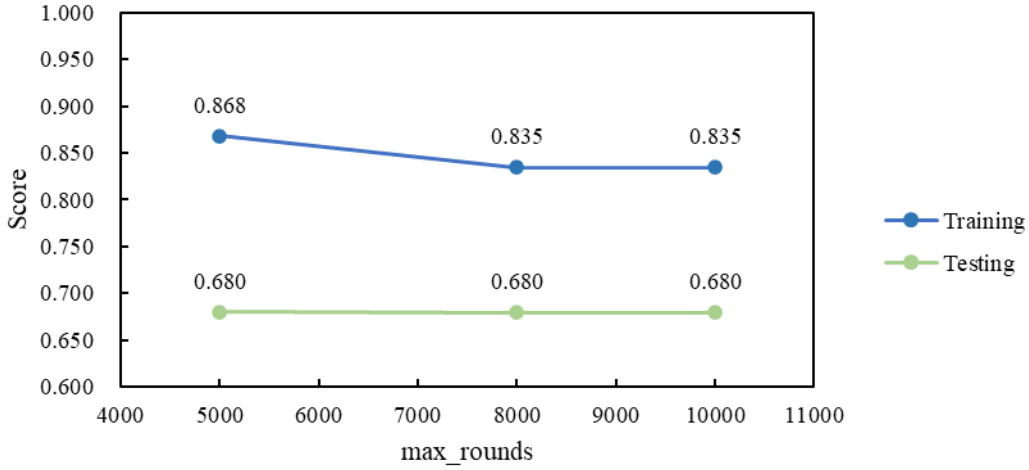


(c)

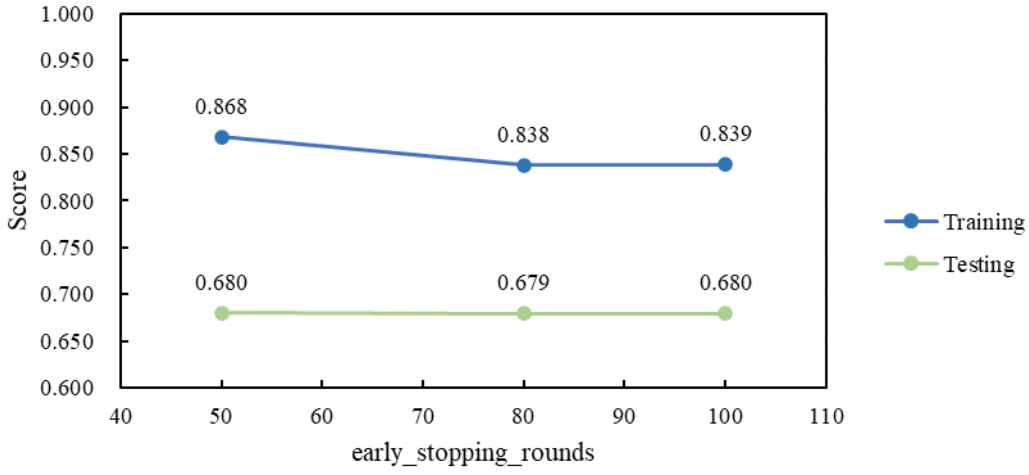
doi:10.6342/NTU202502436



(d)

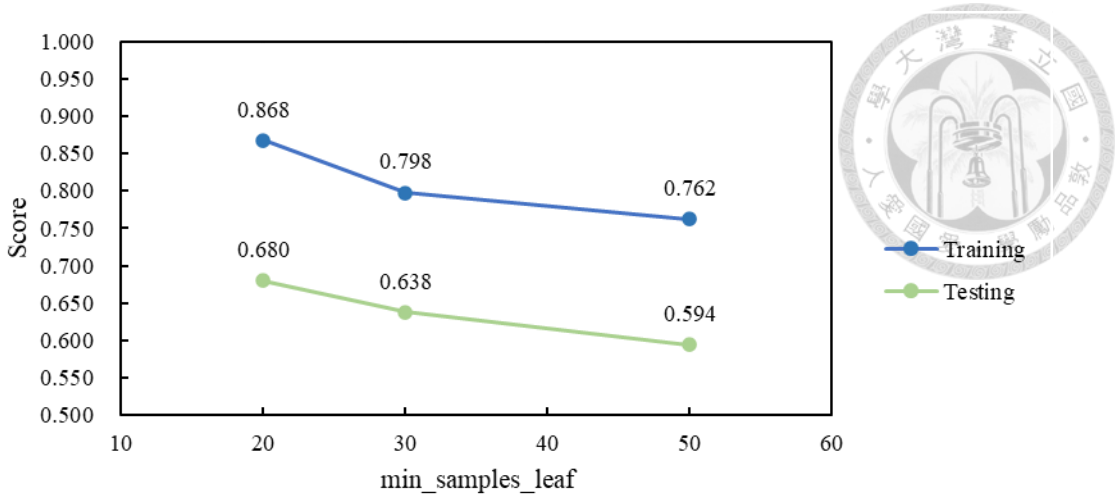


(e)

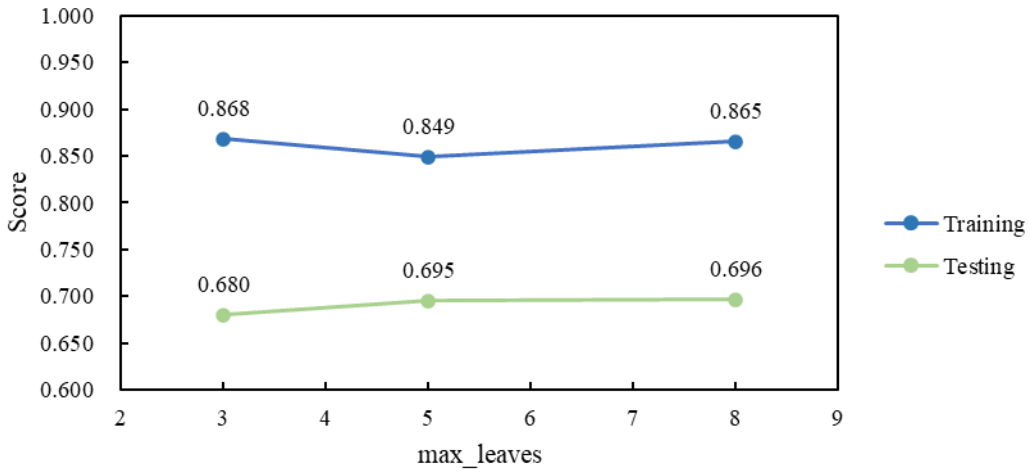


(f)

doi:10.6342/NTU202502436



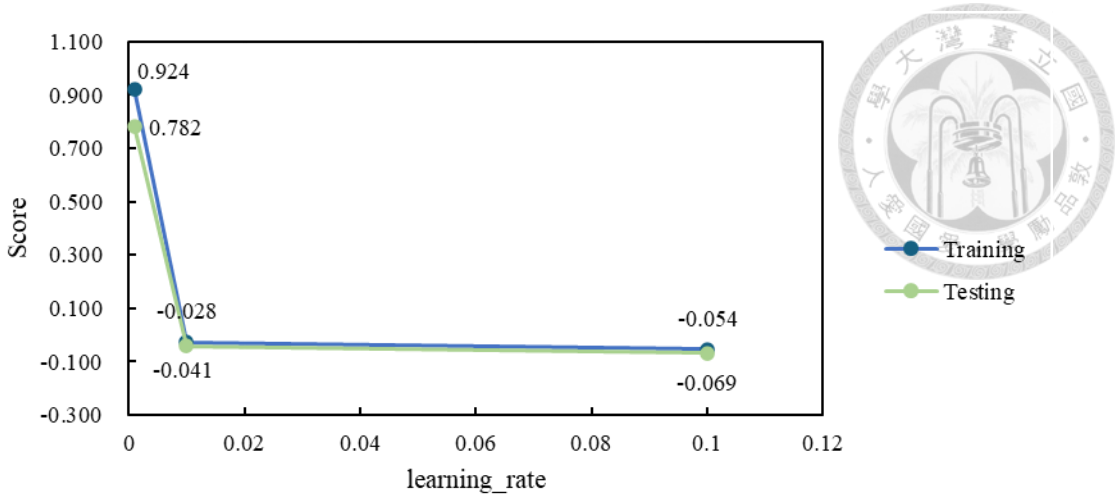
(g)



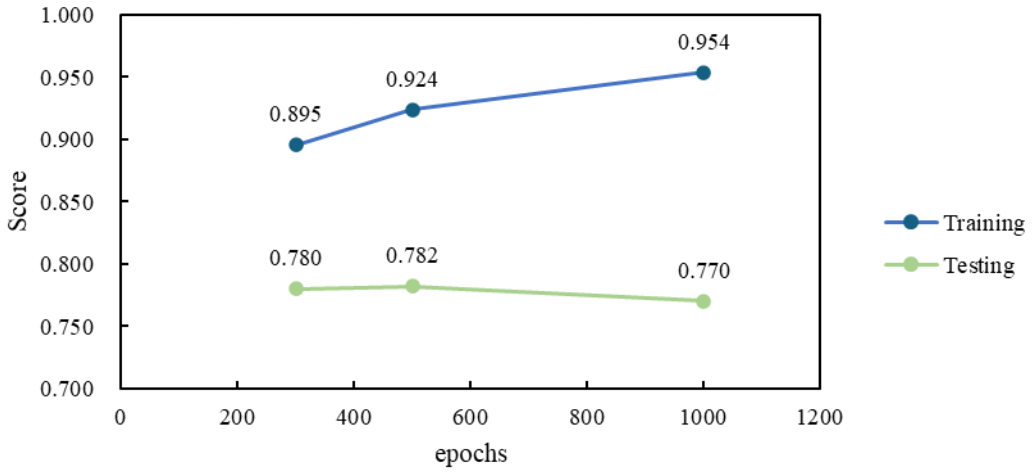
(h)

圖 4.5 EBM 不同超參數之 R^2 分數比較圖 (a) max_bins；(b) max_interactiion_bins；(c) interactions；(d) learning_rate；(e) max_rounds；(f) early_stoppping_rounds；(g) min_samples_leaf；(h) max_leaves

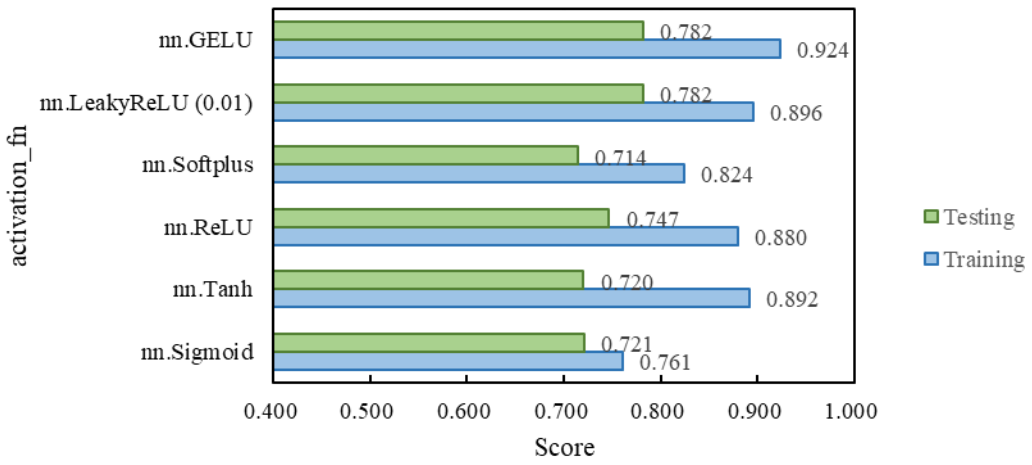
doi:10.6342/NTU202502436



(a)

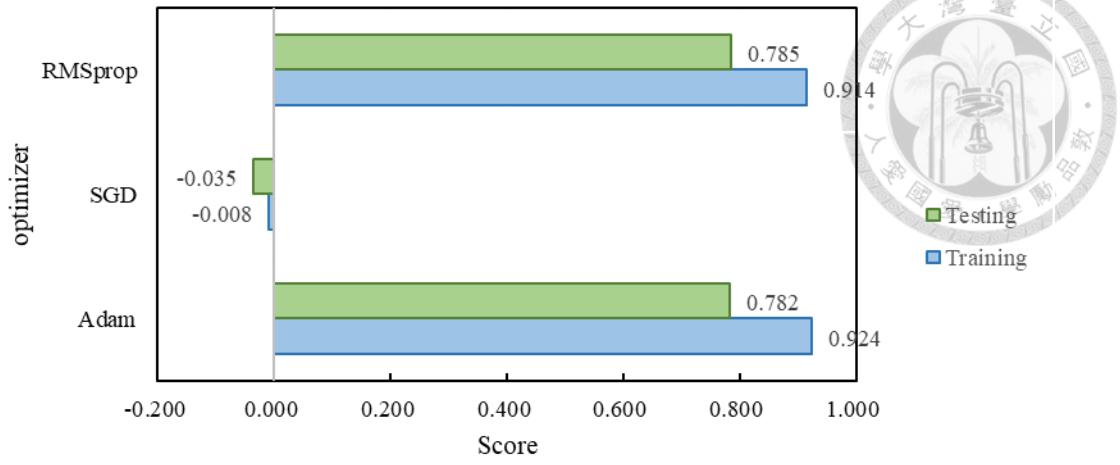


(b)



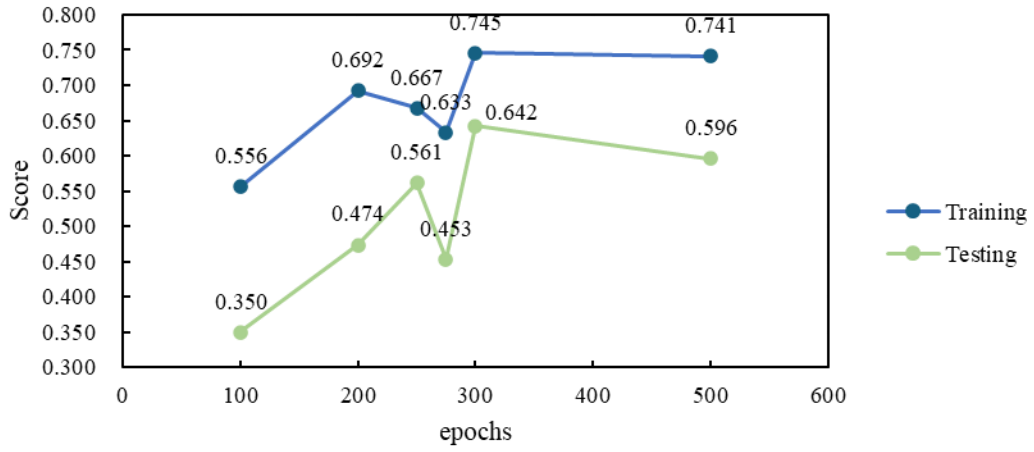
(c)

doi:10.6342/NTU202502436

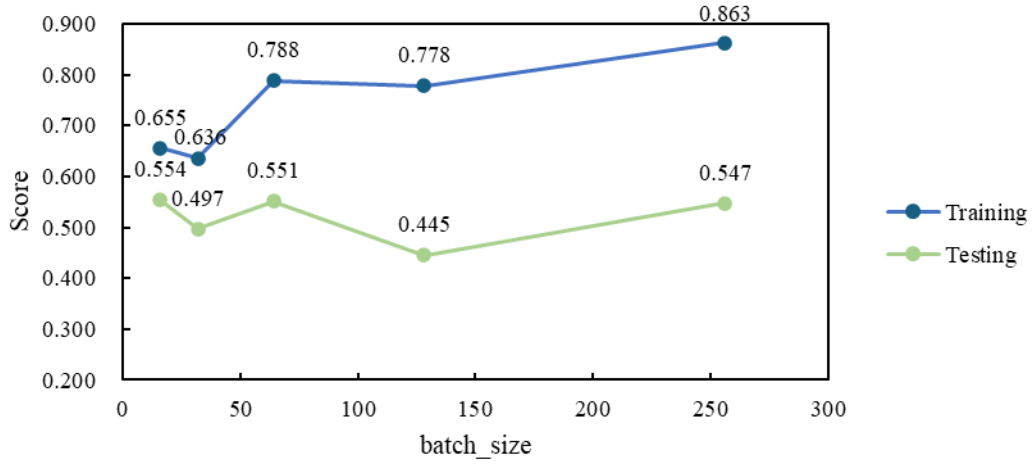


(d)

圖 4.6 BNN 不同超參數之 R^2 分數比較圖 (a) learning_rate；(b) epoch；(c) activaiton_fn；(d) optimizer

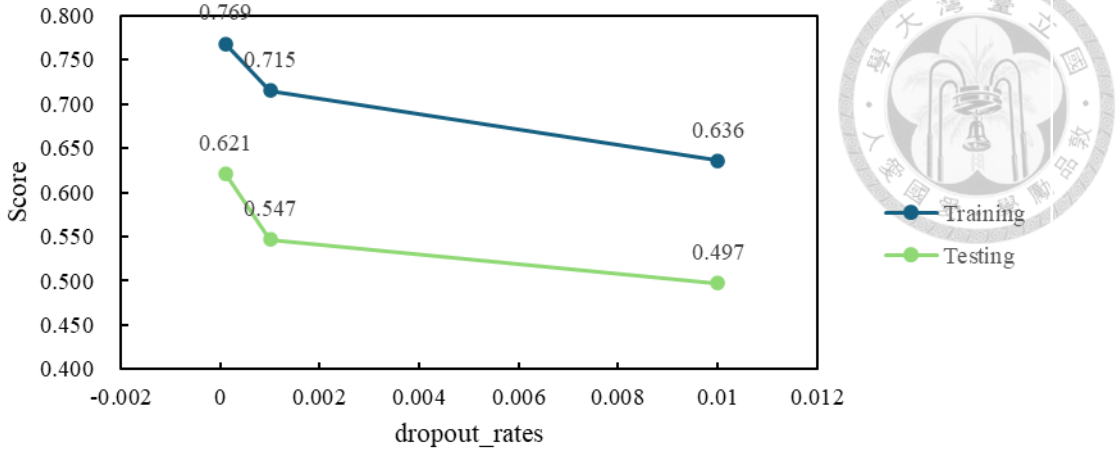


(a)

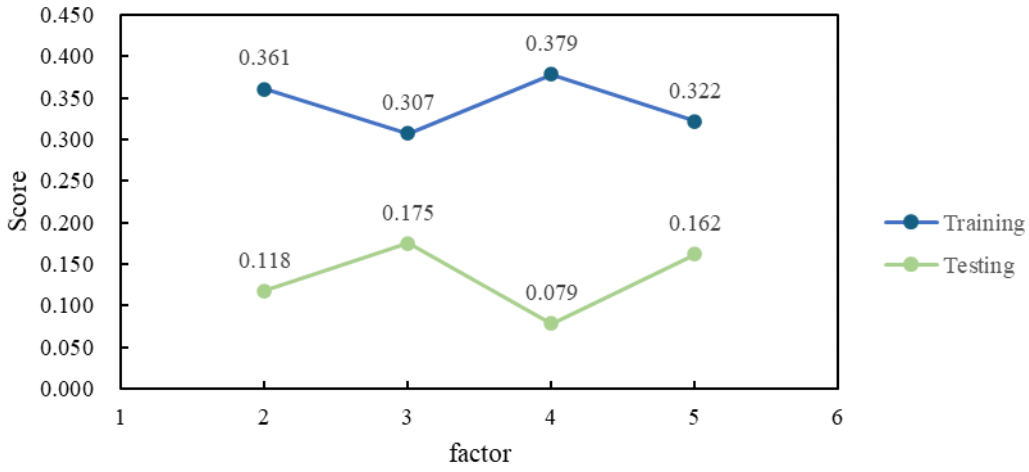


(b)

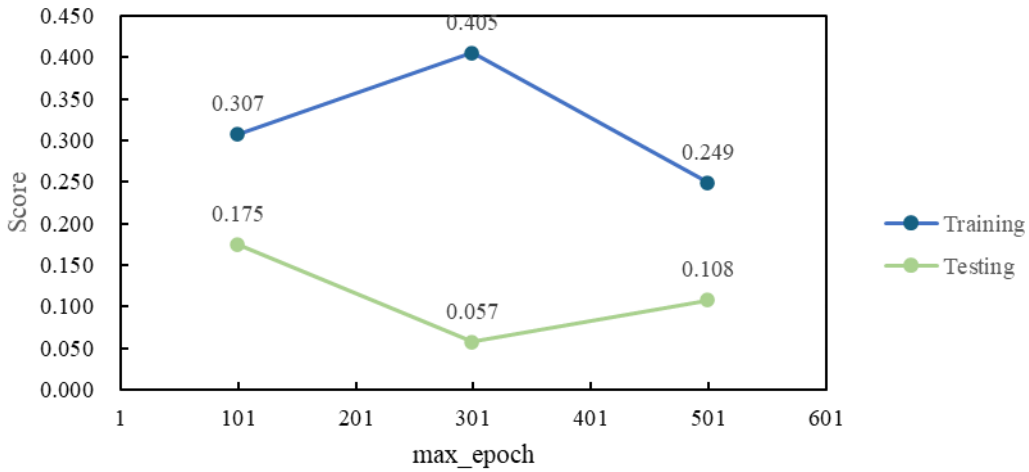
doi:10.6342/NTU202502436



(c)

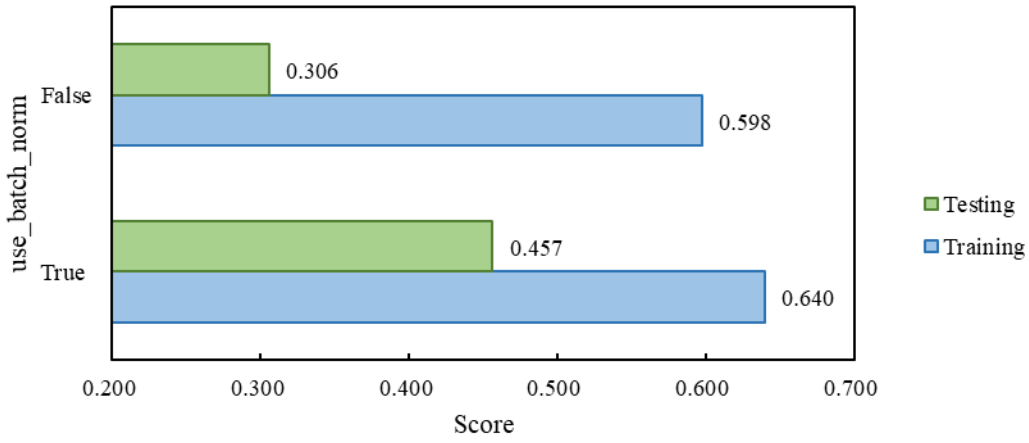


(d)

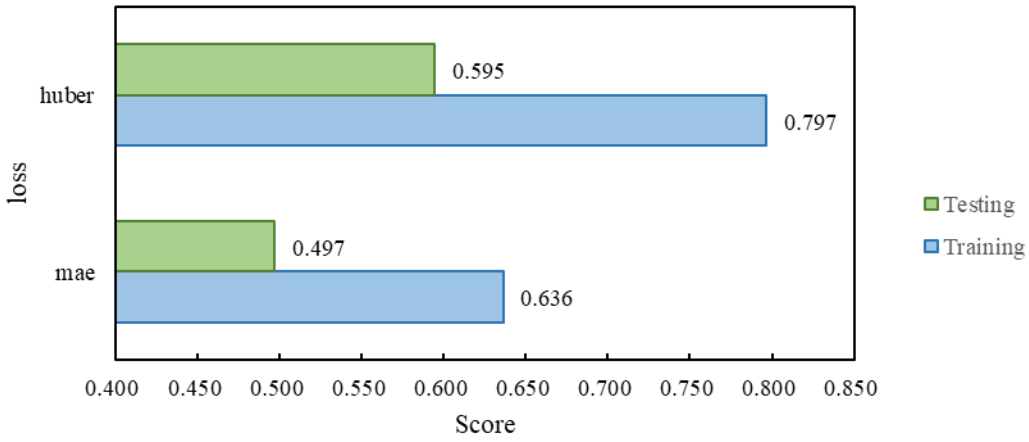


(e)

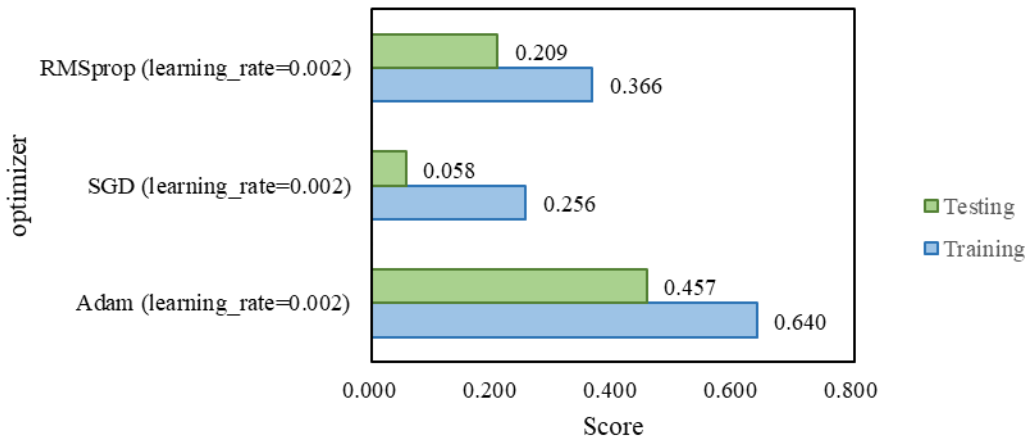
doi:10.6342/NTU202502436



(f)

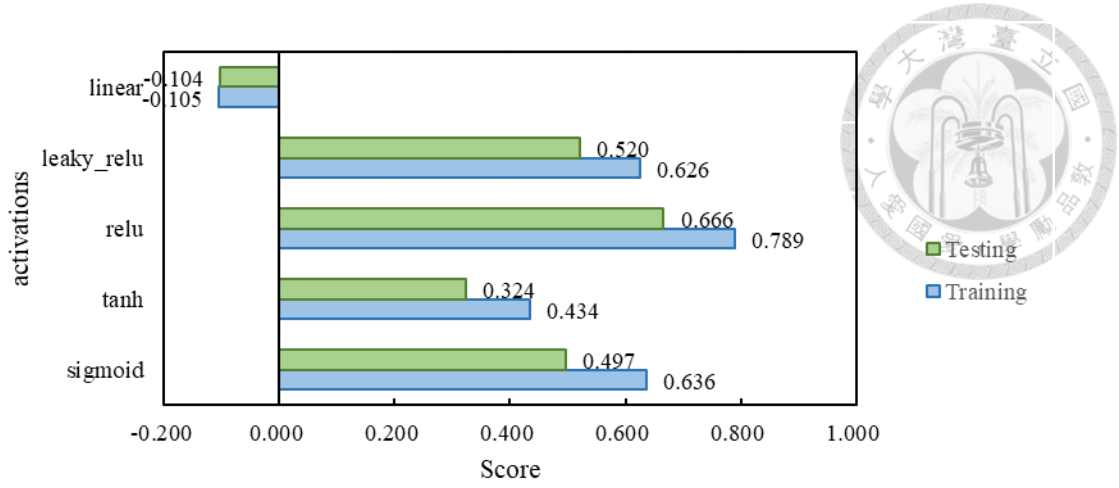


(g)



(h)

doi:10.6342/NTU202502436



(i)

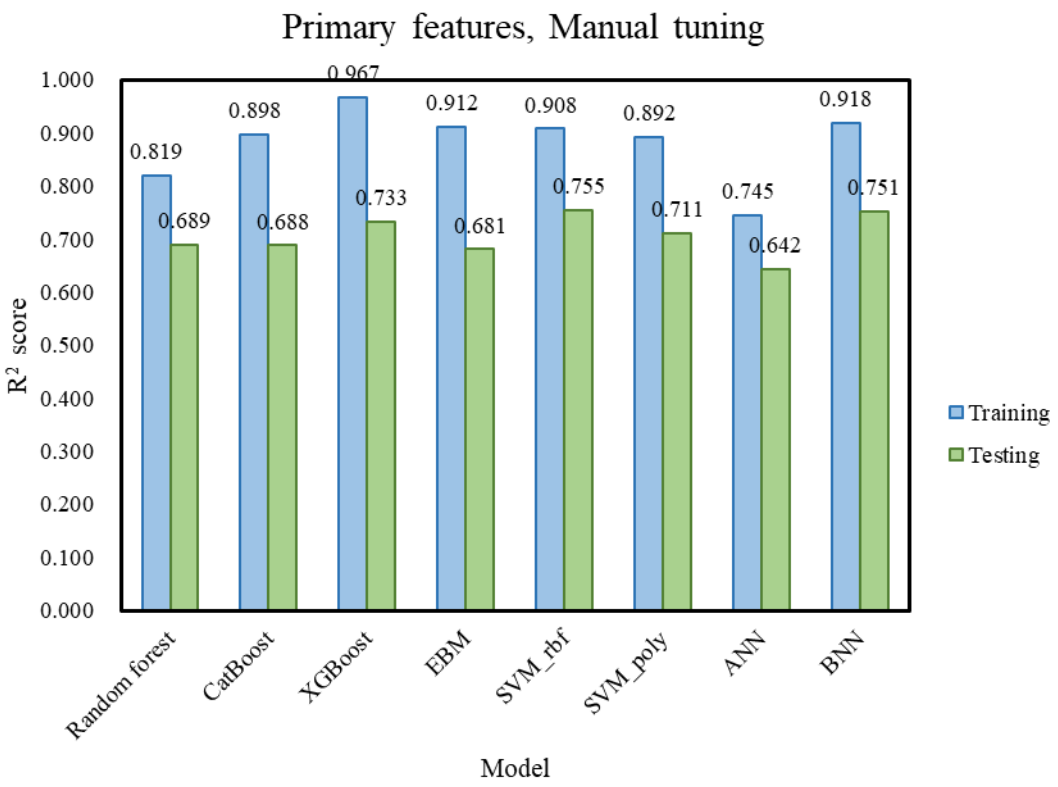
圖 4.7 ANN 不同超參數之 R^2 分數比較圖 (a) epochs； (b) batch_size；(c) drop_rates；(d) factor；(e) max_epoch；(f) use_batch_norm；(g) loss；(h) optimizer；(i) activations

4.2 預測結果比較

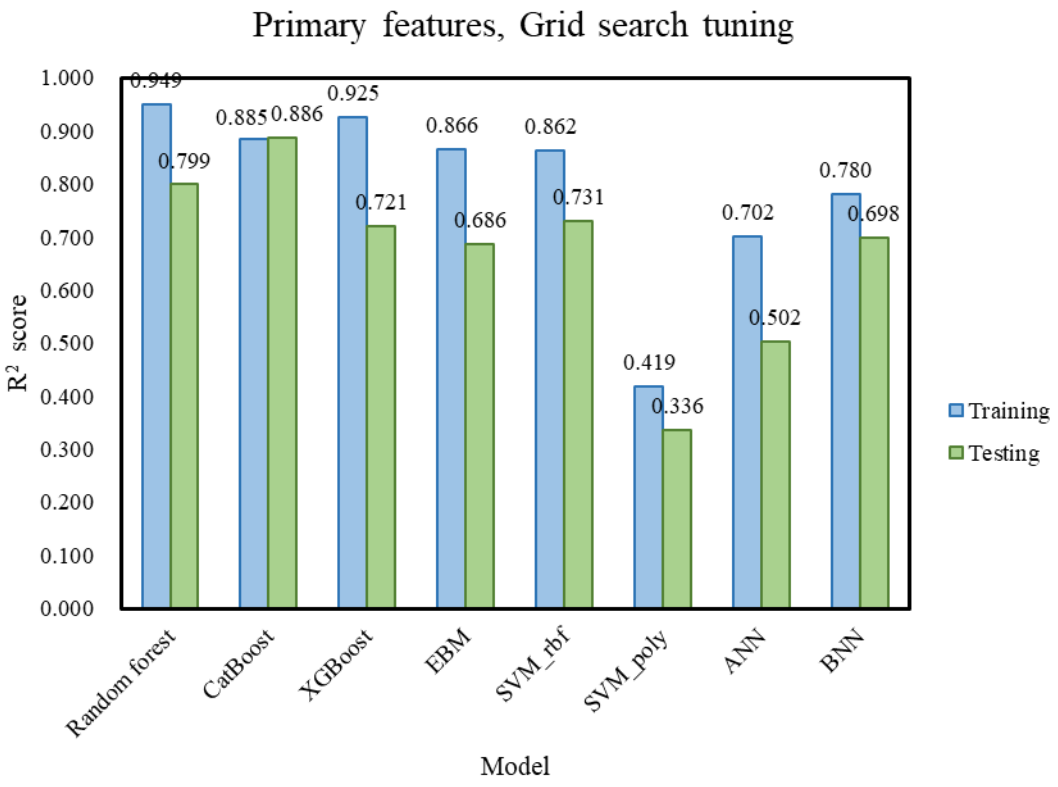
為比較不同輸入條件對模型預測效能之影響，本節分別針對「調參方式」、「特徵組合」與「缺失值填補方式」三個面向進行分析，並篩選出各條件下表現最佳之模型，整理為八組模型預測結果，每組均有八種不同的機器學習模型，統一呈現於圖 4.8 至圖 4.11 中。各組模型之最佳超參數設定，因篇幅限制，已另外整理於附錄 B。

圖 4.8 統整各模型表現，橫軸為各模型名稱，縱軸為其對應之 R^2分數，圖標則標示不同的資料處理條件，涵蓋前述三個面向。為輔助 R^2評估結果之解讀，圖 4.9 至圖 4.11 分別呈現每個模型之 MAE、MSE 與 RMSE 等誤差型指標，提供更全面之預測效能評估依據。

doi:10.6342/NTU202502436

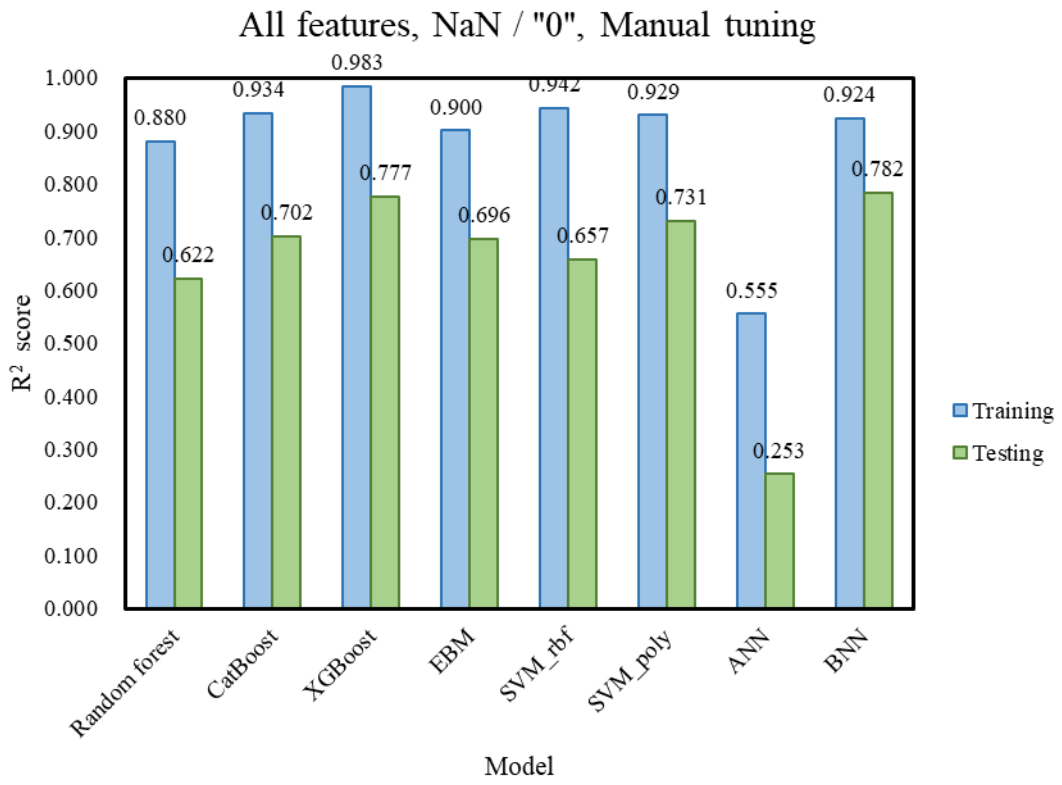


(a)

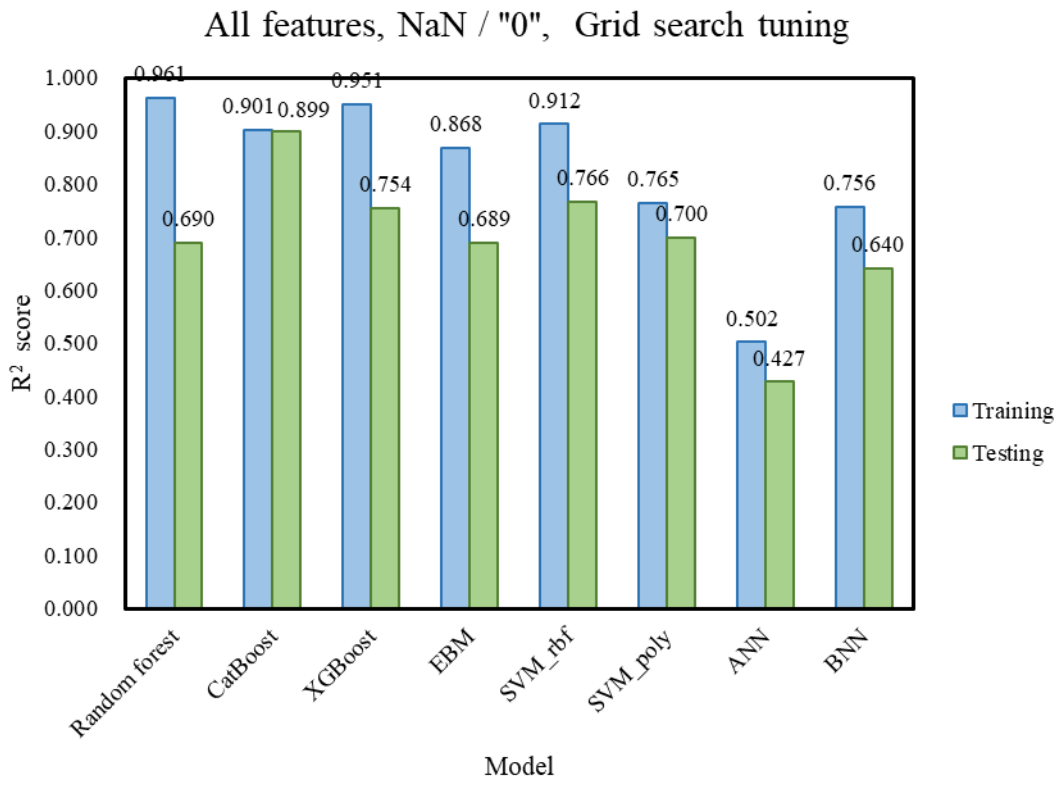


(b)

doi:10.6342/NTU202502436

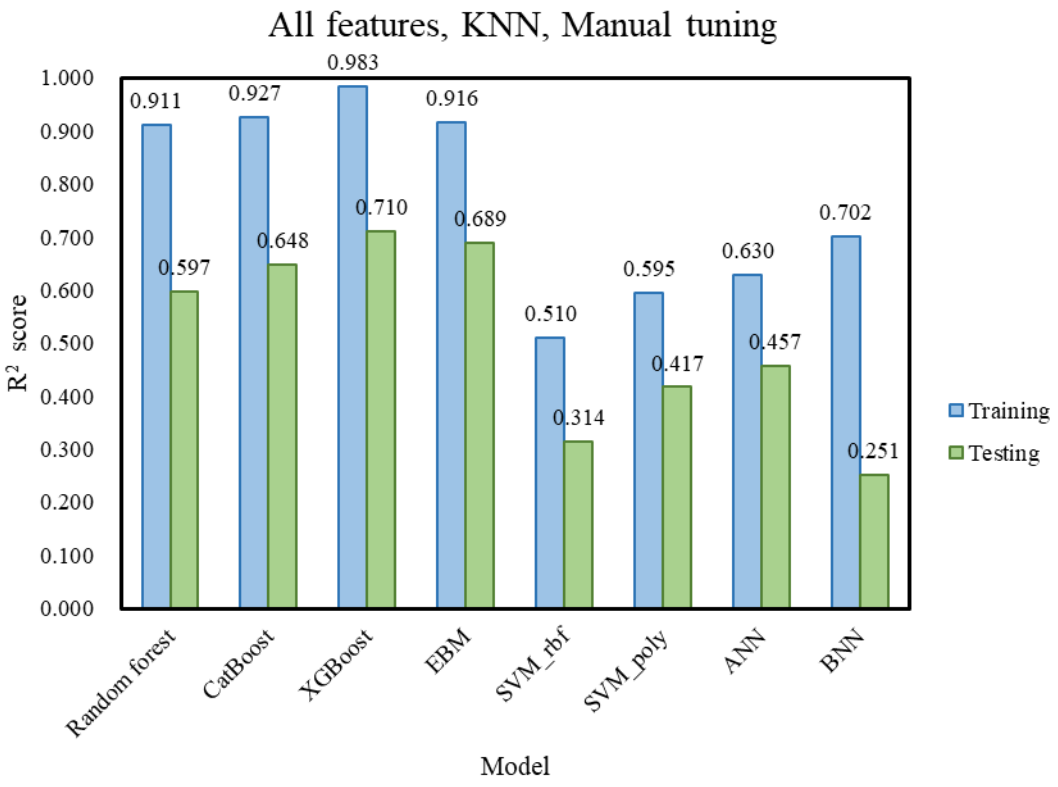


(c)

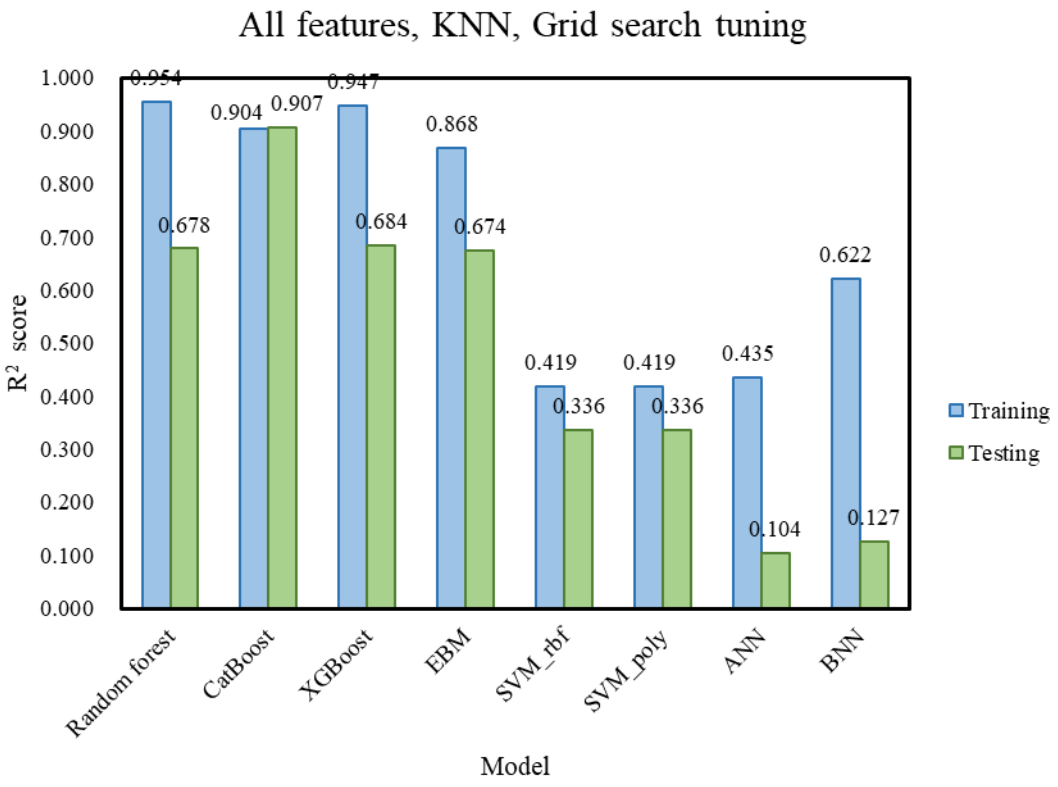


(d)

doi:10.6342/NTU202502436

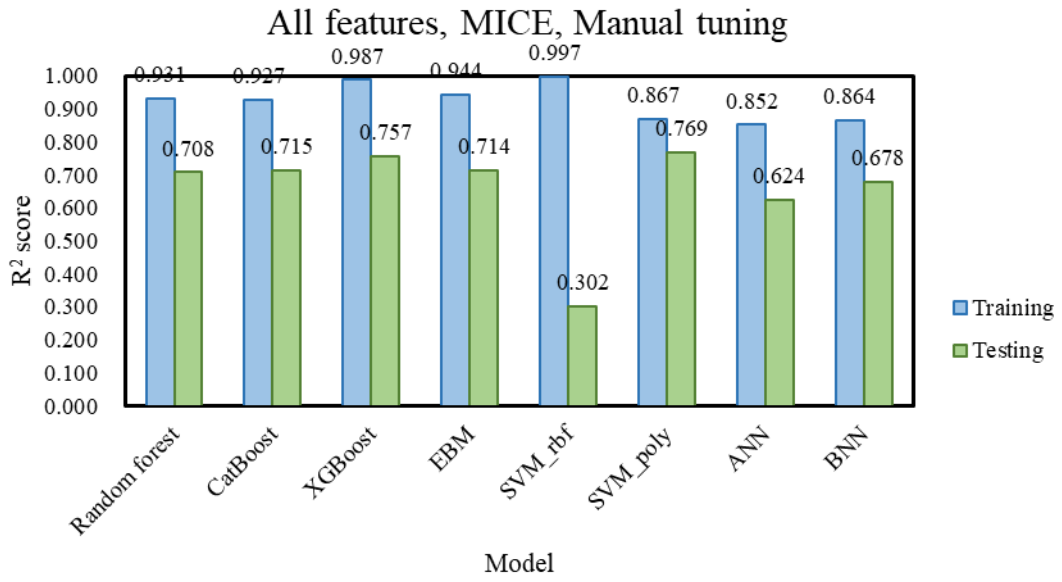


(e)

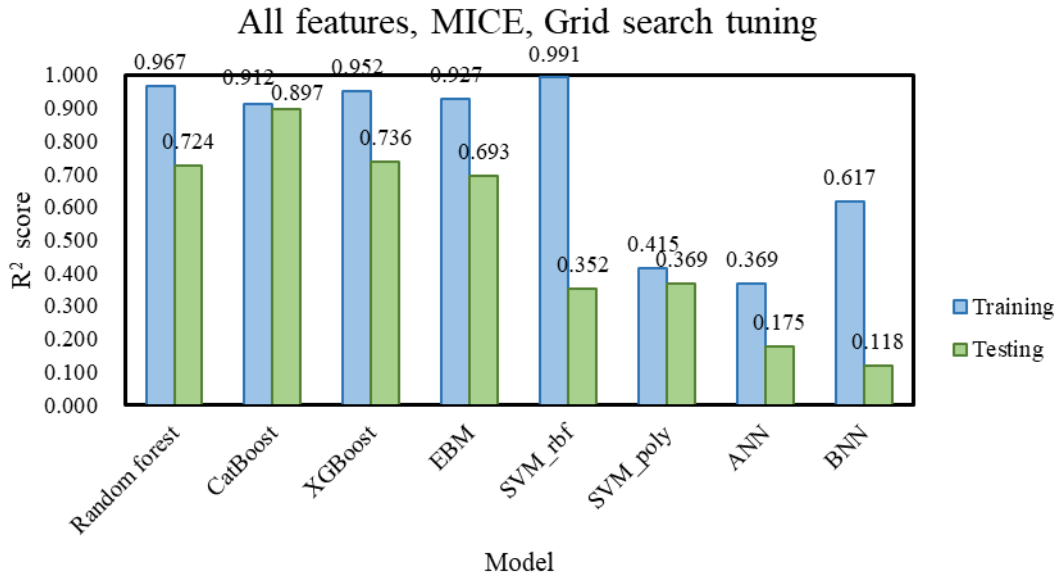


(f)

doi:10.6342/NTU202502436



(g)

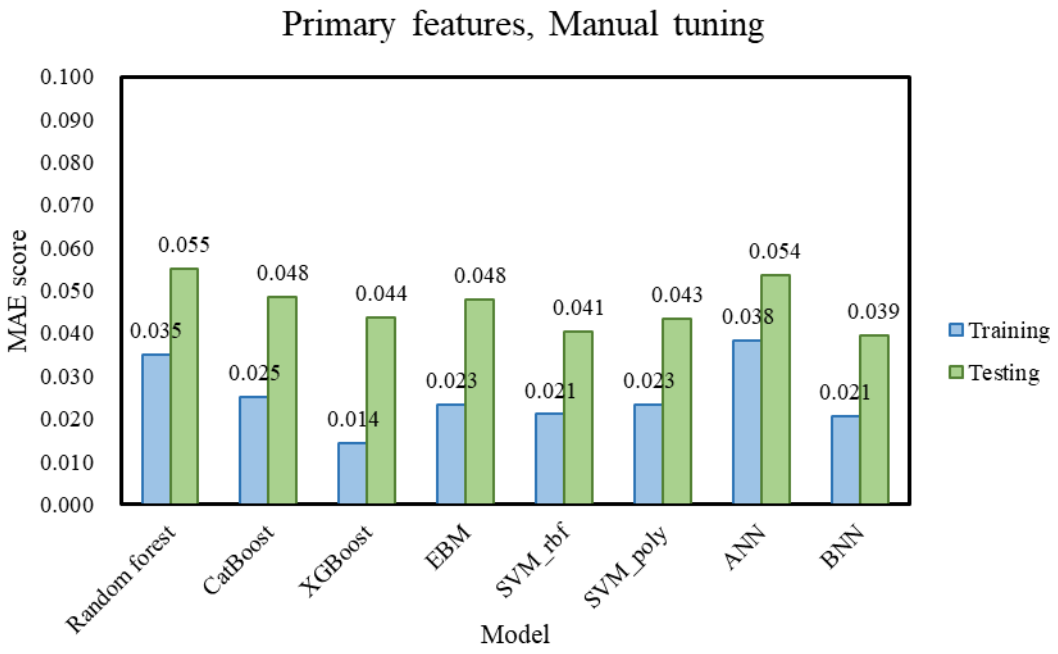


(h)

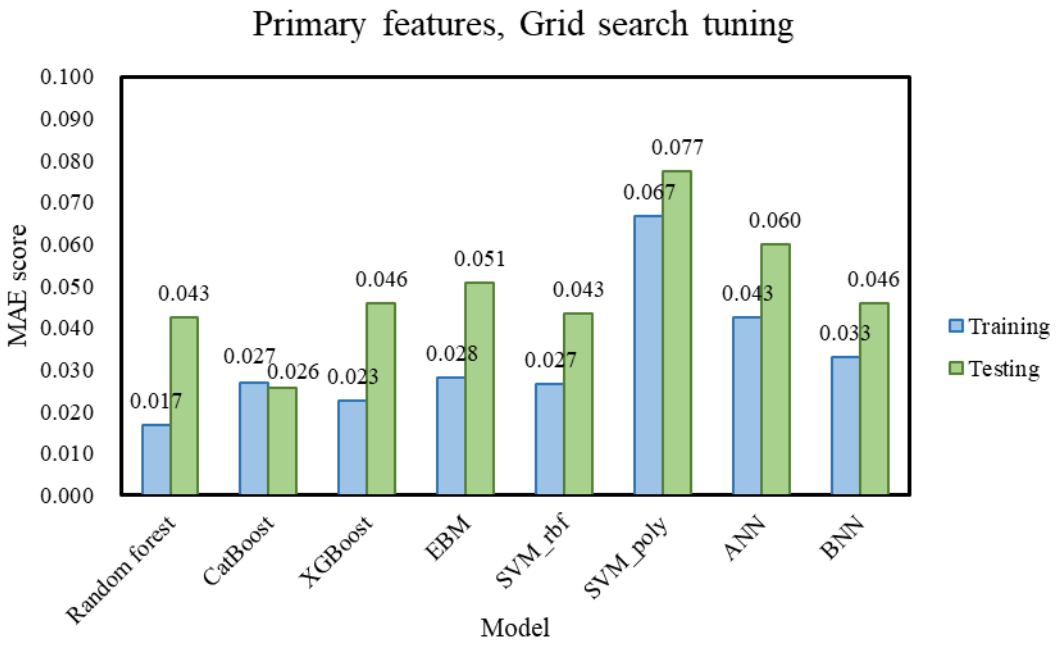
圖 4.8 不同特徵工程下各類模型最佳 R^2 分數比較圖 (a) 包含主要特徵與手動調參；(b) 包含主要特徵與 Grid search 調參；(c) 包含所有特徵、以 NaN 或 0 填補與手動調參；(d) 包含所有特徵、以 NaN 或 0 填補與 Grid search 調參；(e) 包含所有特徵、以 KNN 填補與手動調參；(f) 包含所有特徵、以 KNN 填補與 Grid search 調參；(g) 包含所有特徵、以 MICE 填補與手動調參；(h) 包含所有特徵、以 MICE 填補與 Grid search 調參

doi:10.6342/NTU202502436

圖 4.9 顯示不同資料處理條件下，各模型於訓練與測試階段之 MAE (Mean Absolute Error) 表現。整體而言，誤差值與 R^2 分數呈負相關。多數模型在手動調參下於測試集的誤差明顯小於 Grid search 調參組，顯示其泛化能力佳。而在填補策略中，KNN 填補與部分 Grid search 模型的 MAE 普遍偏高，與 R^2 評估中表現較差的結果一致。

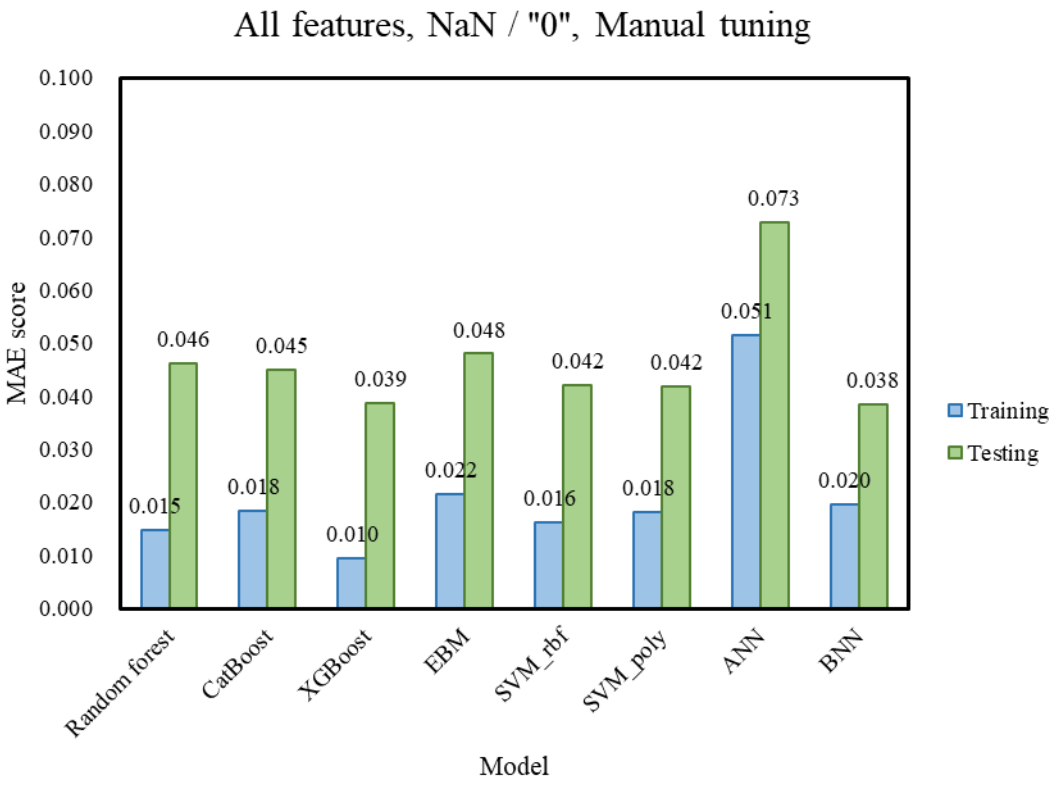


(a)

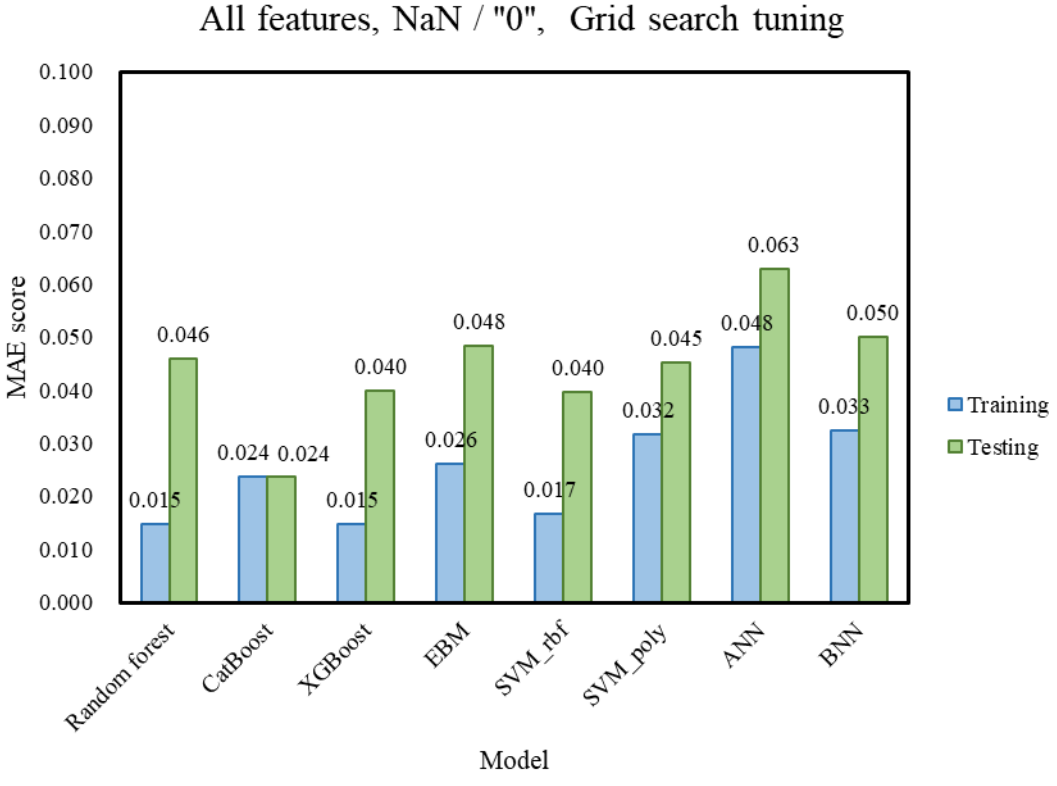


(b)

doi:10.6342/NTU202502436

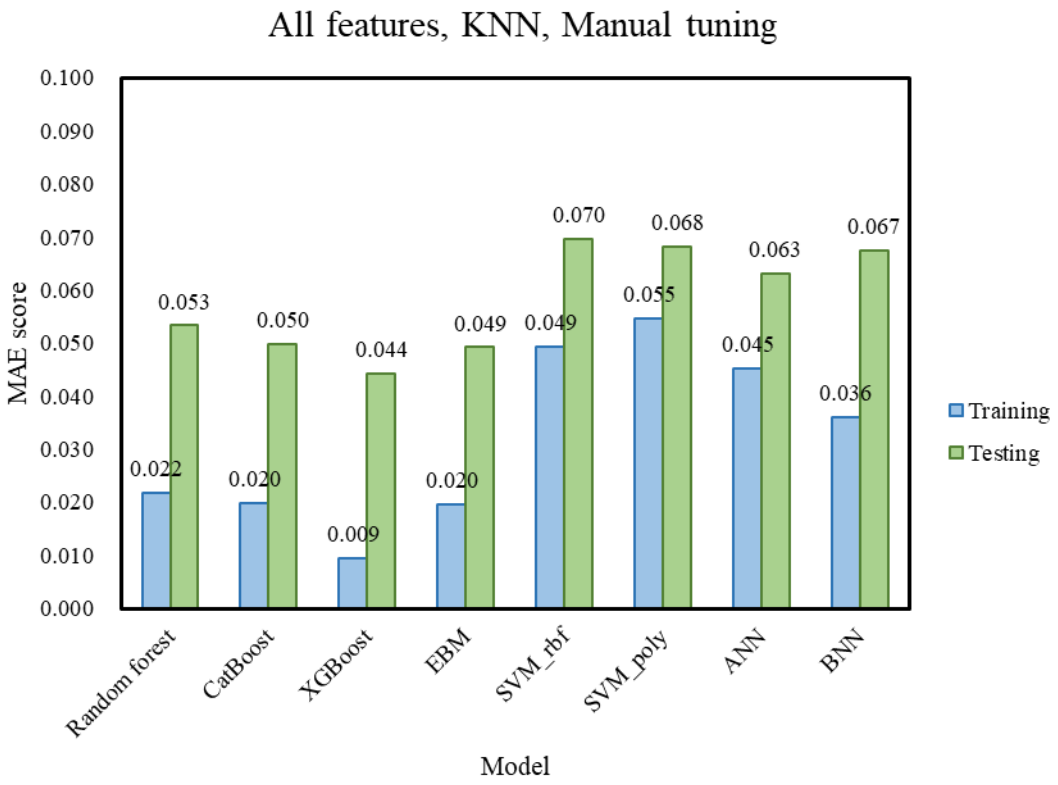


(c)

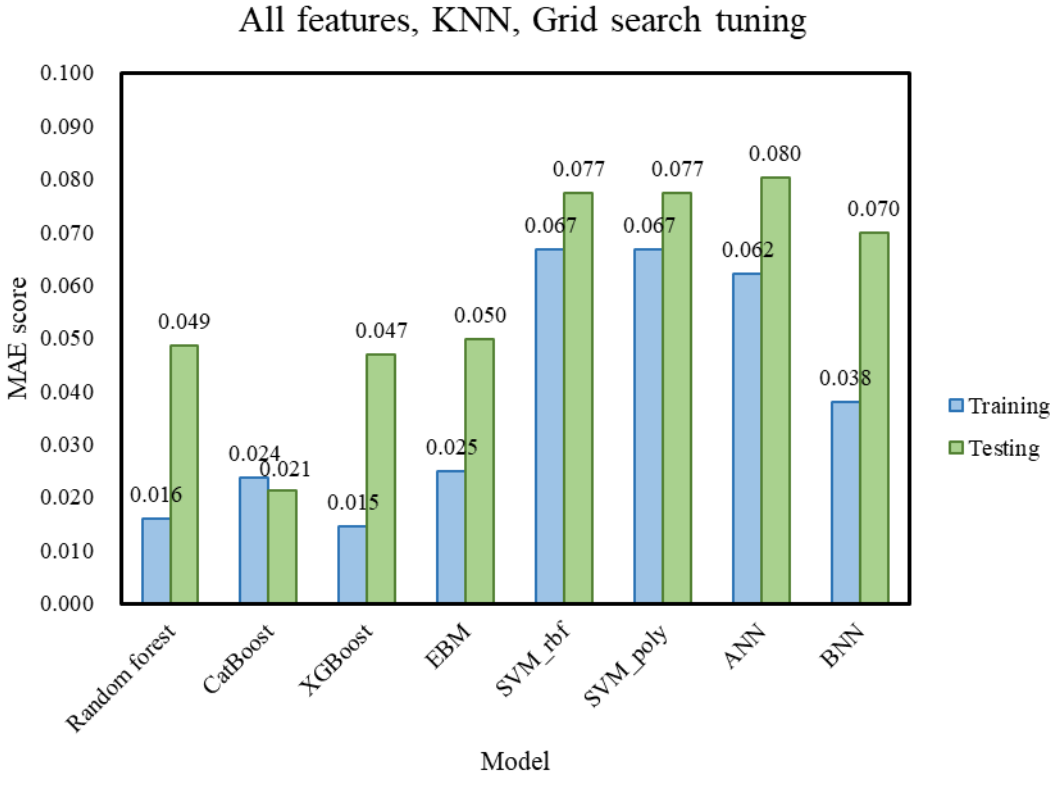


(d)

doi:10.6342/NTU202502436

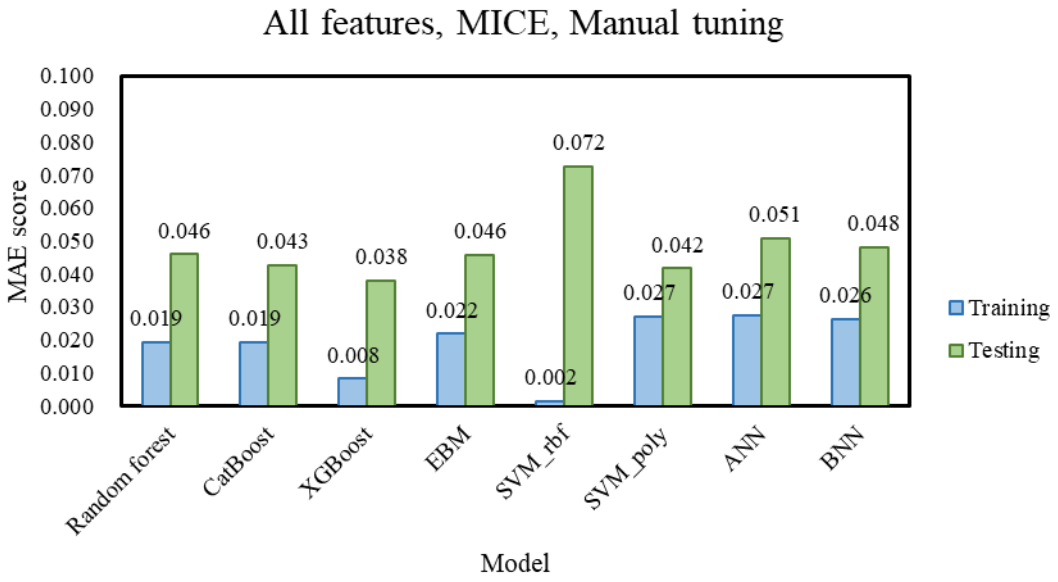


(e)

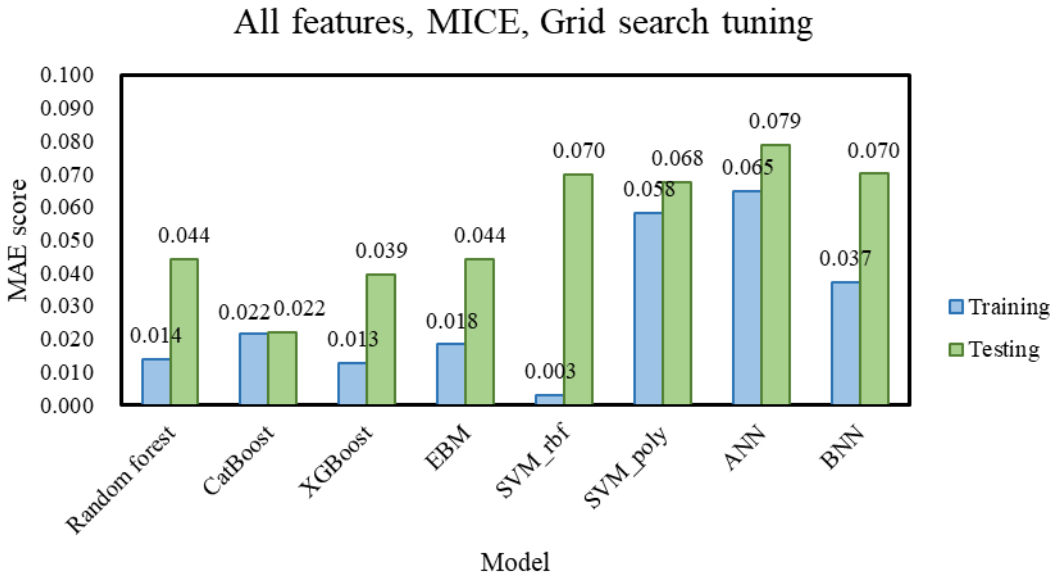


(f)

doi:10.6342/NTU202502436



(g)

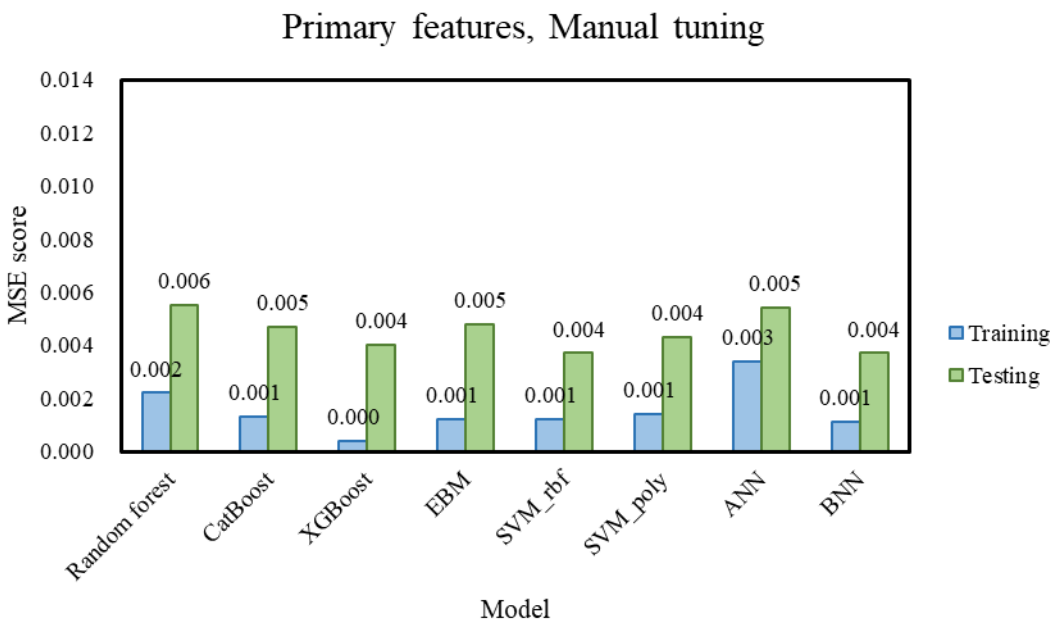


(h)

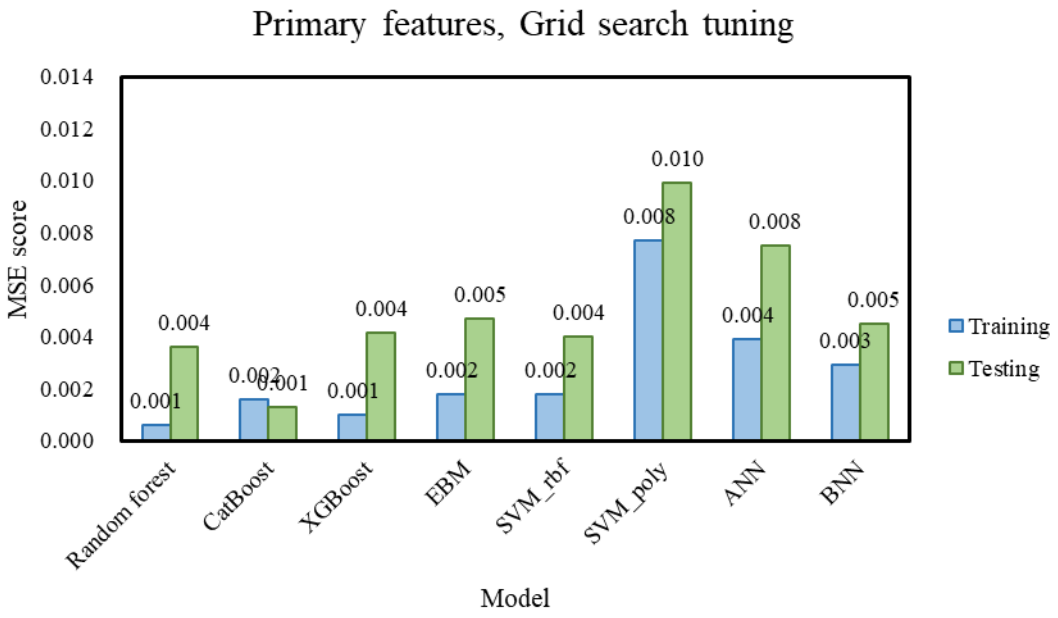
圖 4.9 不同特徵工程下各類模型之 MAE 分數比較圖；(a) 包含主要特徵與手動調參；(b) 包含主要特徵與 Grid search 調參；(c) 包含所有特徵、以 NaN 或 0 填補與手動調參；(d) 包含所有特徵、以 NaN 或 0 填補與 Grid search 調參；(e) 包含所有特徵、以 KNN 填補與手動調參；(f) 包含所有特徵、以 KNN 填補與 Grid search 調參；(g) 包含所有特徵、以 MICE 填補與手動調參；(h) 包含所有特徵、以 MICE 填補與 Grid search 調參

doi:10.6342/NTU202502436

圖 4.10 呈現不同組別在訓練與測試階段的 MSE (Mean Squared Error) 結果，與 MAE 類似，MSE 亦可作為模型穩定性的指標，由於 MSE 對極端誤差更為敏感，可觀察到部分模型 (如 ANN、BNN 在 KNN 或 MICE 填補下) 於測試階段出現明顯誤差放大現象，表示模型學習過程受不穩定資料干擾，進而降低其泛化能力。

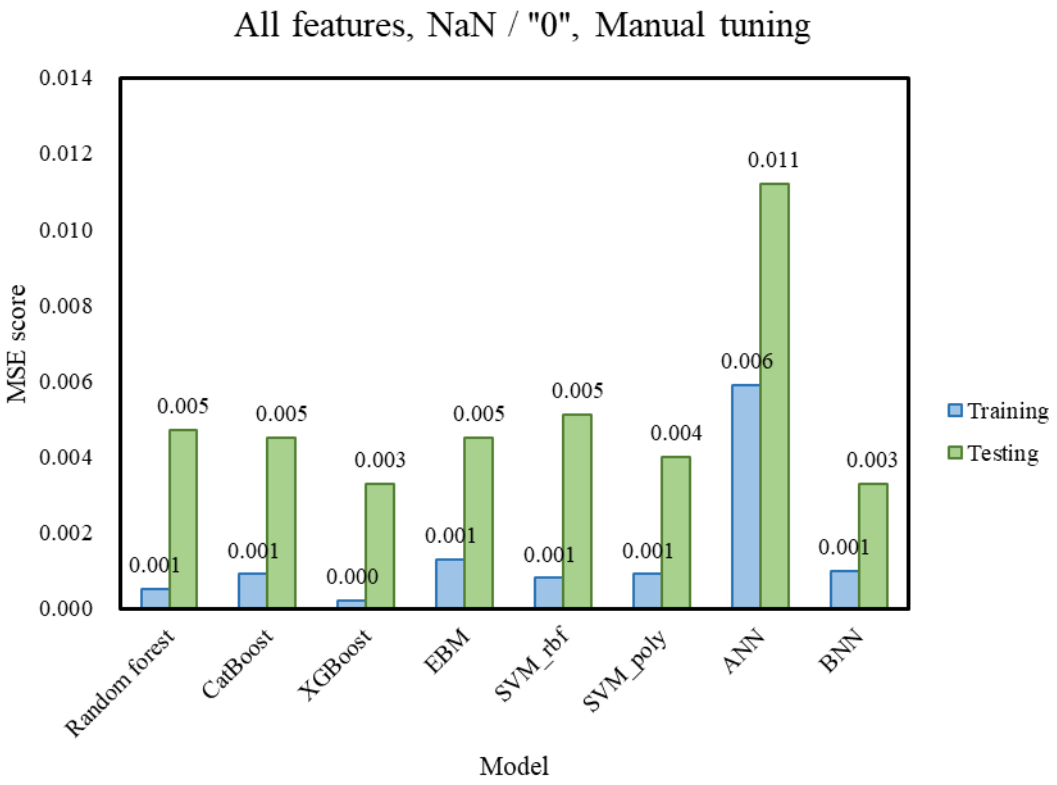


(a)

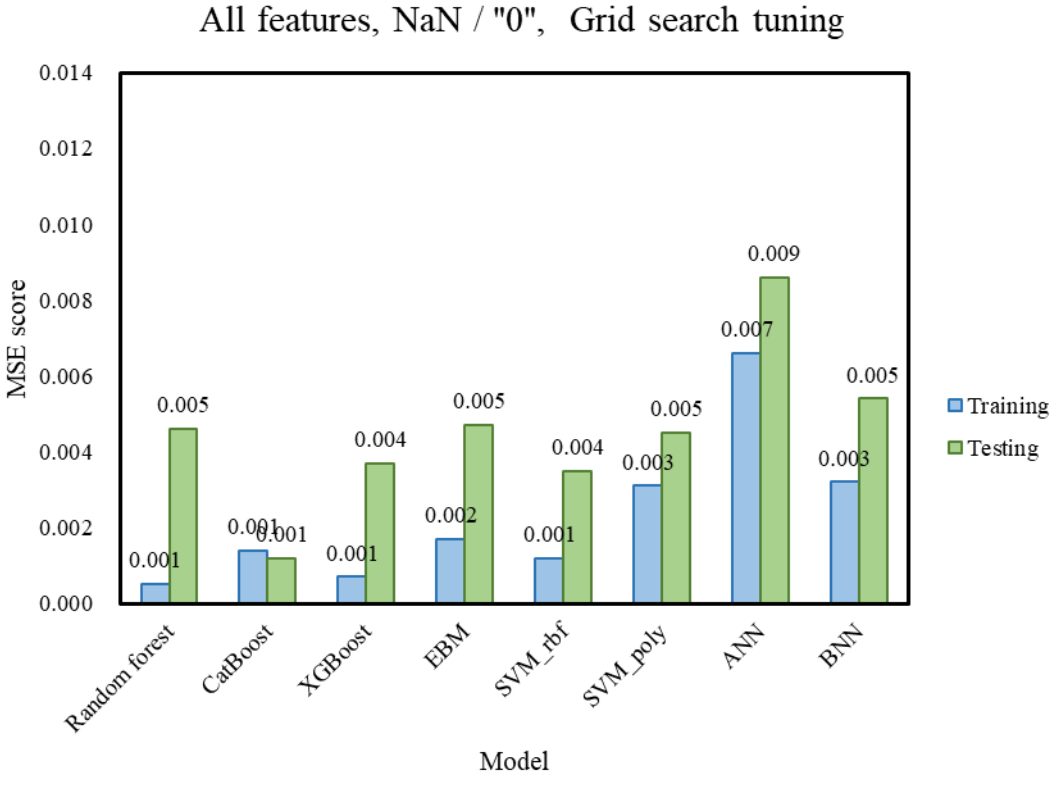


(b)

doi:10.6342/NTU202502436

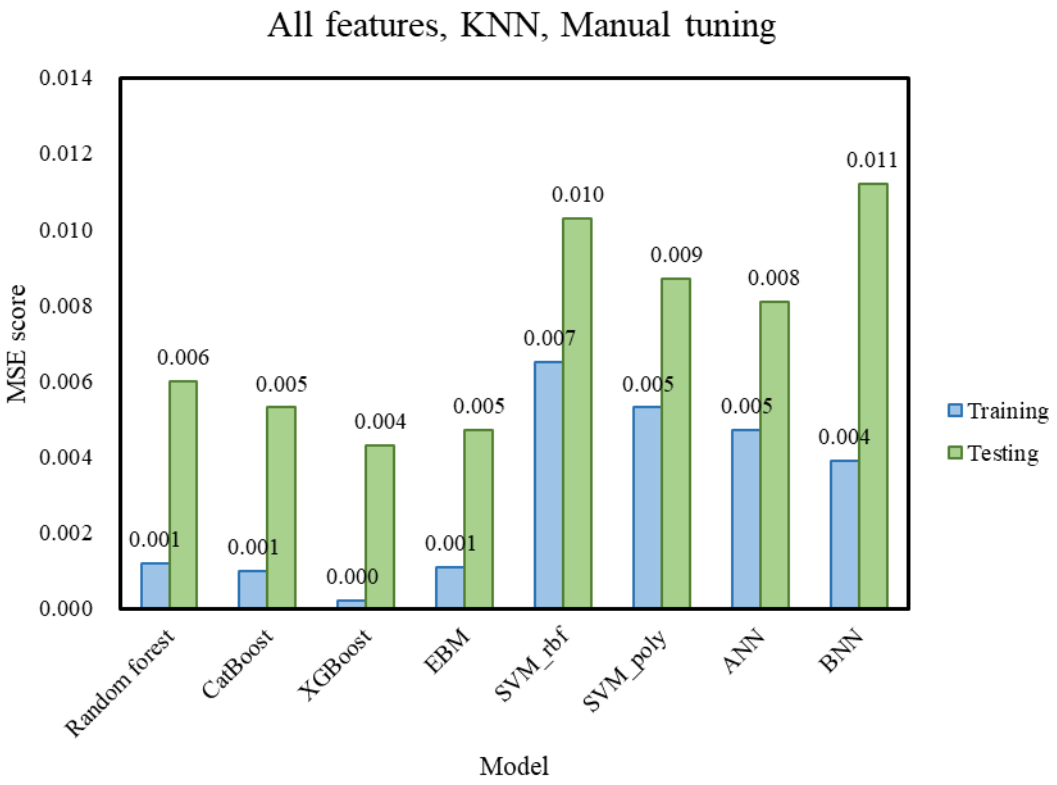


(c)

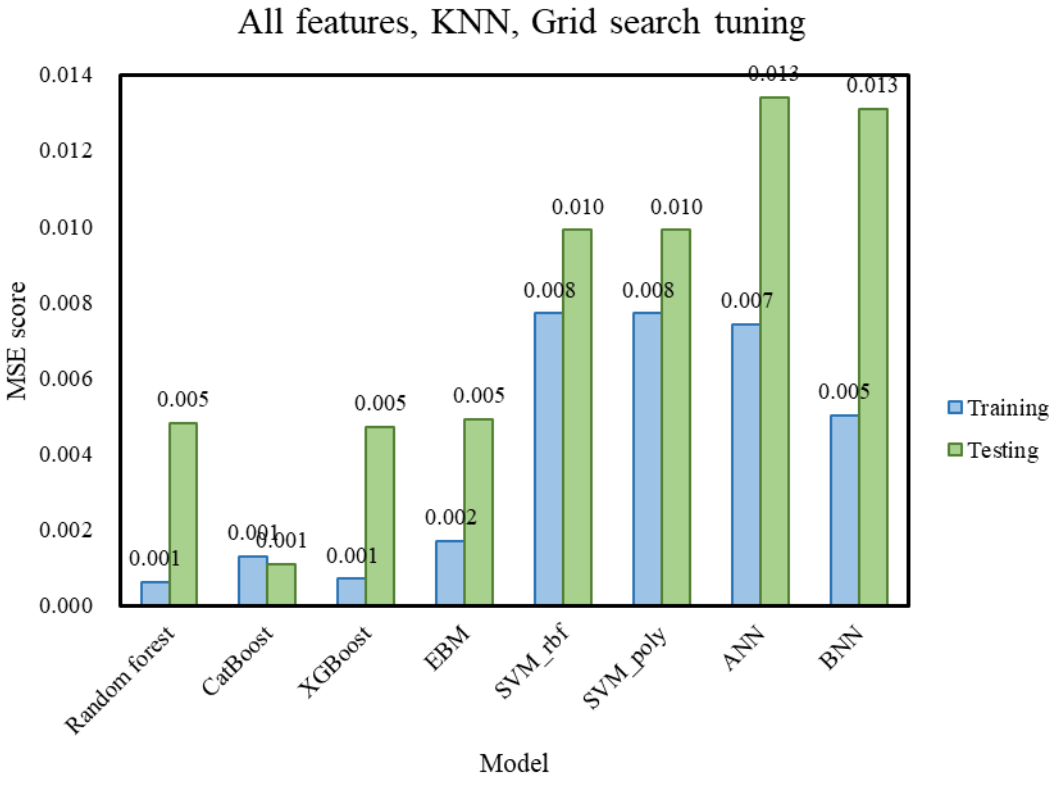


(d)

doi:10.6342/NTU202502436

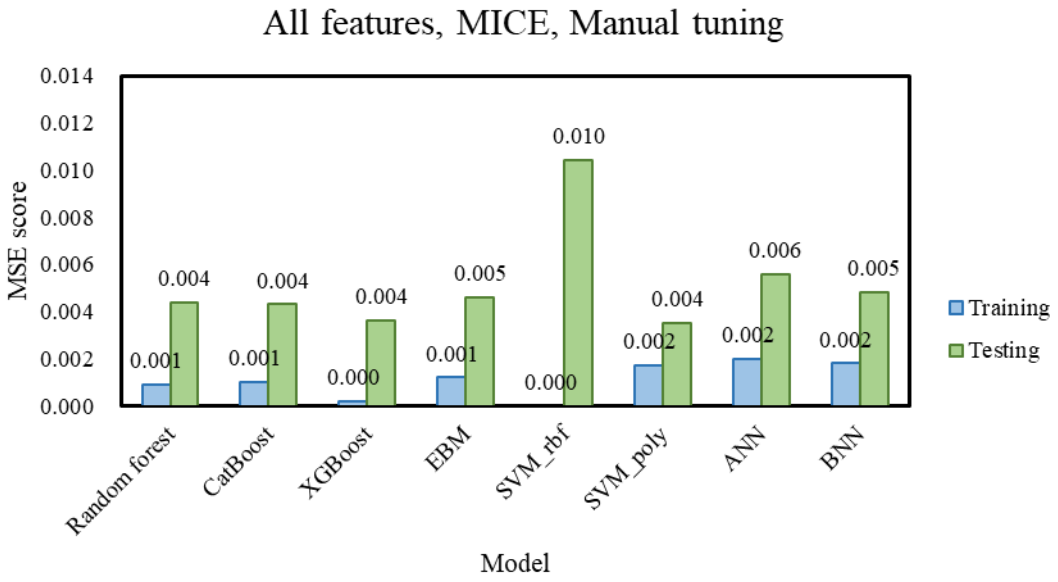


(e)

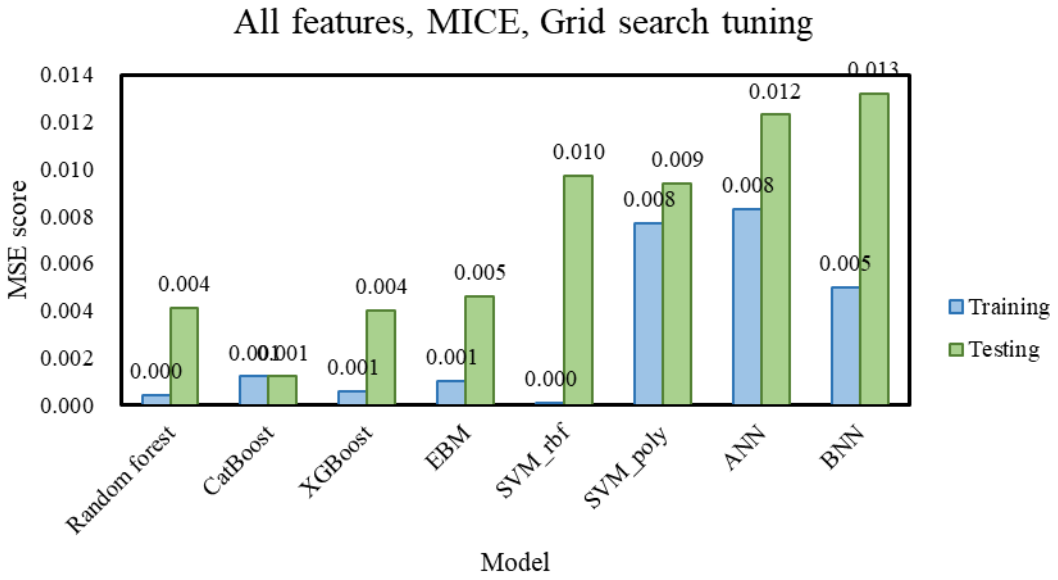


(f)

doi:10.6342/NTU202502436



(g)

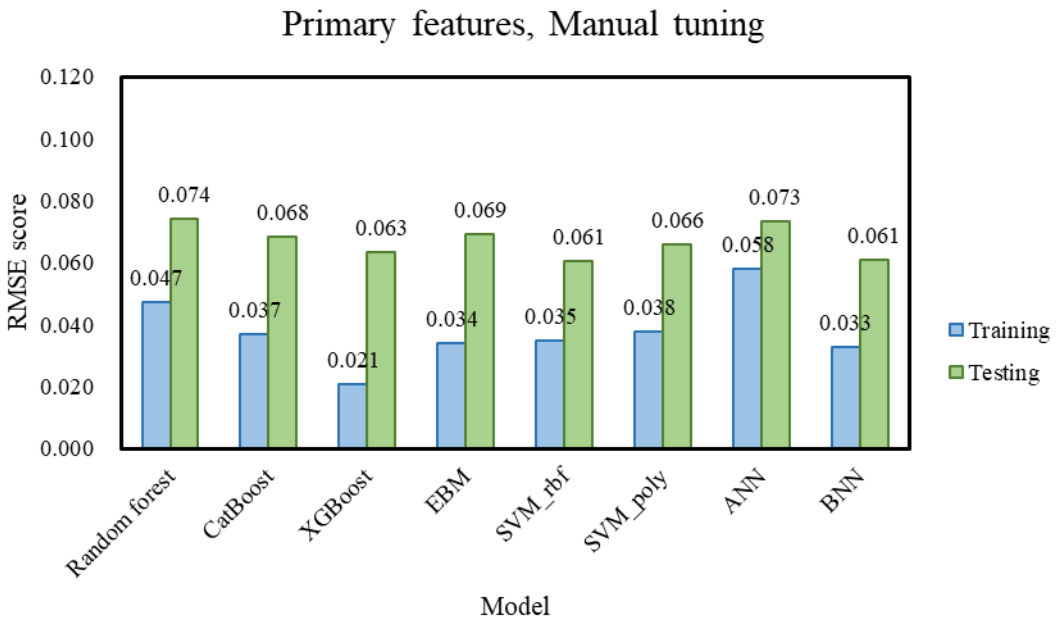


(h)

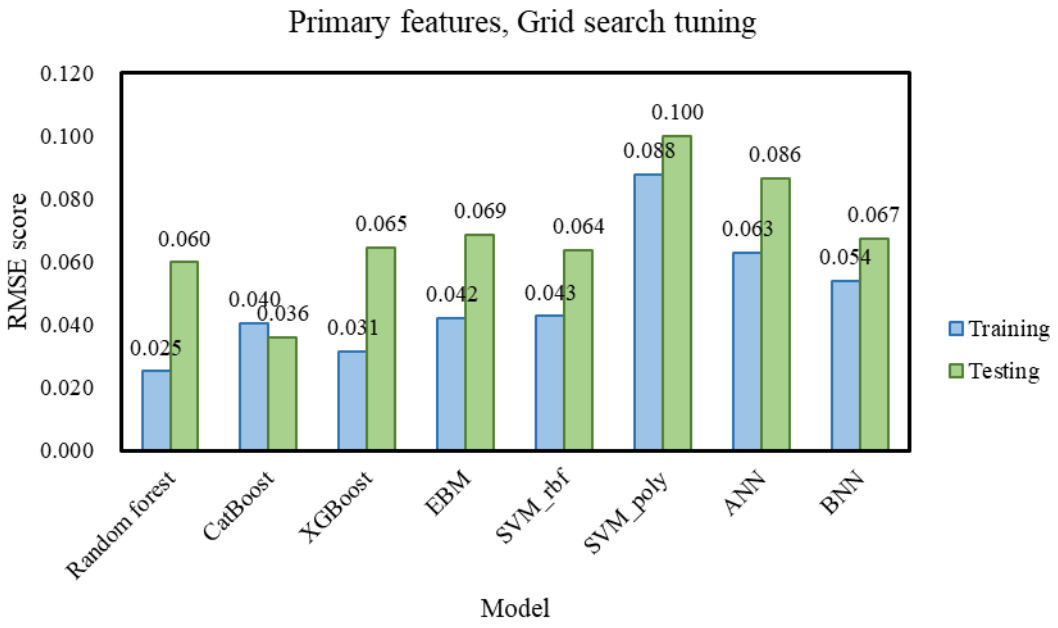
圖 4.10 不同特徵工程下各類模型之 MSE 分數比較圖；(a) 包含主要特徵與手動調參；(b) 包含主要特徵與 Grid search 調參；(c) 包含所有特徵、以 NaN 或 0 填補與手動調參；(d) 包含所有特徵、以 NaN 或 0 填補與 Grid search 調參；(e) 包含所有特徵、以 KNN 填補與手動調參；(f) 包含所有特徵、以 KNN 填補與 Grid search 調參；(g) 包含所有特徵、以 MICE 填補與手動調參；(h) 包含所有特徵、以 MICE 填補與 Grid search 調參

doi:10.6342/NTU202502436

圖 4.11 為 RMSE (Root Mean Squared Error) 比較圖，與 MSE 具有相同的誤差度量概念，但因其單位與原始資料移至，具高直觀解讀性。大部分模型在 RMSE 的表現趨勢 MAE 和 MSE 一致，與 R^2 評估結果同樣具有高度一致性，可做為 R^2 評估結果的補充說明。

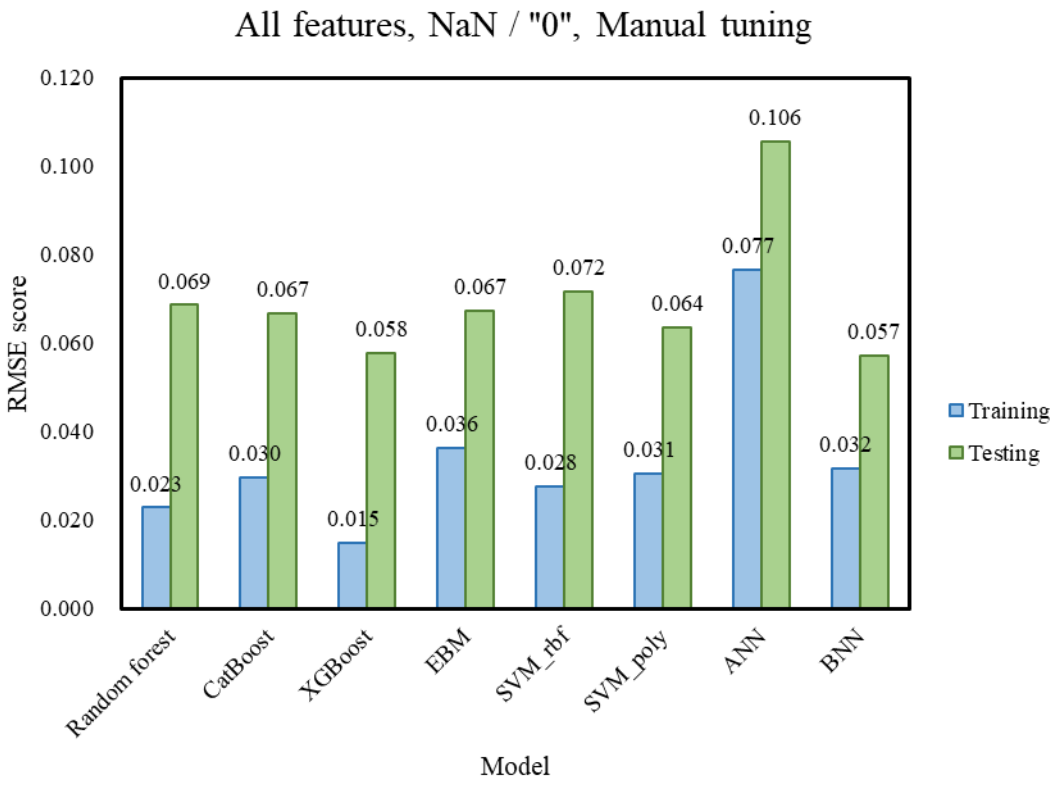


(a)

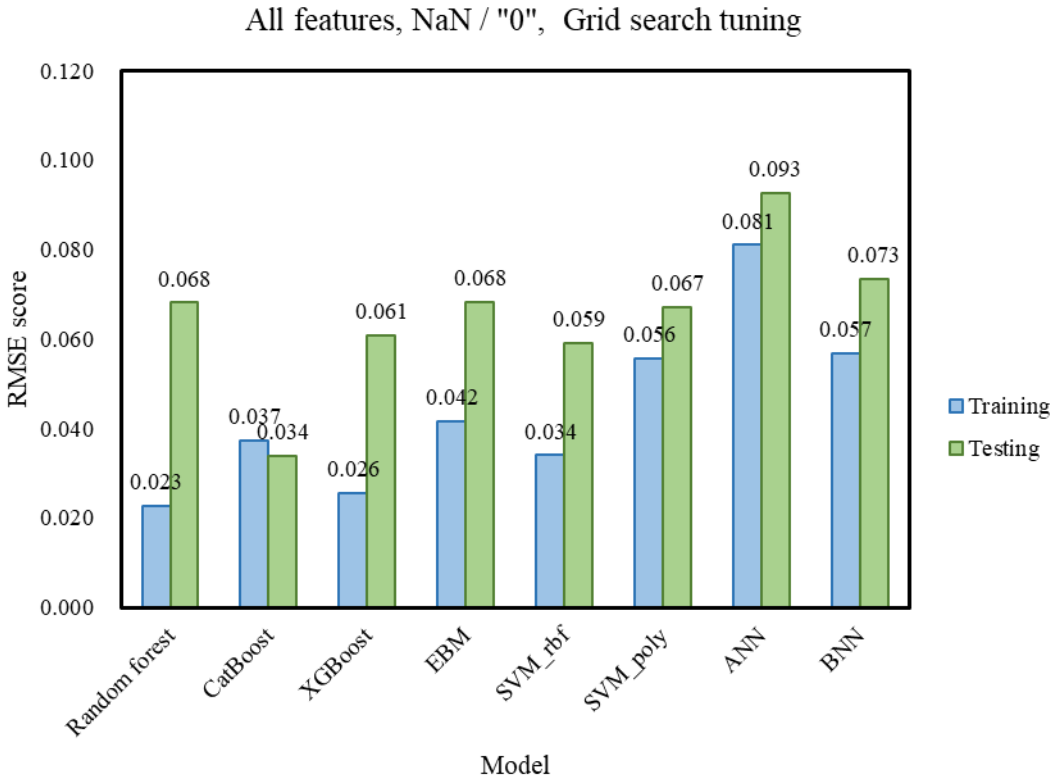


(b)

doi:10.6342/NTU202502436

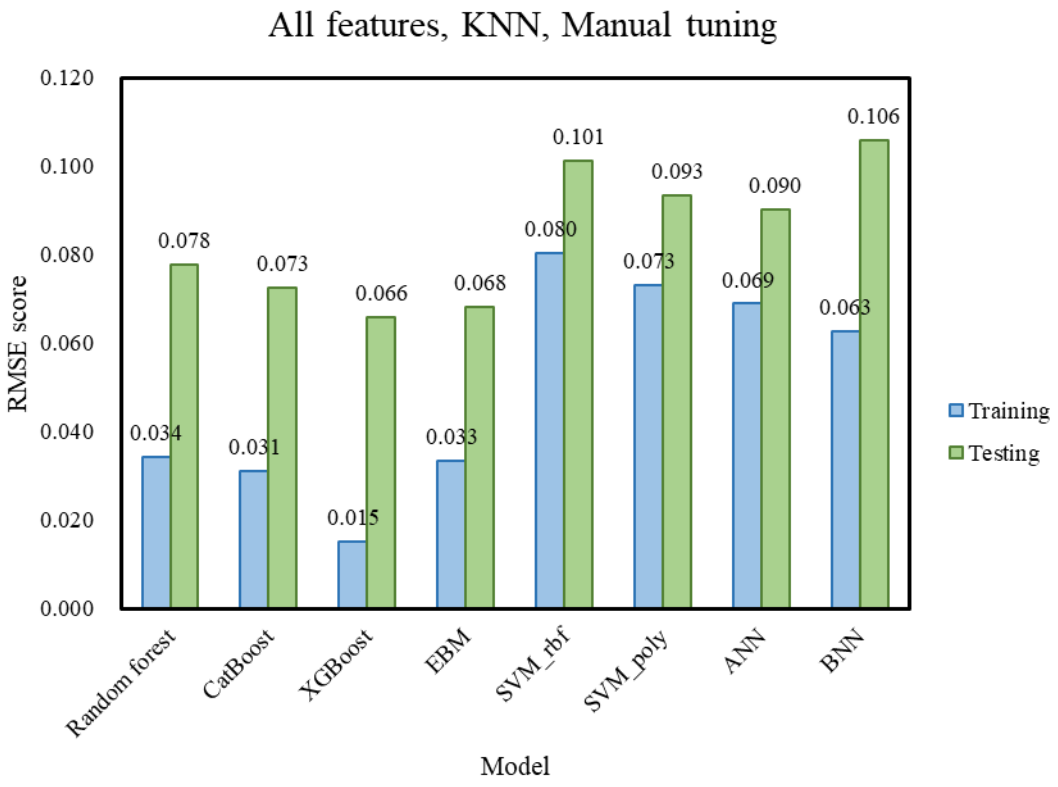


(c)

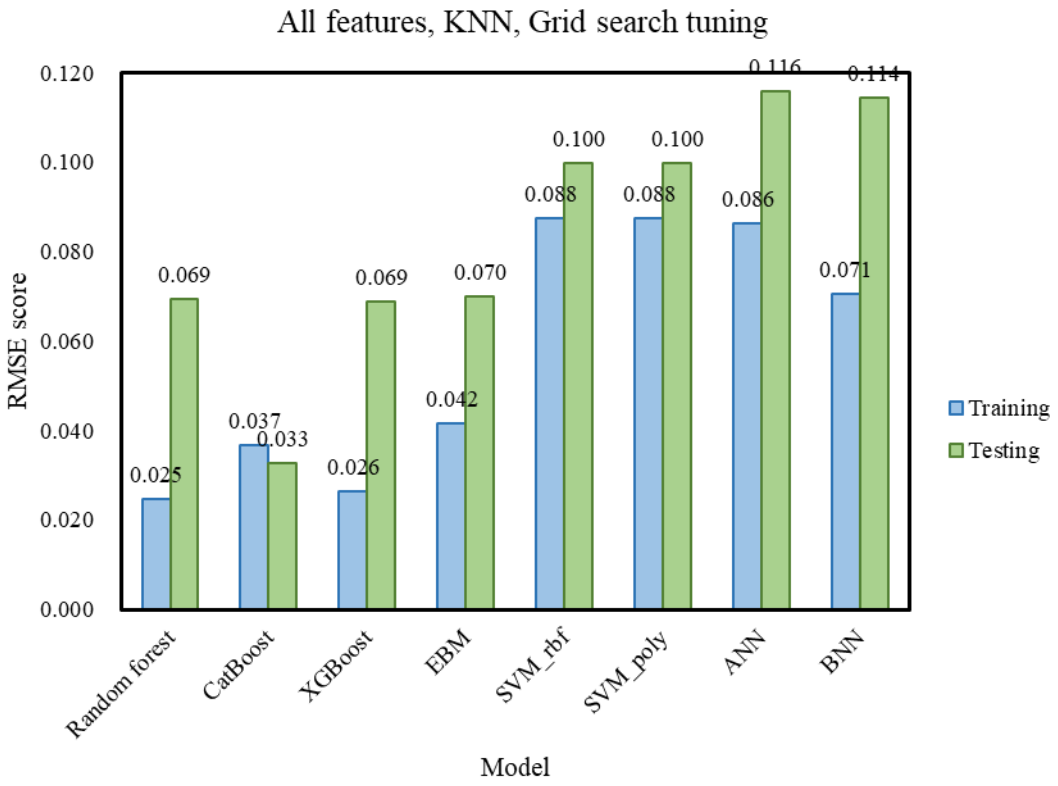


(d)

doi:10.6342/NTU202502436

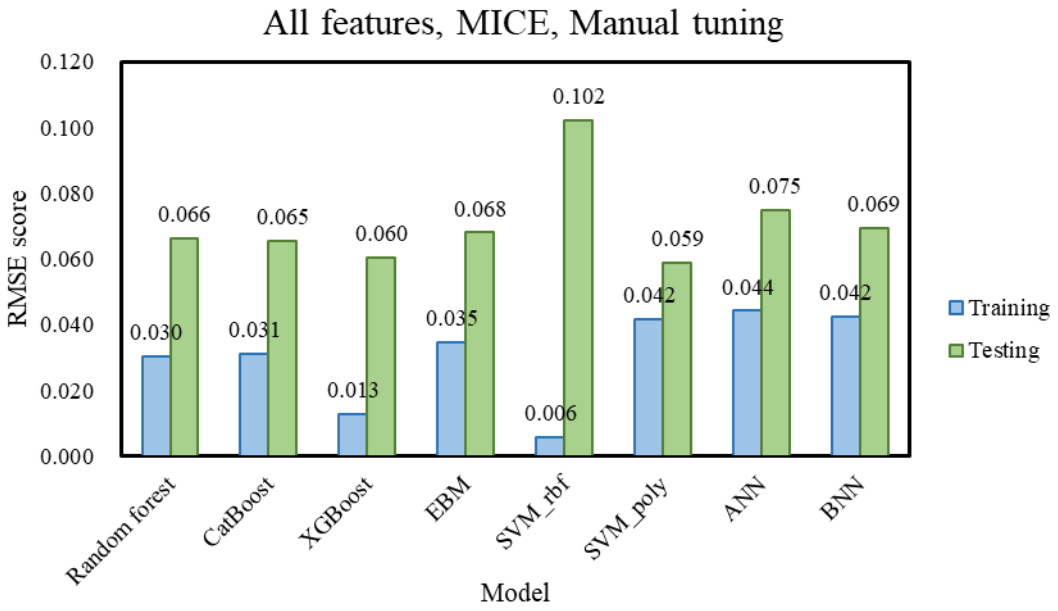


(e)

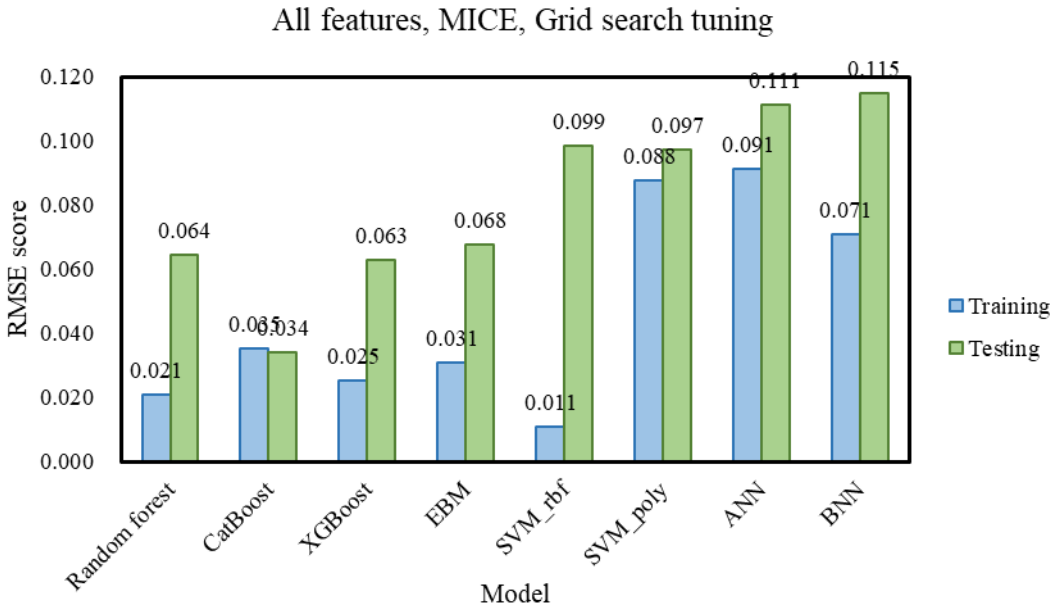


(f)

doi:10.6342/NTU202502436



(g)



(h)

圖 4.11 不同特徵工程下各類模型之 RMSE 分數比較圖 (a) 包含主要特徵與手動調參；(b) 包含主要特徵與 Grid search 調參；(c) 包含所有特徵、以 NaN 或 0 填補與手動調參；(d) 包含所有特徵、以 NaN 或 0 填補與 Grid search 調參；(e) 包含所有特徵、以 KNN 填補與手動調參 (f) 包含所有特徵、以 KNN 填補與 Grid search 調參 (g) 包含所有特徵、以 MICE 填補與手動調參；(h) 包含所有特徵、以 MICE 填補與 Grid search 調參

doi:10.6342/NTU202502436

各組比較結果將於後續小節中依變因依序說明，並統一以 R^2 分數為主要評估指標，逐一探討不同條件對預測表現與模型穩定性之影響。

4.2.1 不同調參方式

結果觀察

由圖 4.8 可觀察到，在相同的資料條件下，手動調參方式 (圖 4.8 (a)、(c)、(e)、(g)) 整體上相較於 Grid search 調參 (圖 4.8 (b)、(d)、(f)、(h)) 表現更為穩定。

整體而言，大部分傳統機器學習模型 (如 Random Forest、CatBoost、XGBoost、EBM) 於兩種調參方式下皆能維持穩定，測試 R^2 分數多介於 0.68 至 0.77，表示其對參數設定之敏感度較低，僅需合理設定即具良好的泛化能力。

相對的，SVM 與深度學習模型 (ANN 與 BNN) 對調參策略較為敏感。SVM 模型在 Grid search 調參方式中常出現過擬合或收斂不穩的狀況。深度學習模型表現更為極端，在 Grid search 條件下常見測試 R^2 分數低於 0.5，甚至低於 0.2，反映出其對超參數組合依賴性高，亦顯示手動調參的重要性。

原因解析

1. Grid search 依賴預設的參數組合，可能錯失最佳解

Grid search 雖然具備系統性與全面性，但其效果受限於使用者所設定的超參數組合範圍與離散程度。若搜尋範圍未涵蓋最佳解所在區段，即使耗費大量時間與計算資源，仍可能找不到合適的參數組合。此外，Grid search 所需之計算成本相對較高，尤其當參數空間維度增加時，所需運算之組合呈指數成長。相較之下，若使用者對模型特性與超參數影響具有一定經驗，透過手動方式進行有策略性的

doi:10.6342/NTU202502436

調整，往往能在更短時間內找出合理且具代表性的最佳參數組合，提升調參效率與實務可行性。

2. Grid search 所選最佳參數組合未必為效能最佳組合

雖然 Grid search 具備系統性地歷遍參數組合的優點，但在實際操作中觀察到手動調參所得之最佳組合，雖在 Grid search 結果中排名第二或第三，卻擁有更高的測試準確率，這可能與 Grid search 所使用之交叉驗證策略傾向選擇誤差較穩定的組合有關，而非僅依賴單一指標之最佳值，藉以避免過擬合風險。

3. 深度學習模型超參數眾多，Grid search 難以涵蓋所有組合

在建立 ANN 與 BNN 等深度模型時，需同時設定多層隱藏層、每層神經元數、學習率、激活函數、Batch size、Dropout rate 等多個維度的超參數，即使每個參數僅設三個值，總組合數亦可能高達數百甚至數千組，導致計算成本急遽上升。在時間成本的限制下，往往會縮小搜尋空間，進一步侷限了參數探索的範圍，進而無法捕捉到與資料特性相符的最佳組合，導致模型學習表現不佳。

4. 深度模型在小樣本情境下容易過擬合

本研究資料量稱不上龐大，在特徵數量與填補後維度較多的情況下，深度學習模型雖能快速擬合訓練資料，卻無法有效泛化至測試資料，導致模型表現在測試資料上顯著下滑。若 Grid search 未有效加入防止過擬合的設計 (如適當的 Dropout rate)，更容易產生此問題。

4.2.2 不同特徵組合

為探討特徵組合對模型預測表現之影響，本節選取兩組未進行缺失值填補之結果進行比較，分別為僅包含主要特徵 (圖 4.8 (a)、(b)) 與包含所有特徵 (圖 4.8 (c)、(d)) 之模型。此兩組模型在資料處理流程上最為單純，均未受填補所造成之雜訊，具備良好的可比較性，適合做為特徵組合影響的討論對象。

doi:10.6342/NTU202502436

結果觀察

由圖 4.8 可見，在測試資料的 R^2 分數上，包含所有特徵之模型 (圖 4.8 (c)、(d)) 表現普遍略優於僅含主要特徵的模型 (圖 4.8 (a)、(b))，此趨勢在傳統機器學習模型 (如 XGBoost 和 EBM) 及深度學習模型 (如 BNN) 中皆有出現。

然而，在訓練資料上的分數方面，兩者差異不大，顯示雖增加特徵可能略微提升模型學習效果，但整體效益有限，無法有效提升模型對測試資料的泛化能力。部分模型如 ANN 與使用 polynomial 核的 SVM 甚至在加入次要特徵後測試分數下降，反映出特徵擴充亦可能導致資料為度增加而干擾模型關注重點。

原因解析

1. 主要特徵已涵蓋大部分關鍵資訊

本研究所挑選之主要特徵 (如孔隙比、含細粒料與循環剪切次數) 多為文獻中出現頻率較高的變數，具備良好的物理意義與解釋性。此類特徵可能已足夠提供模型建立預測能力所需的關鍵資訊，使得即使未加入其他次要特徵，模型亦能達到合理準確度。

2. 次要特徵貢獻有限，且存在特徵擴充風險

次要特徵雖可能包含部分與預測目標相關的資訊，加入資料集後可做為額外輔助，有助於模型掌握更完整的資料特性，進而略微提升預測表現。然而，若其與目標變數之關聯性不強或解釋力有限，則其實質貢獻將相對微弱。此外，特徵數量的增加亦會提升模型複雜度，若未能有效提供額外資訊，反而會導致學習過程受到干擾，影響模型專注於關鍵因子的能力。綜合而言，特徵擴充並非總是有益，若無法加入品質較好的輔助資料，則其對模型表現的提升有限，而本研究中即展現了此特性。

doi:10.6342/NTU202502436

4.2.3 不同填補方式

為分析缺失值填補方式對模型預測表現之影響，本節選取皆使用全部特徵之資料集進行比較，並分別採用填入 NaN 或 0 (圖 4.8 (c)、(d))、KNN 填補 (圖 4.8 (e)、(f)) 與 MICE 填補 (圖 4.8 (g)、(h)) 三種方式進行討論。

結果觀察

整體而言，大部分傳統機器學習模型在各項填補方式下皆表現穩定，訓練與測試 R^2 分數普遍可達 0.6 至 0.9，顯示其對不同填補方法具有一定容錯能力。

相較之下，深度學習模型則對填補方式較為敏感，特別是在使用 KNN 或 MICE 填補時表現差異較大。以 BNN 為例，在填補 NaN 且使用手動調參時，測試集之 R^2 可達 0.782，預測表現優異，但改用 KNN 填補且搭配 Grid search 調參時，測試分數大幅下降至 0.127，顯示該模型對填補值準確性的依賴較高，亦受雜訊影響。類似狀況同樣出現在 ANN 模型中，在 MICE 填補搭配 Grid search 的組合下僅達到 0.175 的測試分數，整體表現明顯下降。

此外，KNN 填補對多數模型之影響整體偏向負面，訓練 R^2 分數普遍略低於其他填補方式，尤其在深度學習中更為明顯。如 ANN 模型於 KNN 填補與 Grid search 條件下，訓練分數僅 0.435，測試分數更只有 0.104，表示模型難以學習有用資訊，可視為學習失敗。

另外值得關注的是，使用 RBF 核的 SVM 模型在使用 MICE 填補時亦產生典型過擬合現象。如在 Grid search 調參下訓練分數高達 0.991，但測試分數僅為 0.352，此明顯差距顯示此模型可能對 MICE 填補後所建立之便數關聯性過度擬和，導致泛化能力下降。

doi:10.6342/NTU202502436

原因解析

1. 傳統機器學習模型具容錯能力，深度學習則對雜訊較為敏感

在資料預處理階段採用 KNN 或 MICE 進行缺失值填補，雖能補齊不完整資料，但所生成之填補值多為基於其他變數所推估之估算結果，未必能真實反映原始資料的分布特性，進而產生潛在的偏差。

傳統機器學習模型中的樹狀模型 (tree-based model，如 Random Forest、XGBoost、EBM) 主要根據特徵閾值進行樹的分裂，僅關注資料是否落於某個區間，而非其數值之精度，對誤差具一定容錯能力。即便填補策略產生些微偏差，模型仍可透過樣本間的區間分類概念，或藉由內部多樹結構之加權與平均機制，維持預測結果的穩定性。此外，這類模型本質上偏向分類架構，而非積極的追求曲線擬合，因此較不易受填補誤差所影響。

相較之下，深度學習模型如 ANN 與 BNN 屬於高度參數化模型，對輸入數據的分布變化敏感度高，當輸入包含不準確填補所產生的雜訊時，模型易受干擾而產生錯誤學習，導致無法有效收斂，產生訓練與測試表現失衡的情形。

2. 填入 NaN 或 0 為最保守處理，對資料整體結構干擾最小

將缺失值填以 NaN 或 0 屬較保守且簡化之處理方式，不會引入額外推估誤差。若整體缺失比例不高，且模型本身具有一定容錯能力，此類填補方式反而可避免錯誤估值對模型造成誤導，維持預測表現穩定性。

3. KNN 填補效果受資料結構限制與高維特徵空間影響

KNN 填補方法依據資料間之距離來推估缺失值，然而在本研究資料中，變數數量多且尺度不一，可能導致距離計算不穩定，進而影響填補準確度。此外，KNN 在高維空間中容易受到維度災難 (curse of dimensionality) 的影響，鄰近點判定失準，也可能導致填補值偏離實際分布，進而降低模型對測試資料的泛化能力。

4. MICE 填補保留統計結構，對簡單模型易引發過擬合

MICE 透過逐步迴歸方式多次推估缺失值，能有效保留變數間的統計關係，填補品質相對穩定，對於線性模型或樹狀模型而言為有效之處理方式，但其填補

doi:10.6342/NTU202502436

結果可能建立了變數間的隱藏關聯性，對於如 SVM (RBF 核) 這類結構簡單的模型而言，容易忽略掉這些關聯性，導致訓練表現極佳但測試表現不佳，形成典型過擬合現象。

4.3 視覺化分析

本節欲探討模型預測行為與解釋其內部邏輯，針對預測結果與特徵重要性進行視覺化分析，透過圖型輔助了解模型在不同資料條件下的預測準確度與學習關注的重點。

4.3.1 預測值與真值對照

為進一步了解模型預測結果與實際值之間的關係，本節採用散佈圖 (Scatter plot) 與殘差圖 (Residual plot) 進行討論。圖 4.12 與圖 4.13 分別為 Random Forest 與使用 RBF 核之 SVM 模型於訓練與測試階段的預測對照情形。圖中紅色虛線為理想預測線，點越集中於此線，代表預測結果越準確。

圖 4.12 顯示 Random Forest 模型在使用所有特徵、手動調參並採用 MICE 進行填補的條件下，於訓練與測試階段皆呈現良好的一致性。預測值大致集中於理想線附近，殘插圖中亦無明顯偏誤，表示模型具備穩定且良好的泛化能力。

相對的，圖 4.13 中的 SVM_rbf 模型則呈現典型的過擬合現象。儘管訓練階段的預測幾乎完美貼合理想線、殘差極小，但在測試階段卻出現明顯偏差與離散，顯示模型過度學習訓練資料中的細節與雜訊，導致泛化能力不足，無法有效了解新資料。

doi:10.6342/NTU202502436

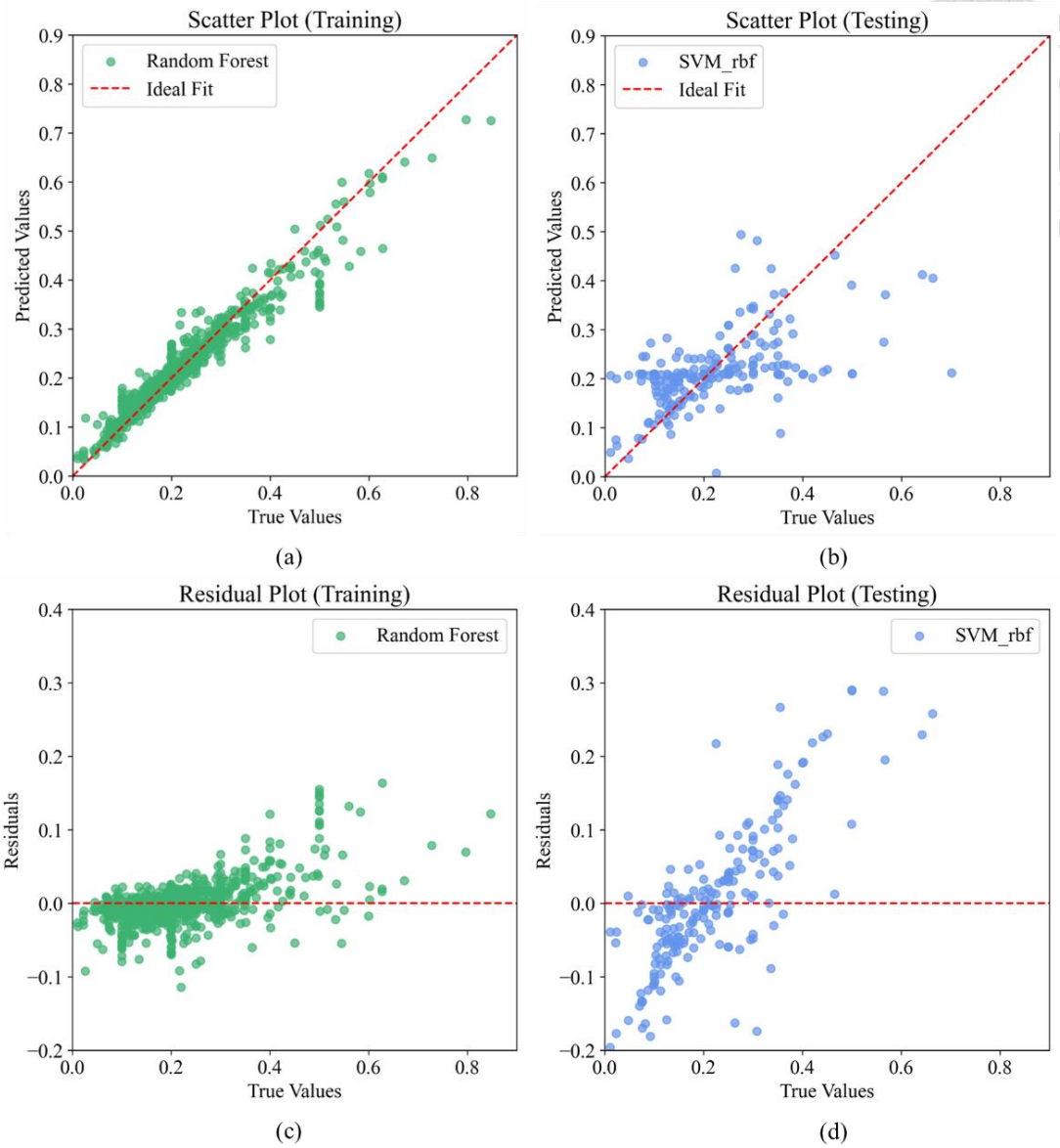


圖 4.12 Random Forest (包含所有特徵，手動調參，MICE 填補) 模型散佈與殘差於訓練與測試集之對照圖

doi:10.6342/NTU202502436

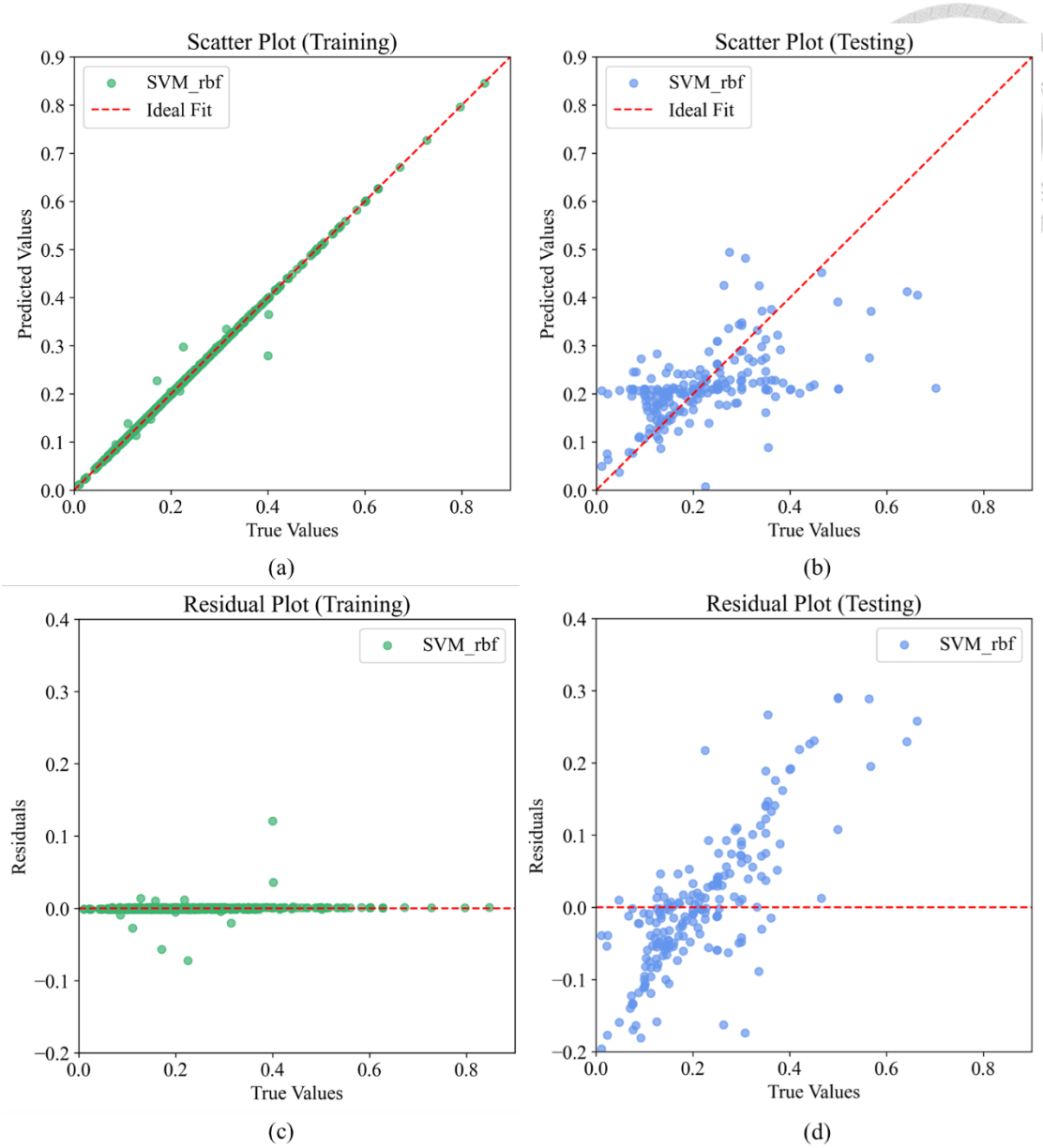


圖 4.13 使用 RBF 核 SVM (包含所有特徵，手動調參，MICE 填補) 模型散佈與殘差於訓練與測試集之對照圖

此外，由圖 4.8 至圖 4.11 所對應的 (g) 小圖可見 MAE、MSE 與 RMSE 誤差指標與圖 4.8(g) 所示之 R^2 分數具有高度負相關，顯示視覺化分析能有效輔助定量指標之判讀。

doi:10.6342/NTU202502436

4.3.2 特徵重要性

本節將說明模型在進行預測時所依賴之資料特徵分佈情形。所謂特徵重要性 (Feature importance) 是指各輸入變數對模型預測結果的影響程度，可協助判斷模型在學習過程中較為重要的訊息來源，同時具備可解釋性與應用價值。由於不同模型內部機制不一，其評估特徵重要性的方式亦有所差異，依本研究所使用之模型類型，大致可分為以下兩類:

樹狀模型 (Tree-based model)

包含 Random Forest、XGBoost、CatBoost 和 EBM 等樹狀模型，通常內建以節點分裂統計為基礎的特徵重要性計算機制，常見作法包含分裂次數法 (Frequency-based) 以及損失減少加總法 (Gain-based)。分裂次數法為統計各特徵在所有決策樹中作為分裂節點的出現次數，次數越多代表該特徵對迴歸或分類決策中越常被引用，重要性越高；損失減少加總法則是計算每個特徵在分裂節上所貢獻的損失函數下降量 (如 MSE)，並將其加總，作為該特徵的整體貢獻指標。以上兩種計算機制的重要性分數通常以相對比例表示（加總為 1），數值越高表示該特徵越常被用來進行有效分裂，對預測結果影響越大。

以 Random Forest 模型為例，在只有使用主要特徵模型中，如圖 4.14，分析出來的重要特徵前五名依序為 e_0、N、f_c、Liquefaction criteria 和 p'_0；在包含所有特徵的模型中，如圖 4 15，分析出來的重要特徵前五名會因而外特徵的加入而有所變動，為 e_0、N、e_{max}、p'_0和 D_{50}。

doi:10.6342/NTU202502436

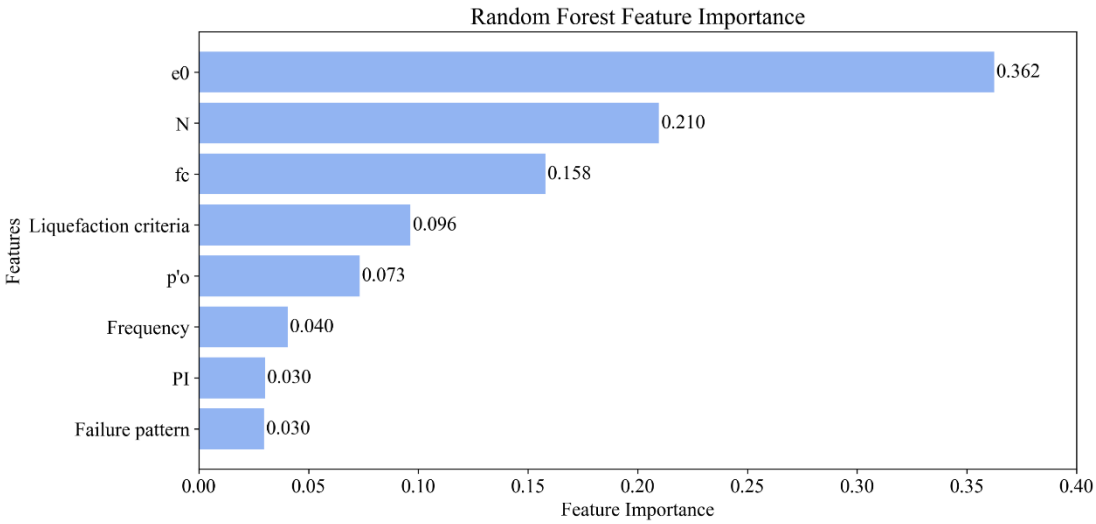


圖 4.14 Random Forest (包含主要特徵，手動調參) 之特徵重要性圖

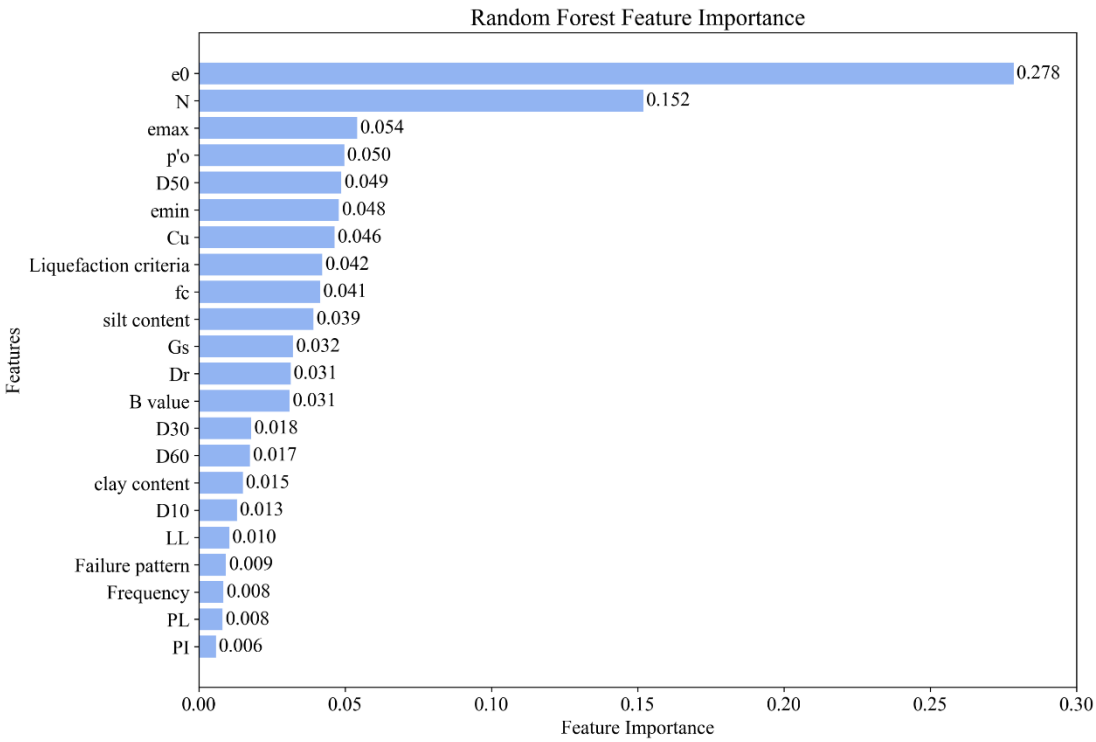


圖 4 15 Random Forest (包含所有特徵，手動調參，NaN 填補) 之特徵重要性圖

黑箱模型 (Black box model)

包含 SVM、ANN 與 BNN 等模型，由於本身不具內建的特徵重要性推估機制，通常使用置換重要性 (Permutation importance) 作為後處理方法進行特徵解釋。其評估流程如下：

doi:10.6342/NTU202502436

1. 基準計算：以原始資料進行預測，並記錄模型在測試資料上的評估指標 (如 R^2、MSE)。
2. 隨機打亂單一特徵：對指定特徵欄位的數值進行隨機置換，使其與目標變數的關聯性消失，其他特徵保持不變。
3. 再次預測：以打亂後的資料再次執行模型預測，並重新計算評估指標。
4. 比較預測效能下降程度：若打亂後模型表現明顯變差，表示該特徵在原始模型中具有較高的重要性。

以不同特徵組合作為輸入，能顯著影響模型結果的特徵排名會有所改變。以 SVM 模型為例，於只使用主要特徵之模型中，如圖 4.16，特徵重要性前五名依序為 f_c、e_0、Frequency、N、p'_0 和 Liquefaction criteria；在包含所有特徵的模型中，如圖 4.17，前五名依序為 e_0、N、silt content、f_c 和 Back pressure。

特徵重要性取決於模型結構、參數設定與輸入資料等多項因素，因此不具結果穩定性。即使為相同模型，當特徵組合或超參數設定不同時，特徵重要性排序能可能產生變動。此特性在使用置換重要性 (Permutation importance) 作為估算方式的黑箱模型中更為明顯，由於此法相當依賴資料在模型中的預測貢獻，當輸入的特徵組合不同時，模型對資訊來源的依賴順序也將隨之重新構建，進而影響特徵重要性分數與排序結果。

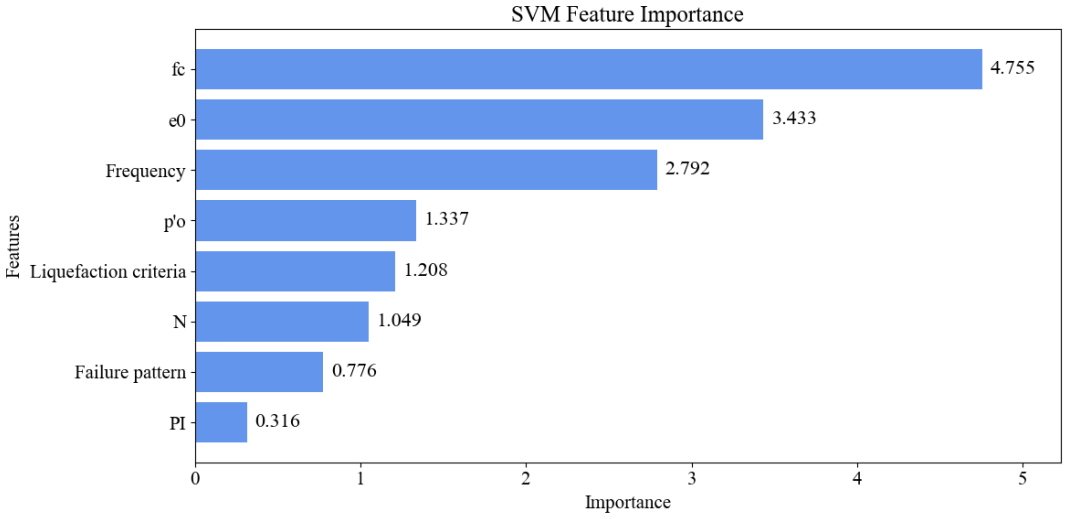


圖 4.16 使用 RBF 核之 SVM (包含主要特徵，手動調參) 之特徵重要性圖

doi:10.6342/NTU202502436

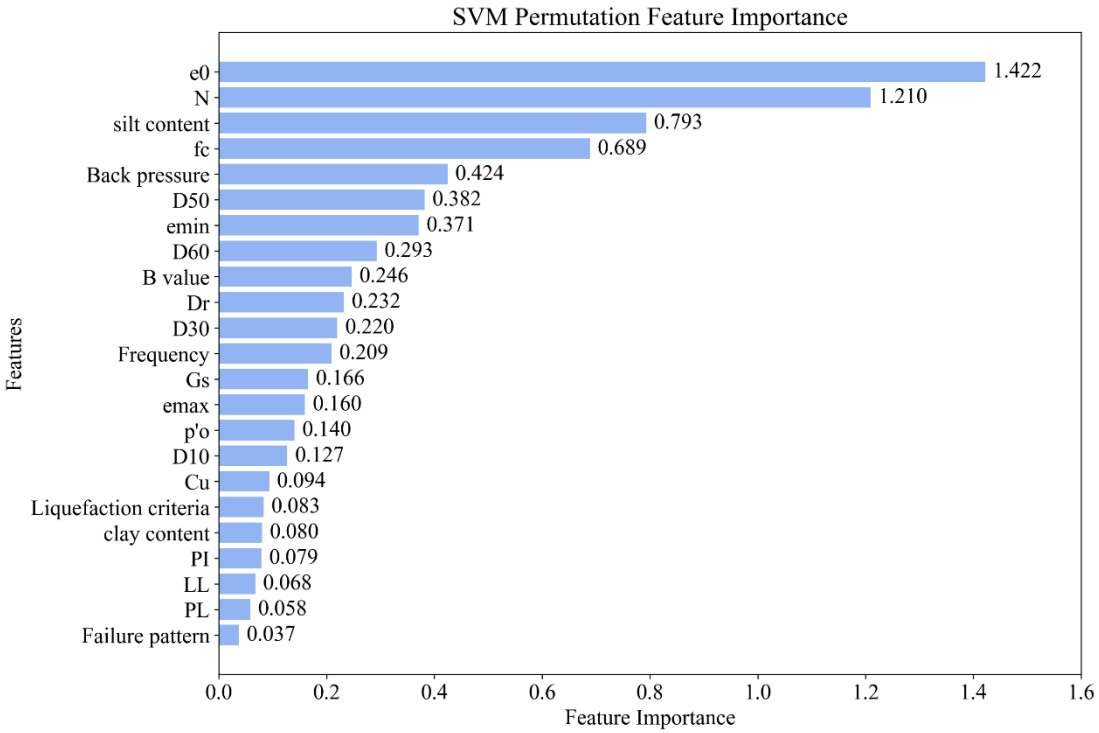


圖 4.17 SVM (包含主要特徵，手動調參，以 0 填補) 之特徵重要性圖

此外，EBM 模型所繪製之特徵重要性圖亦可呈現特徵交互作用之重要性排序，使其具備更高的可解釋性。如圖 4.18，當經 One-Hoe Encoding 處理後，原本為 object 類型的特徵將被展開為多個欄位，再加上模型在學習過程中所產生的交互作用項，整體特徵重要性之數量及排序變得相對複雜。若將特徵交互作用去除，如圖 4.19，則會得到較為簡潔單一的特徵重要性排序。進一步則可將因 One-Hot Encoding 拆分出去的特徵重要性做相加，便可得到與其他模型相近與項目較為整潔的特徵重要性圖，如圖 4.20 所示。

doi:10.6342/NTU202502436

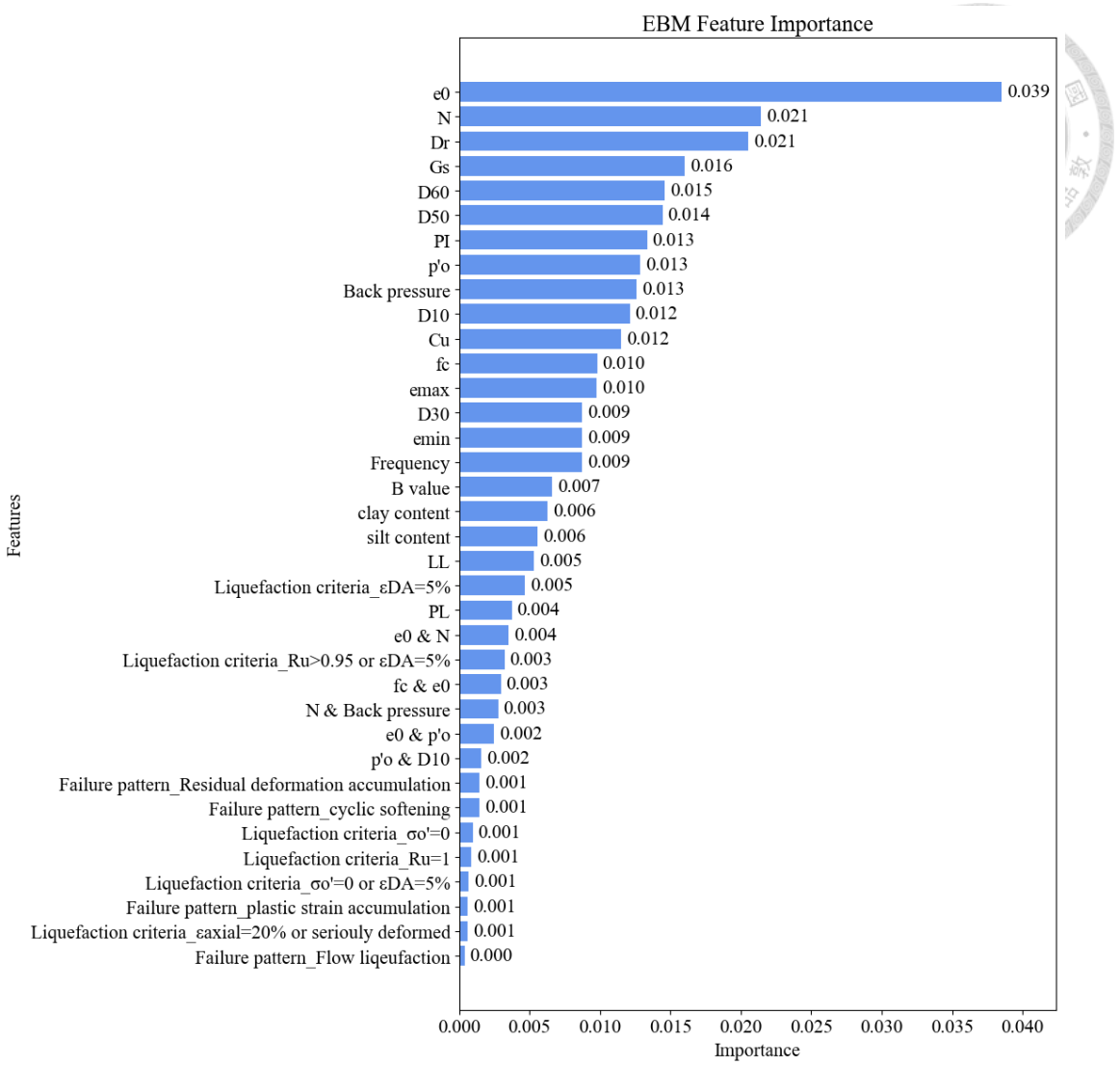


圖 4.18 EBM 含特徵交互作用之特徵重要性圖

doi:10.6342/NTU202502436

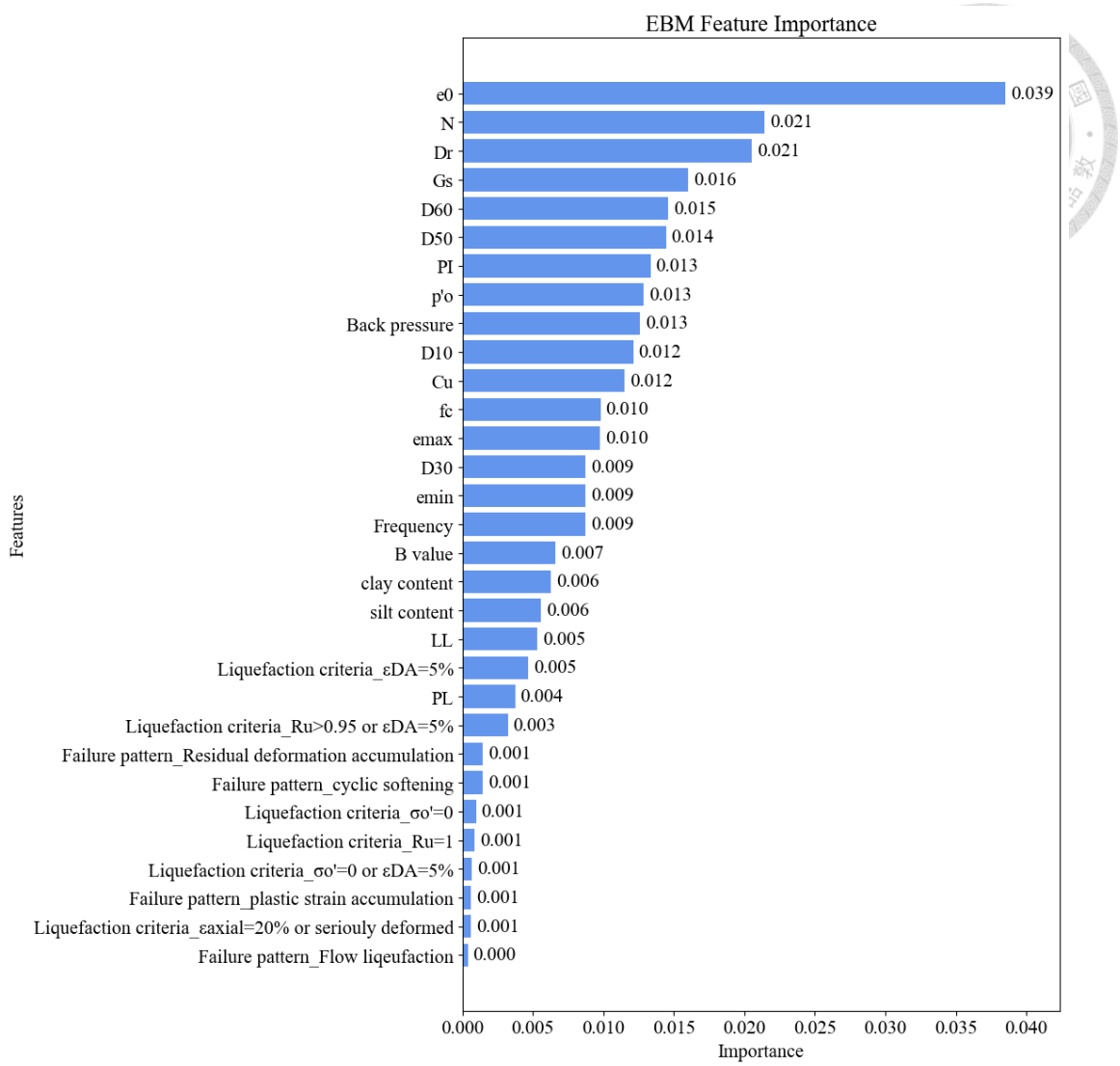


圖 4.19 EBM 不含特徵交互作用之特徵重要性圖

doi:10.6342/NTU202502436

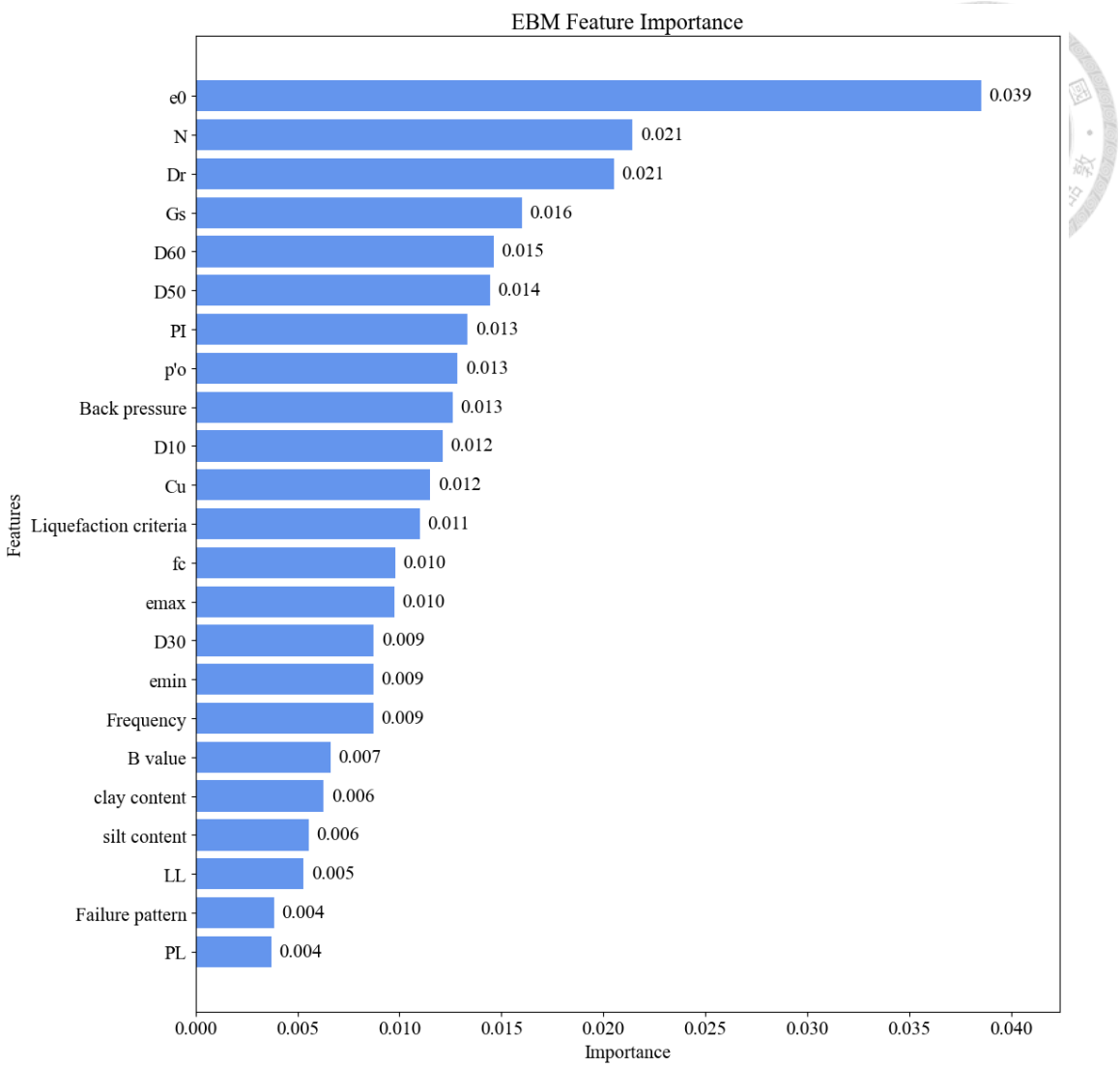


圖 4.20 EBM 不含特徵交互作用且將 dummy variable 合併之特徵重要性圖

所有模型的重要性排序結果於表 4.2，依照輸入特徵組合（只有 Primary feature 之模型與包含 All features 之模型）的不同去做統計，本研究將特徵歸納為以下三類，分別為核心特徵、次核心特徵與次級特徵。

doi:10.6342/NTU202502436

表 4.2 所有模型於不同參數組合及設定下之前五名特徵重要性整理表

Model	Hyperparameters	Primary features	All features		
			NaN / "0"	KNN	MICE
Random Forest	Manual tuning	e_0	e_0	e_0	e_0
		N	N	C_u	N
		f_c	e_{max}	N	PI
		Liquefaction criteria	p'_0	D_{10}	C_u
		p'_0	D_{50}	B value	B value
	Grid search tuning	e_0	e_0	e_0	e_0
		N	N	C_u	N
		f_c	C_u	N	PI
		Liquefaction criteria	B value	D_{10}	C_u
		p'_0	e_{min}	PI	B value
CatBoost	Manual tuning	e_0	e_0	e_0	D_r
		f_c	N	N	e_0
		N	Liquefaction criteria	D_{10}	N
		Liquefaction criteria	f_c	Liquefaction criteria	Back pressure
		Failure pattern	Failure pattern	C_u	D_{50}
	Grid search tuning	e_0	e_0	e_0	D_r
		f_c	N	N	e_0
		N	e_{min}	D_{10}	N
		Liquefaction criteria	f_c	Liquefaction criteria	Liquefaction criteria
		Failure pattern	Liquefaction criteria	C_u	Back pressure

doi:10.6342/NTU202502436

Model	Hyperparameters	Primary features	All features		
			NaN / "0"	KNN	MICE
XGBoost (gain)	Manual tuning	Liquefaction criteria	Liquefaction criteria	Liquefaction criteria	Liquefaction criteria
		Frequency	B value	C_u	B value
		e_0	C_u	e_0	e_0
		f_c	e_0	D_{10}	f_c
		N	D_{60}	B value	D_{50}
	Grid search tuning	Liquefaction criteria	Liquefaction criteria	Liquefaction criteria	Liquefaction criteria
		e_0	C_u	D_{10}	e_0
		f_c	B value	e_0	B value
		Frequency	e_0	C_u	e_{max}
		N	D_{10}	f_c	C_u
EBM (without interaction)	Manual tuning	e_0	e_0	e_0	e_0
		f_c	N	N	D_r
		N	D_r	D_r	N
		Frequency	e_{min}	e_{min}	G_s
		Liquefaction criteria	PI	Liquefaction criteria	D_{50}
	Grid search tuning	e_0	e_0	e_0	e_0
		f_c	N	N	D_r
		N	e_{min}	e_{min}	N
		Frequency	f_c	Liquefaction criteria	G_s
		p'_0	e_{max}	D_r	D_{50}

doi:10.6342/NTU202502436

Model	Hyperparameters	Primary features	All features		
			NaN / "0"	KNN	MICE
SVM_rbf	Manual tuning	f_c	D_{60}	silt content	D_r
		e_0	N	f_c	Back pressure
		Frequency	B value	Back pressure	p'_0
		p'_0	e_{min}	LL	silt content
		Liquefaction criteria	silt content	N	f_c
	Grid search tuning	f_c	e_0	D_r	D_r
		e_0	N	Back pressure	Back pressure
		Frequency	silt content	C_u	p'_0
		p'_0	f_c	p'_0	silt content
		N	Back pressure	f_c	f_c
SVM_poly	Manual tuning	Frequency	e_{min}	silt content	LL
		Liquefaction criteria	D_{60}	f_c	PL
		f_c	G_s	Back pressure	C_u
		e_0	N	D_r	Back pressure
		N	B value	LL	PI
	Grid search tuning	N	e_0	D_r	Back pressure
		D_r	N	Back pressure	LL
		Back pressure	Frequency	C_u	silt content
		p'_0	D_r	p'_0	PL
		B value	D_{60}	f_c	f_c

doi:10.6342/NTU202502436

Model	Hyperparameters	Primary features	All features		
			NaN / "0"	KNN	MICE
ANN	Manual tuning	e_0	f_c	silt content	D_r
		f_c	p'_0	f_c	Back pressure
		N	silt content	e_0	silt content
		p'_0	Back pressure	PL	p'_0
		Liquefaction criteria	C_u	C_u	N
	Grid search tuning	e_0	Back pressure	C_u	f_c
		f_c	f_c	silt content	Back pressure
		N	D_r	Back pressure	N
		Liquefaction criteria	p'_0	p'_0	p'_0
		p'_0	silt content	D_r	D_r
BNN	Manual tuning	Liquefaction criteria	Liquefaction criteria	Failure pattern	Failure pattern
		Failure pattern	Failure pattern	Liquefaction criteria	Liquefaction criteria
		N	N	e_0	e_0
		f_c	e_0	D_{50}	D_{50}
		e_0	p'_0	Frequency	Frequency
	Grid search tuning	Liquefaction criteria	Liquefaction criteria	Failure pattern	Failure pattern
		Failure pattern	Failure pattern	Liquefaction criteria	Liquefaction criteria
		N	N	e_0	e_0
		f_c	e_0	D_{50}	D_{50}
		e_0	C_u	N	D_{10}

doi:10.6342/NTU202502436

核心特徵 (N、e_0、f_c、Liquefaction criteria)

包含初始孔隙比 (e_0)、循環剪切次數 (N)、細粒料含量 (f_c) 與液化判斷標準 (Liquefaction criteria) 四項特徵，於所有模型與不同特徵組合中皆穩定出現在前五名，顯示其在本研究中具有高度的預測影響力與穩定性。

- 初始孔隙比 (e_0)

於不同特徵組合中總共分別出現 15 與 28 次，且於樹模型中大多位居第一，即使在加入次要特徵 (Secondary features) 後仍榜上有名，其與土壤結構鬆緊程度有關，為影響液化行為的關鍵參數，不僅在基本模型中扮演重要角色，也能在高維度特徵模型中維持解釋力。

- 循環剪切次數 (N)

於僅包含主要特徵與包含所有特徵之模型中，分別出現 15 與 28 次，且在不受填補值方法與特徵擴充影響，仍保有模型對其的依賴性，顯示其為模型預測時的重要指標。

- 細粒料含量 (f_c)

於不同特徵組合中分別出現 15 與 17 次，與土壤之塑性與滲透性有關。然而，於包含所有特徵的模型中，出現次數明顯降低，顯示可能有其他更具影響力的特徵取代其預測貢獻。

- 液化判斷標準 (Liquefaction criteria)

於所有模型中共出現 33 次，作為核心特徵的其中之一，亦常被模型視為重點學習指標。同樣於包含所有特徵的模型中，重要性排名出現次數明顯降低，表示可能受到其他特徵的影響而降低其預測貢獻。

doi:10.6342/NTU202502436

次核心特徵 (C_u、D_r、p'_0、Back pressure)

包含平均粒徑 (C_u)、相對密度 (D_r)、試驗圍壓 (p'_0) 與反水壓 (Back pressure) 三項特徵。主要在包含所有特徵之模型中被列入排名，顯示其在模型預測中亦為重要指標。

- 平均粒徑 (C_u)

  於包含所有特徵之模型中，為出現次數最高的非核心變數，共 17 次。

- 相對密度 (D_r)

  亦出現 17 次，與孔隙比與顆粒排列鬆緊相關，在高維特徵資料下表現出其潛在重要性。

- 試驗圍壓 (p'_0)

  同樣出現 17 次，為試驗中模擬試體於不同深度的試驗參數，即使在加入次要特徵後仍榜上有名，不失為預測模型中的重要指標。

- 反水壓 (Back pressure)

  為試驗中控制試體有效應力的試驗參數，可視為最關鍵的試驗控制變數。

次級特徵 (Failure pattern、Frequency)

包含破壞模式 (Failure pattern) 與剪切頻率 (Frequency) 兩項特徵。這些變數於包含與不包含次要特徵的模型中皆有出現，但總體排序位置較後，重要性較不穩定。

doi:10.6342/NTU202502436

- 破壞模式 (Failure pattern)

總出現次數僅 11 次，屬資料標註或試驗後觀察之變數類別，在部分模型中提供額外解釋力。

- 剪切頻率 (Frequency)

於所有模型中只出現 11 次，於動態剪切試驗中屬加載控制變因，僅於部分模型中具顯著性。

此結果呼應動態三軸試驗設計中參數間的關聯性與模型泛化能力的重要性，資料品質、填補缺失值方式與特徵工程將會顯著影響模型的學習結果與對特徵的依賴性。本節研究結果顯示，若欲建立動態三軸試驗的參數預測模型，可優先考慮輸入參數如初始孔隙比(e_0)、循環剪切次數 (N)、細粒料含量 (f_c) 與液化判斷標準 (Liquefaction criteria) 等核心特徵，並依資料取得條件加入其他具潛在影響力的變數如平均粒徑 (C_u)、相對密度 (D_r)、試驗圍壓 (p'_0)與反水壓 (Back pressure) 等，以提升模型準確性與穩定性。值得注意的是，部分特徵彼此之間存在相關性或交互作用 (Interaction effects)，例如初始孔隙比 (e_0) 與相對密度 (D_r) 常在土壤結構與密實程度的判定中同時使用，這可能造成模型特徵重要性評估結果出現偏差。

進一步觀察特徵缺失率與重要性之間的關係。部分模型 (如樹模型) 中，為試圖了解缺失比例較高的特徵，而賦予較高的權重，這可能增加特徵的重要性結果。然而，在本研究結果中此影響相對有限，即便核心特徵缺失率為零，在多數模型中仍具備穩定且顯著的重要性貢獻，如圖 4.21 所示。因此，本研究認為缺失值對特徵重要性排序之影響不大，上述所建立之特徵分層 (核心、次核心與次級特徵) 結果具備一定的可信度與解釋力。

doi:10.6342/NTU202502436

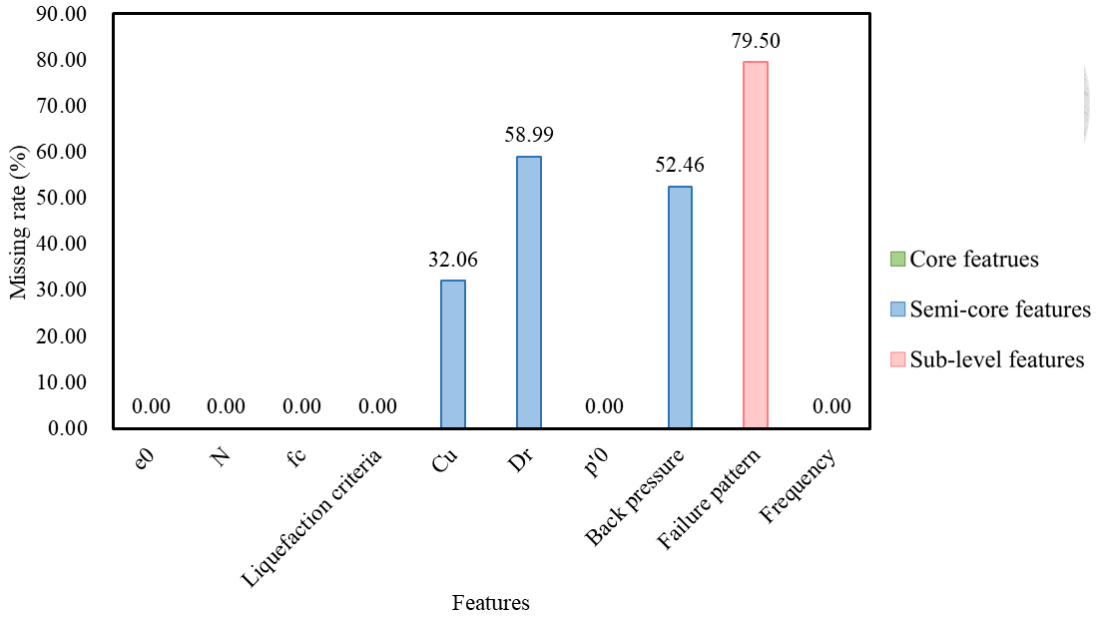


圖 4.21 不同特徵重要性階層之缺失率長條圖

此外，為了瞭解特徵與目標特徵之間的關係，本研究亦針對分數較高之手動調參組合，採用未進行缺失值填補之設定，分別以「僅包含主要特徵」與「包含所有特徵」兩種特徵組合，繪製 SHAP 總結圖（Shapley Additive explanation summary plot, SHAP），並整理於附錄 C。SHAP 圖可進一步顯示各特徵對模型預測結果之相對影響程度與方向性，圖示之閱讀方式於第五章有詳細說明，亦可與表 4.2 所列各模型之特徵重要性排序進行對照，作為模型解釋性與特徵穩定性之補充分析。

4.3.3 模型訓練時間比較

在實務應用中，除了模型的預測準確度外，訓練所需時建議是模型選擇時的重要考量。因此，透過比較不同模型在各特徵選擇和補值方式下的訓練時間，可

doi:10.6342/NTU202502436

以幫助了解模型的可行性與運算效率，進而做出兼顧成果與時間成本的模型篩選。圖 4.22 為各模型的訓練時間比較圖，可以由下列三個面向做討論：

1. 調參策略影響

大部分的模型在 Grid search 調參方式 (圖 4.22 (b)、(d)、(f)、(h)) 時，其訓練時間明顯高於手動調參 (圖 4.22 (a)、(c)、(e)、(g))。特別是 BNN 模型僅使用主要特徵時，在 Grid search 下訓練時間超過 10000 秒，顯示其計算成本極高。而在包含所有特徵的模型中，傳統機器學習模型方面，CatBoost 在手動調參下 (圖 4.22 (c)、(e)、(g)) 耗時較久；EBM 在 Grid search 調參下 (圖 4.22 (d)、(f)、(h)) 的訓練時間則顯著上升。

2. 填補策略影響

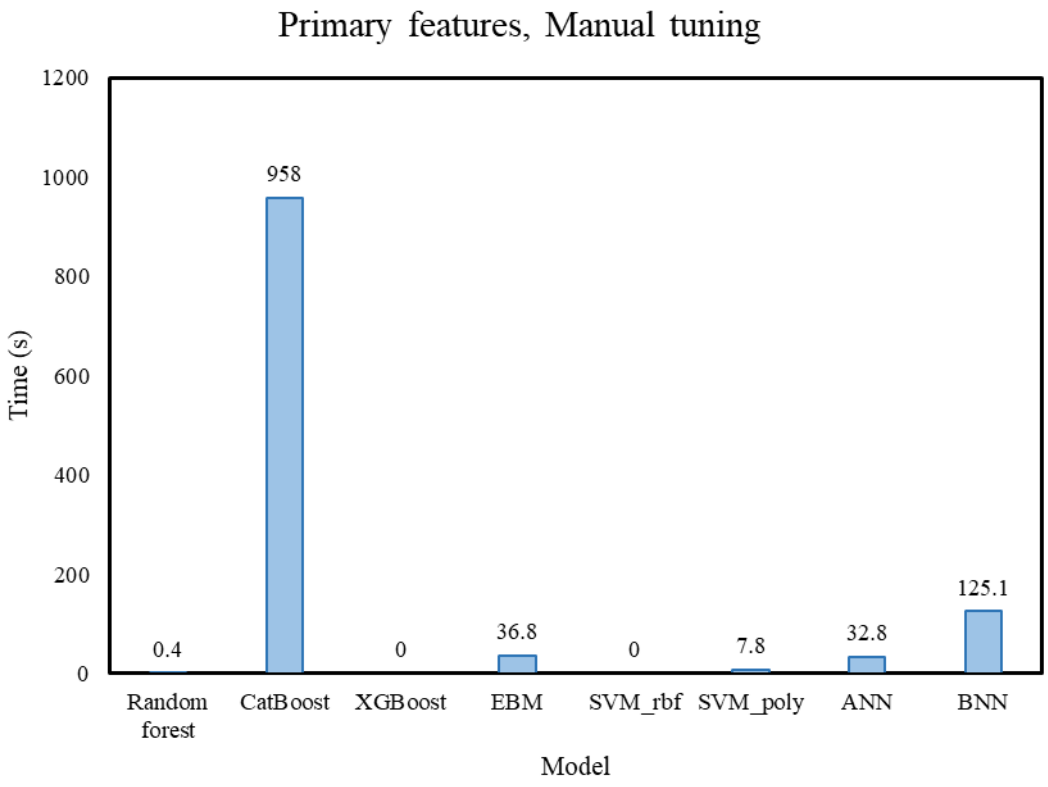
在手動調參的情況下，若採用 NaN 或 0 作為缺失值填補策略 (圖 4.22 (c))，CatBoost 的訓練時間明顯高於其他模型；而改用 KNN 或 MICE 補值 (圖 4.22 (e)、(g)) 後雖時間有所下降，但整體仍相對較長。除 BNN 外，其餘模型在不同補值策略下訓練時間變化相對穩定 BNN 則在採用 KNN 或 MICE 補值時，其訓練時間有明顯增加。

3. 模型個別觀察

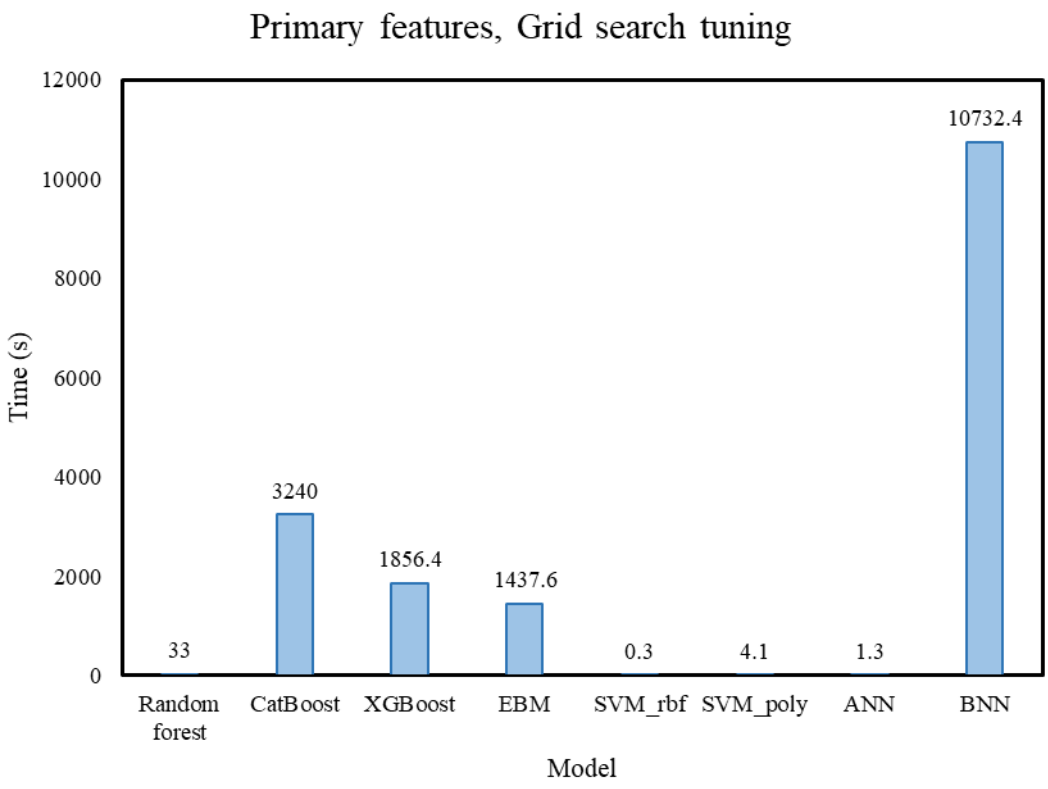
除上述幾個在訓練時間上變化較大的模型外，其餘模型如 Random Forest、XGBoost、SVM (RBF 與 polynomial 核)、以及 ANN 模型，其訓練時間在各情境下均相對穩定。

透過比較不同模型在相同資料與設定下的訓練時間，可以初步判斷其計算資源消耗與潛在複雜度。例如 Random Forest 和 SVM 等結構相對簡單的模型，其訓練時間多能在秒或數分鐘內完成；而 CatBoost、EBM 或 BNN 等複雜模型，在特定條件下則需顯著較長的時間。特別是本研究所設定的 BNN 模型，是在所有模型中，在所有模型中平均訓練時間最長，顯示其模型規模與計算負擔相對較大。

doi:10.6342/NTU202502436

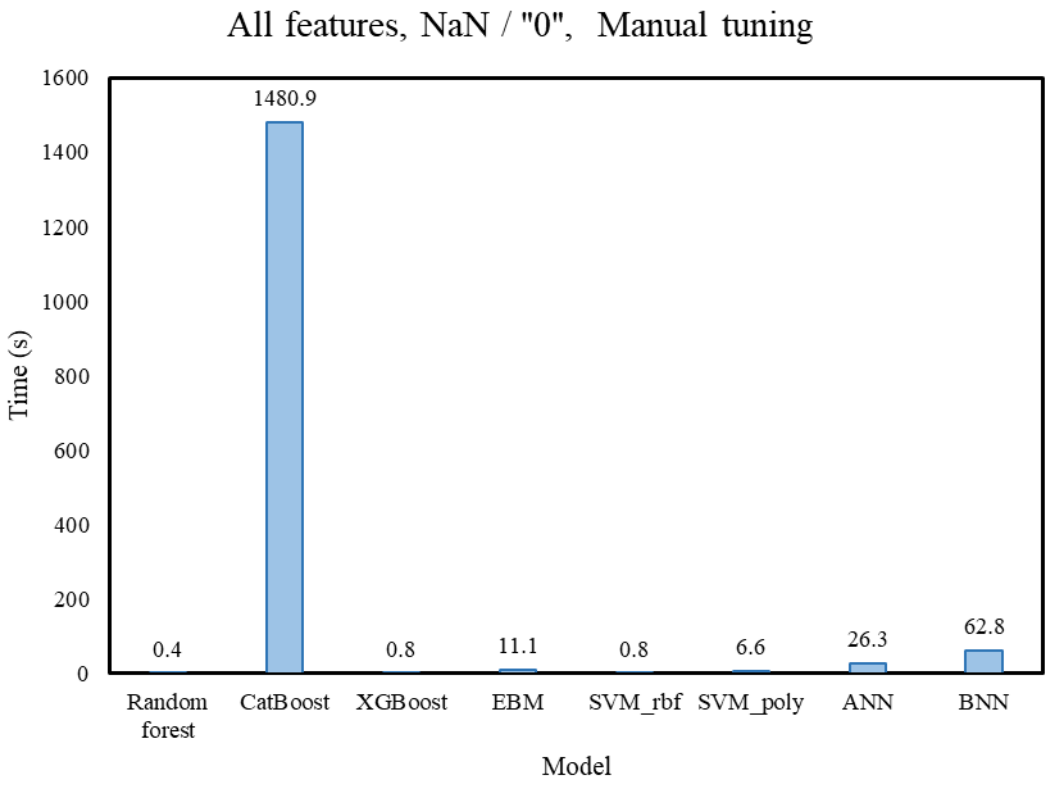


(a)

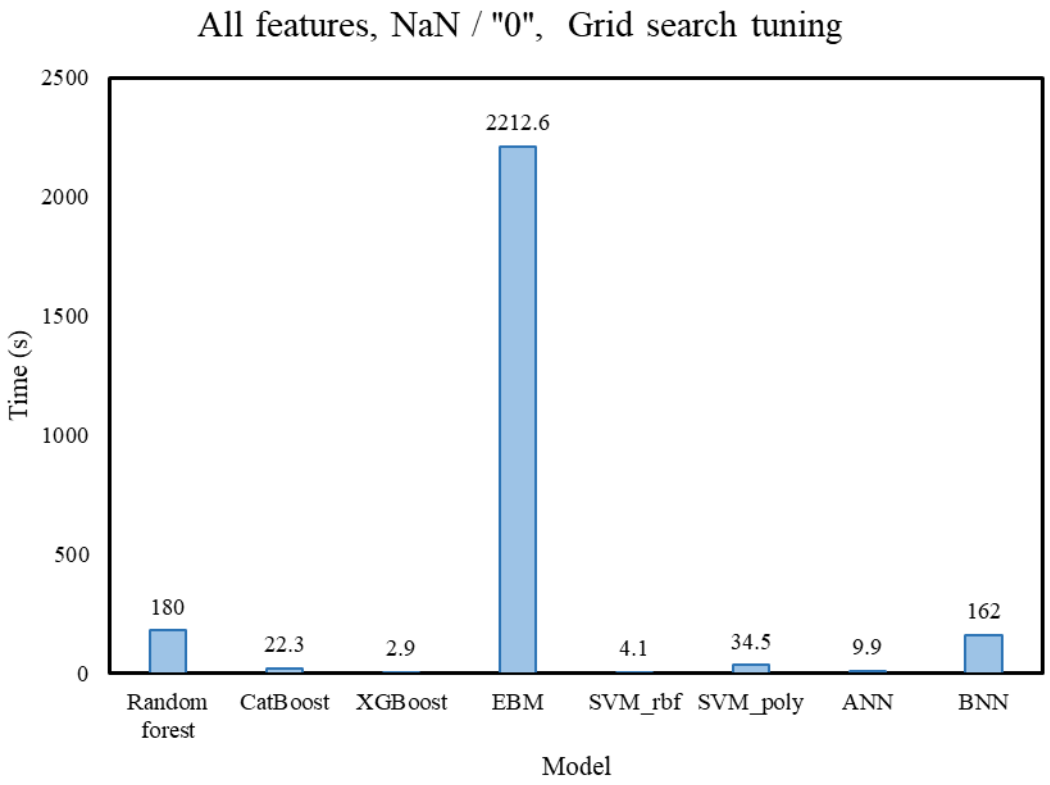


(b)

doi:10.6342/NTU202502436

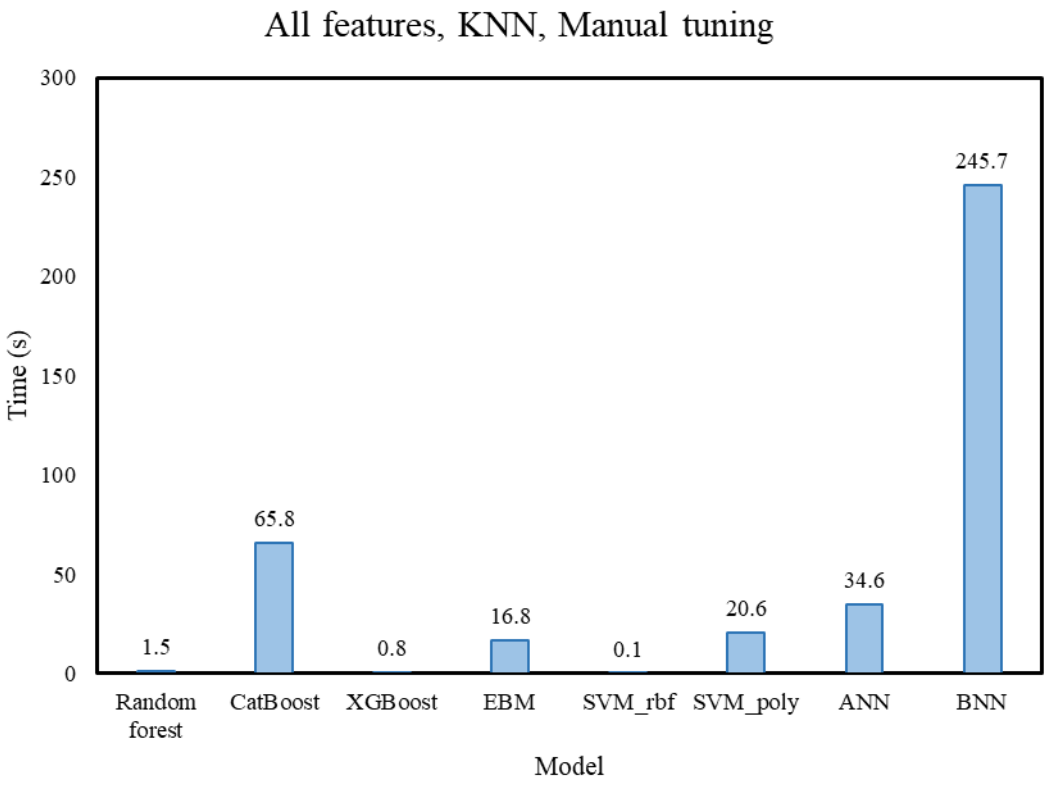


(c)

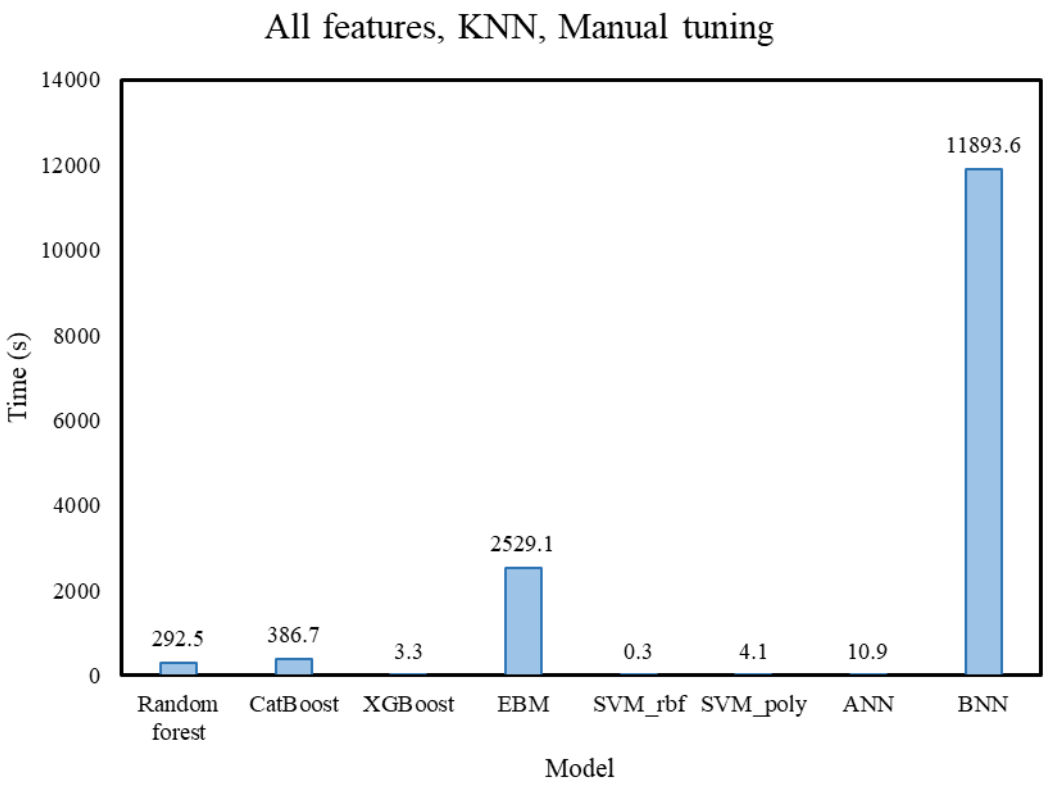


(d)

doi:10.6342/NTU202502436

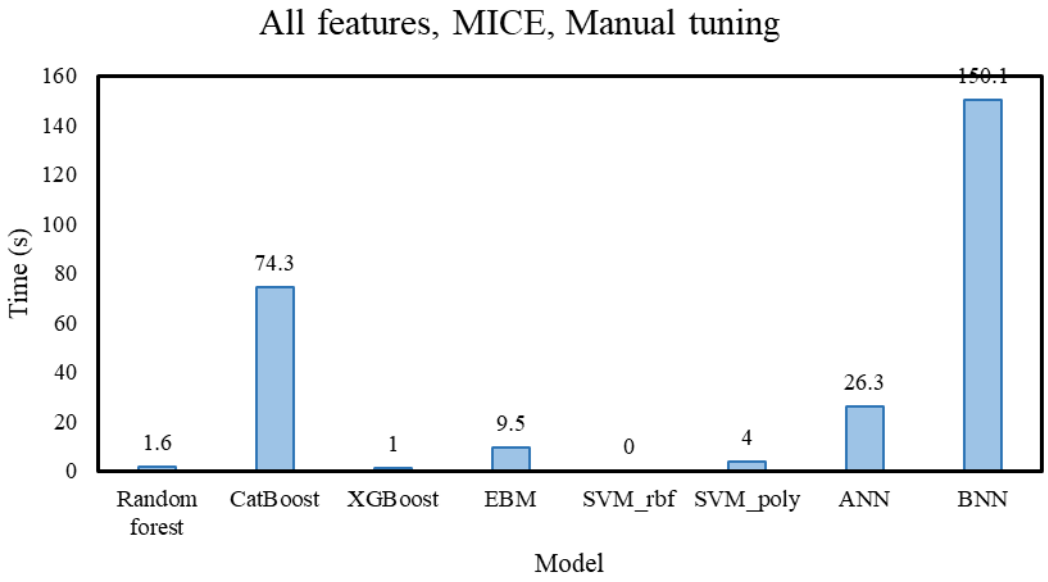


(e)

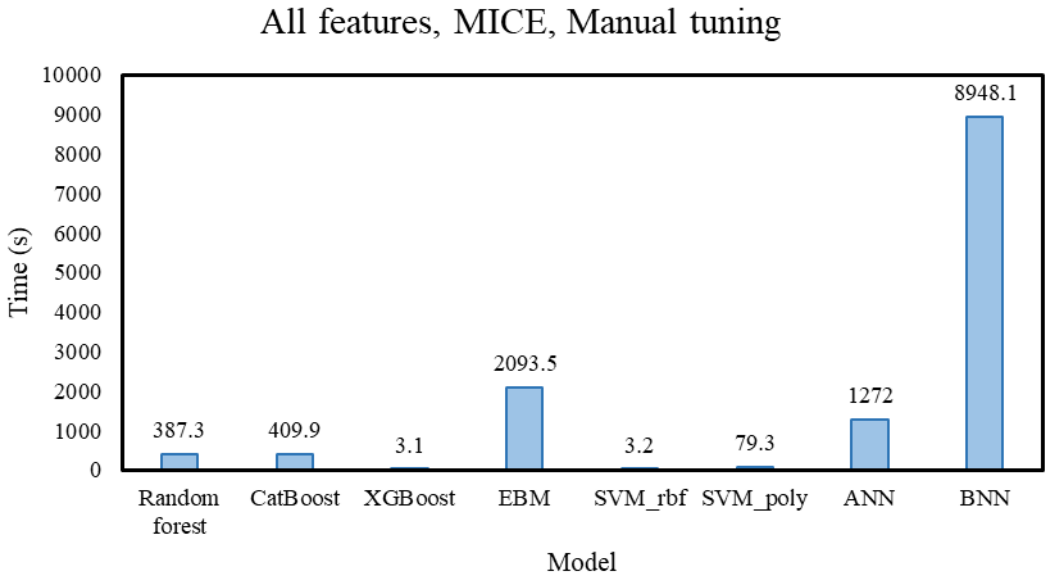


(f)

doi:10.6342/NTU202502436



(g)



(h)

圖 4.22 不同特徵工程下各類模型之訓練時間比較圖 (a) 包含主要特徵與手動調參；(b) 包含主要特徵與 Grid search 調參；(c) 包含所有特徵、以 NaN 或 0 填補與手動調參；(d) 包含所有特徵、以 NaN 或 0 填補與 Grid search 調參；(e) 包含所有特徵、以 KNN 填補與手動調參；(f) 包含所有特徵、以 KNN 填補與 Grid search 調參；(g) 包含所有特徵、以 MICE 填補與手動調參；(h) 包含所有特徵、以 MICE 填補與 Grid search 調參

doi:10.6342/NTU202502436

第五章 EBM 與 BNN 於液化數據之應用

為提升本研究模型之應用廣度，第五章延續第四章對多種機器學習模型進行之迴歸預測分析，額外針對分類問題進行探討。雖然 CSR 預測本身已具應用價值，但若能針對「是否發生液化」進行二元分類，可讓模型應用更直觀，可提供工程實務上常見的判斷需求，因此本章新增分類分析，提供另一種角度進行模型評估。於傳統模型方面選用可解釋增強模型 (Explainable Boosting Machine, EBM)；在深度學習方面則選擇貝葉斯神經網路 (Bayesian Neural Network, BNN)。本章將分別針對兩者進行迴歸分析、分類分析、解釋性探討與不確定性評估之說明與比較，以提升模型在實務中的應用性與資料的利用價值。

5.1 分類模型數據說明

在分類模型的設計上，首先考慮於 2.1.3 節中提到的抗液化曲線觀念，並基於所蒐集資料皆為實際發生液化之動態三軸試驗結果，藉由 CSR 與 N 之間的反比關係建立分類依據。圖 5.1 為本研究所建立之分類標準示意圖，橫軸為取對數之循環剪切次數 (N)，縱軸為循環剪應力比 (CSR)，兩者為動態三軸試驗中常見且高度相關之參數。

為界定分類標準本研究以試誤法擬合出一條能包覆所有資料點之下限界線 (Lower Bound, LB，紅色曲線)，作為是否會發生液化的判斷基準。該下限線之公式如式(5.1):

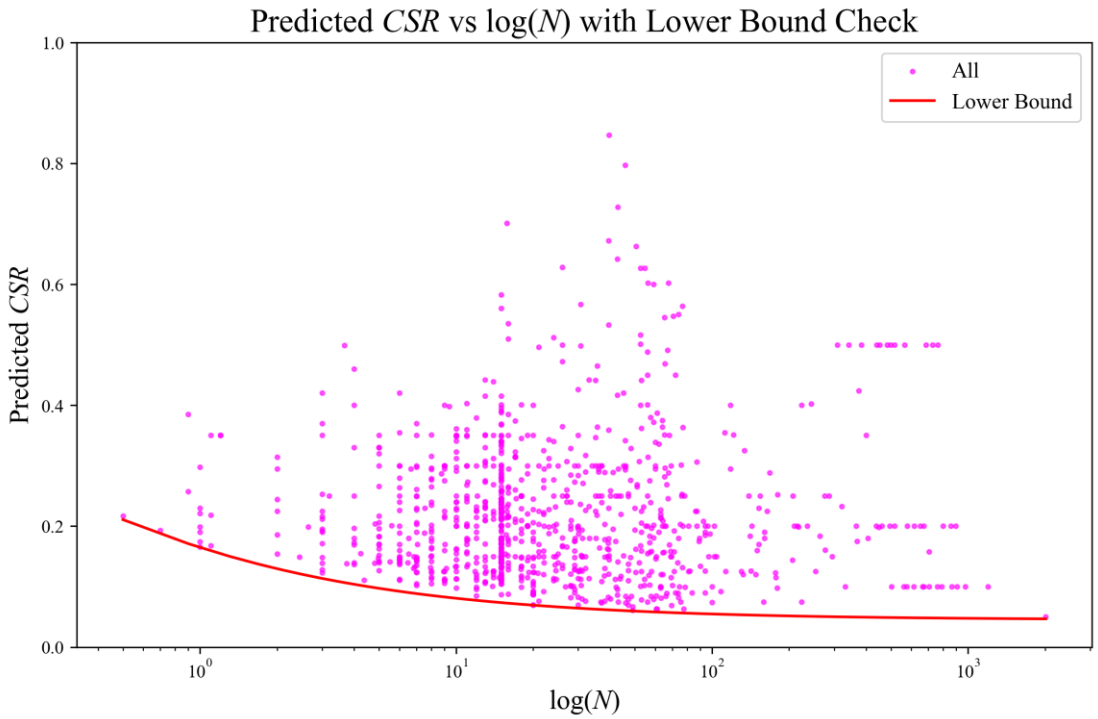
$$CSR = \frac{1}{1.35 + 2N^{0.58}} + 0.045 \tag{5.1}$$

若模型預測之 CSR 值低於此下限曲線，即判斷為「不會液化」；反之，若落於曲線以上，則判斷為「會液化」，以此方式進行分類標籤之設定。

doi:10.6342/NTU202502436

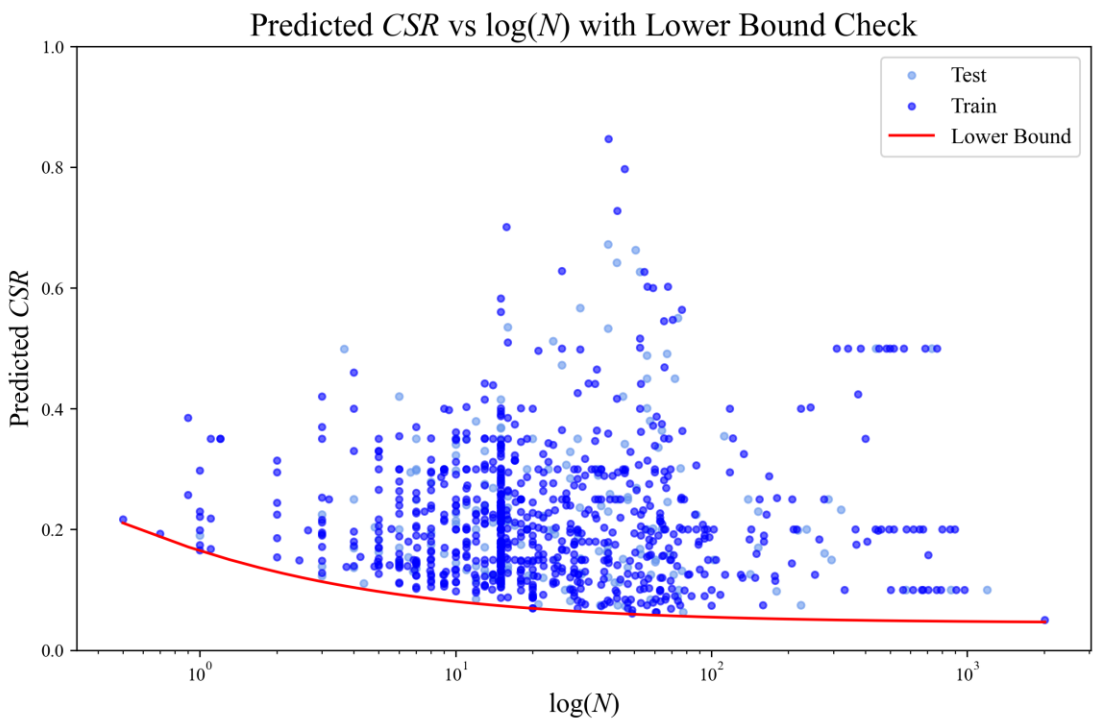
在實作上，本研究將原本迴歸模型所預測之 CSR 結果，根據其位置相對於下界線進行標籤劃分，轉換為二元分類問題。此轉換方式兼具模型連續性與分類的特性，可用於實務上篩選潛在液化樣本，亦有助於了解模型對不同輸入條件的分類反應。

圖 5.1 (a) 中為本研究所蒐集並用來做預測分析之樣本，且均於試驗中實際發生液化 (All, 桃紅色)；圖 5.1 (b) 中深藍色與淺藍色分別代表訓練資料集 (Train) 與測試資料集 (Test)；圖 5.1 (c) 中綠色與橘色點則代表測試資料中模型預測結果與分類判斷之比對結果，綠色點為預測正確；橘色點則表示預測錯誤。此分類方式可將迴歸模型的輸出結果轉變為二元分類結果，能進一步的評估模型在液化辨識上的實用性。

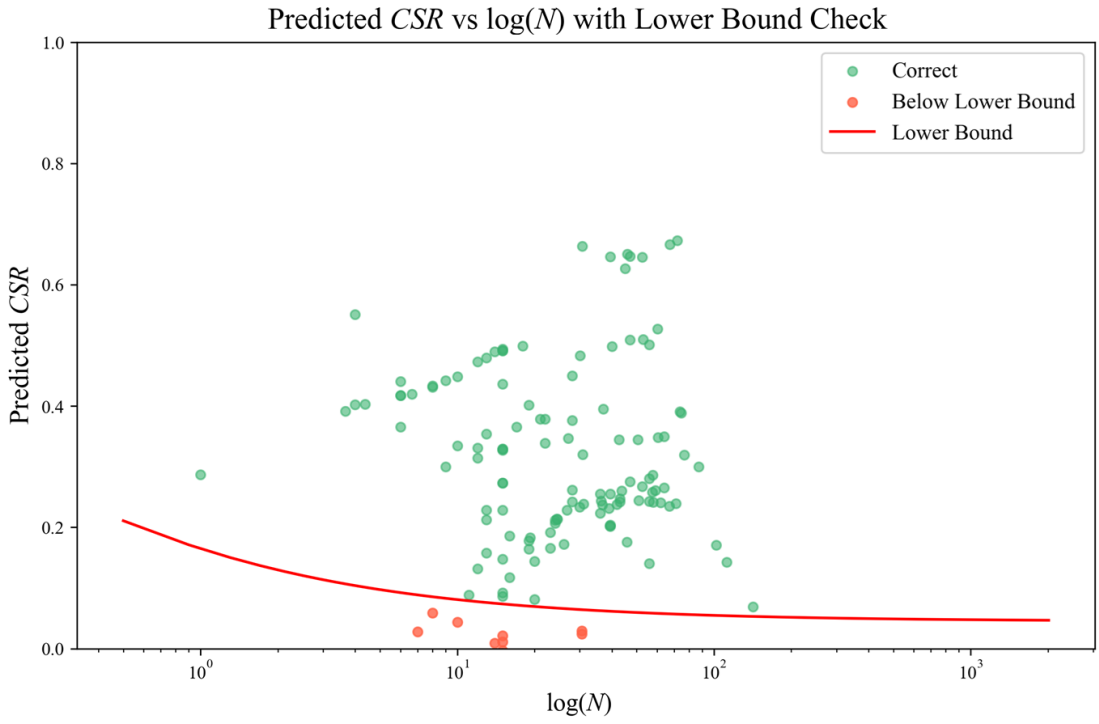


(a)

doi:10.6342/NTU202502436



(b)



(c)

圖 5.1 CSR 與 log(*N*)之關係圖與分類判斷線示意圖 (a) 所有資料點 (b) 訓練資料集與測試資料集 (c) 模型預測結果

doi:10.6342/NTU202502436

本章所使用的資料初步包含所有特徵，並未進行特徵篩選，旨在避免干涉模型所接觸到的資訊，使其能自行學習並辨識具預測能力的特徵。在缺失值處理方面，統一採用 MICE 填補 (Multiple Imputation by Chained Equations) 方法，根據先前模型比較結果，MICE 為整體表現最穩定之填補方式，亦能有效保留資料的變異性與分佈特性。

分析流程上，首先以所有特徵建立模型進行迴歸預測，取得各變數之重要性排序，接著取排名前五之特徵重新訓練模型，並利用該模型進行預測與分類分析，最後再依模型特性分別進行解釋性與不確定性分析。

5.2 EBM 模型分析

本節選用可解釋增強模型 (Explainable Boosting Machine, EBM) 作為代表性傳統機器學習模型，針對試驗資料進行迴歸預設與分類判斷。如第三章所述，EBM 為一種白箱模型，其核心概念來自廣義加法模型 (Generalized Additive Model, GAM)，並結合梯度提升 (Boosting) 技術提升模型穩定性與表現。

作為補充說明，加法模型 GAM 是一種統計學中的可解釋模型，假設輸出變數 y 可由一組非線性函數 f_j 與輸入特徵 x_j 的加總所構成，即式(5.2):

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) + \epsilon \tag{5.2}$$

其中，β_0 是截距項 (intercept)，為所有特徵值都是 0 時模型預測的基礎值，亦可視為平均預測值；ϵ 為誤差項或殘差，為實際值與模型預測值之間的誤差；而每個函數 $f_j(x_j)$ 可根據資料學習而得，不須假設其與輸出變數間存在線性關係。在後續的解釋性分析中，將使上述 GAM 概念呈現各個 $f_j(x_j)$ 所對應的貢獻函數曲線，說明不同特徵對 CSR 的影響趨勢。

以所有特徵建立之模型進行迴歸預測結果於表 5.1，由表可見，此 EBM 初始模型於訓練與測試資料皆展現良好的預測能力。如圖 5.2 所示，其分析出來之

doi:10.6342/NTU202502436

特徵重要性前五名為初始孔隙比 (e_0)、相對密度 (D_r)、通過百分比為 60%對應之土壤粒徑 (D_{60}) 、循環剪切次數 (N) 以及比重 (G_s)。

表 5.1 EBM 初始模型之預測分數結果表

Index	Training score	Testing score
R^2	0.8946	0.7804
MAE	0.0226	0.0397
MSE	0.0013	0.0035
RMSE	0.0360	0.0591

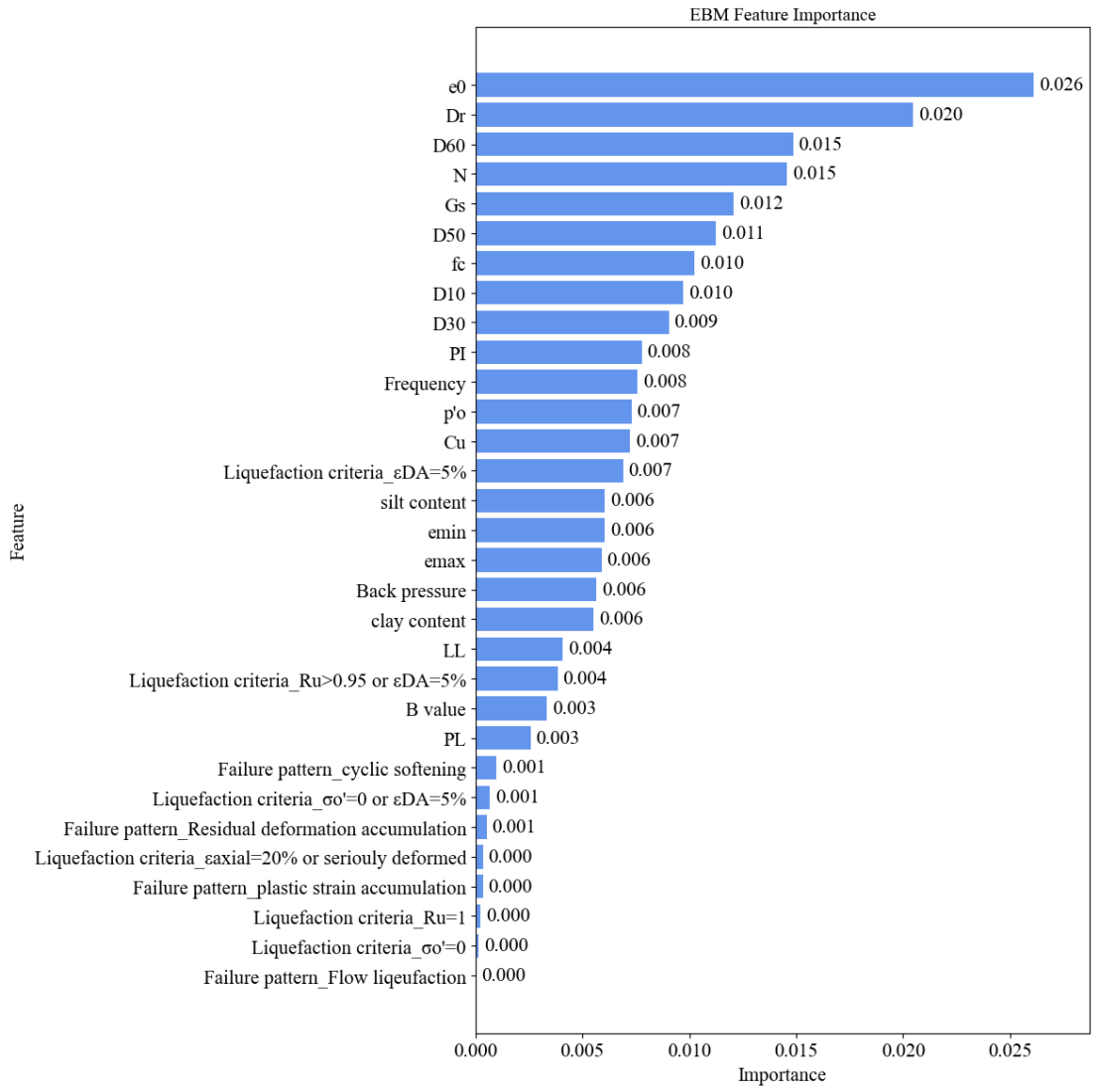


圖 5.2 EBM 初始模型之特徵重要性圖

doi:10.6342/NTU202502436

5.2.1 回歸預測結果

依前小節所挑選之五項特徵 (e_0、D_r、D_{60}、N、G_s) 重新訓練 EBM 模型，其迴歸預測表現整理於表 5.2。此新模型之結果雖未涵蓋所有特徵，預測指標略有降低，但整體結果仍不差，表示所挑選之五項特徵已涵蓋多數資訊，具有足夠預測能力。

表 5.2 EBM 模型之預測分數結果表

Index	Training score	Testing score
R^2	0.8303	0.6117
MAE	0.0314	0.0560
MSE	0.0021	0.0062
RMSE	0.0456	0.0786

圖 5.3 為訓練及與測試資料之預測散佈圖與殘差圖。於散佈圖 (圖 5.3(a)、(b))中可見預測結果與實際值整體呈現接近對角線之分佈，表示預測準確性良好；而於殘差圖 (圖 5.3(c)、(d))中可觀察誤差分佈無明顯偏移，進一步確認模型具穩定之預測表現。

圖 5.4 與圖 5.5 分別顯示 EBM 模型之迴歸分析中，包含與不包含交互作用項 (interaction terms) 之特徵重要性排序圖，可用以比較各特徵在模型中的平均貢獻程度。

doi:10.6342/NTU202502436

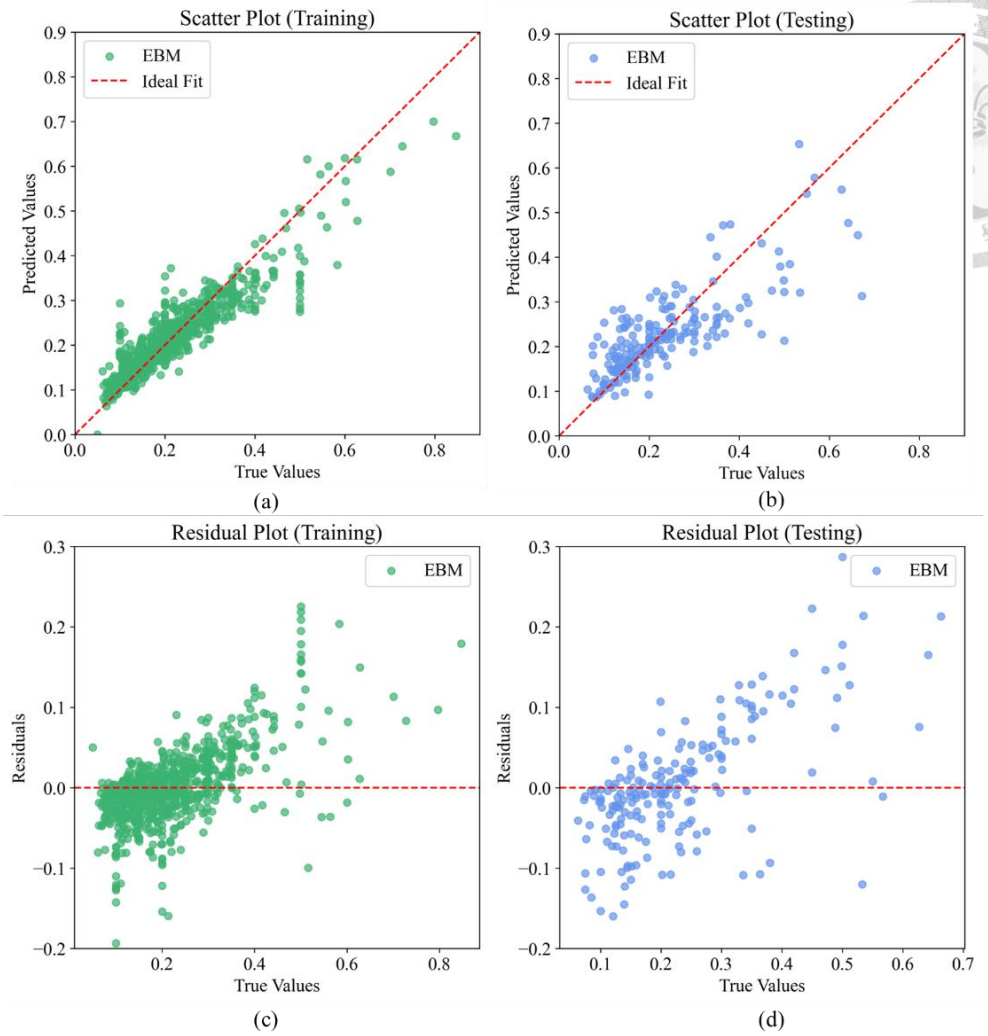


圖 5.3 EBM 模型於訓練與測試集之散佈與殘差圖

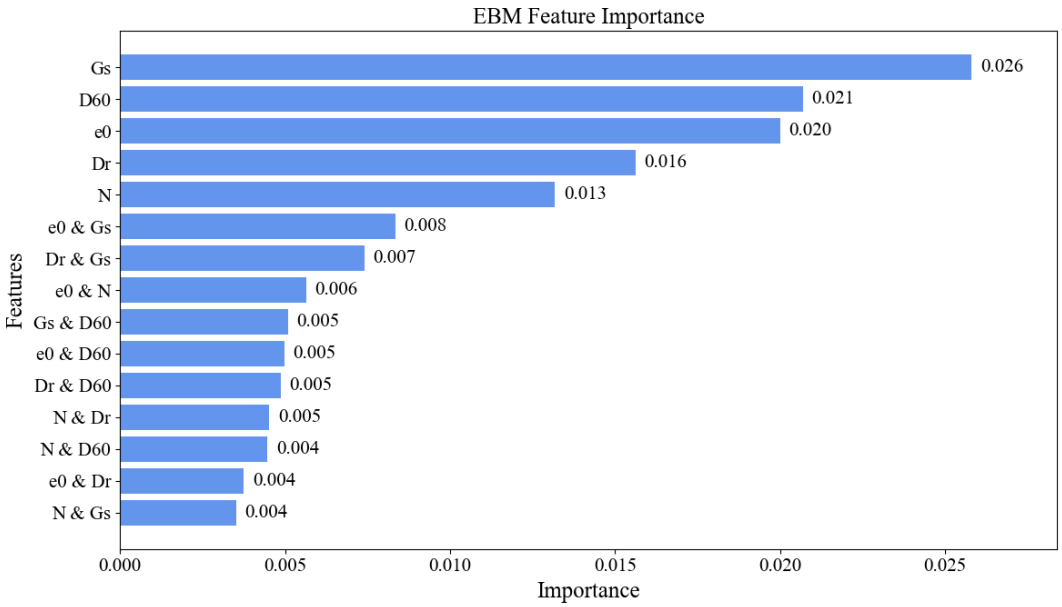


圖 5.4 EBM 模型含交互作用項之重要性排序圖

doi:10.6342/NTU202502436

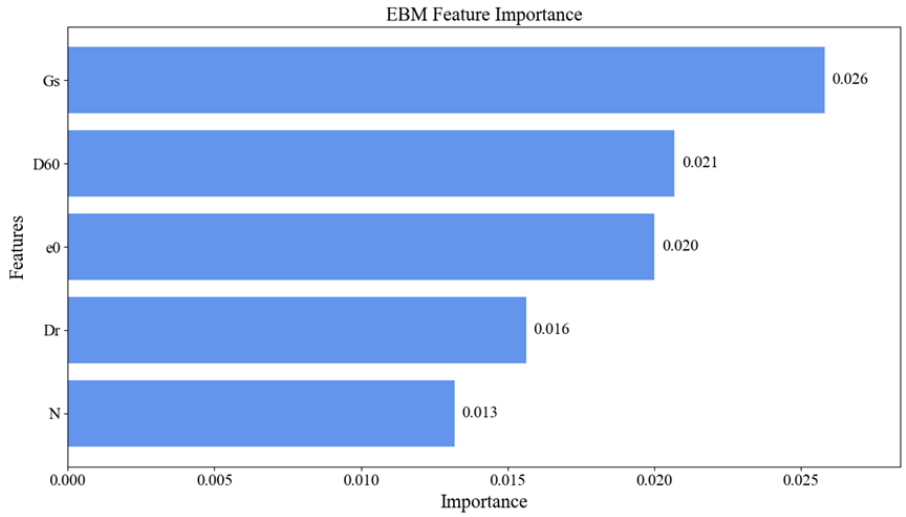


圖 5.5 EBM 模型不含交互作用項之重要性排序圖

5.2.2 模型分類結果

本節討論 EBM 模型於二元分類問題中的表現。分類依據為第 5.1 節所述之 CSR-*N* 下限曲線，做為液化與非液化之判斷基準。圖 5.6 呈現綠色的測試資料集中各樣本之 log(*N*) 與預測 CSR 值之分佈，並標示紅色的下限曲線。可見所有樣本之預測值均位於下限曲線上方，表示模型未預測出任何低於下限曲線的樣本，也就是模型未誤判任何非液化樣本。

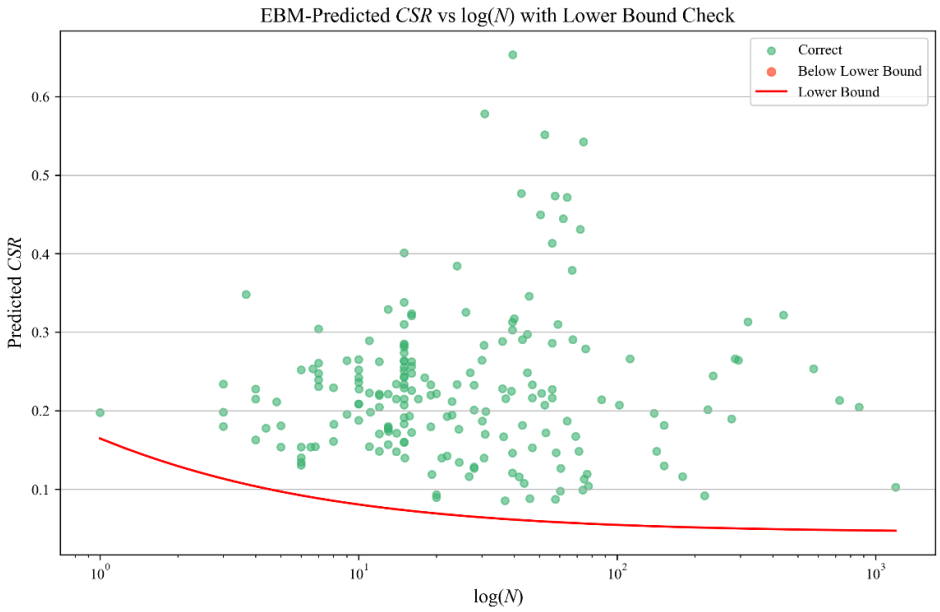


圖 5.6 EBM 模型預測 CSR 與 log(*N*)關係圖及下限線分類結果

doi:10.6342/NTU202502436

圖 5.7 為對應之混淆矩陣 (Confusion matrix)，將真實標籤 (True label) 與模型預測結果 (Predicted label) 進行對照。為符合本研究「預測是否發生液化」的分析目標，將 CSR 值位於下限曲線之上 (Above LB) 定義為正樣本。由於本研究所蒐集的資料來源皆為動態三軸試驗中實際發生液化之樣本，因此在真實標籤中不會出現低於下限曲線之下的樣本 (Below LB)。於混淆矩陣中，所有的測試樣本均被正確判斷為 Above LB，無誤判 (False Positive) 或漏判 (False Negative) 等情形。

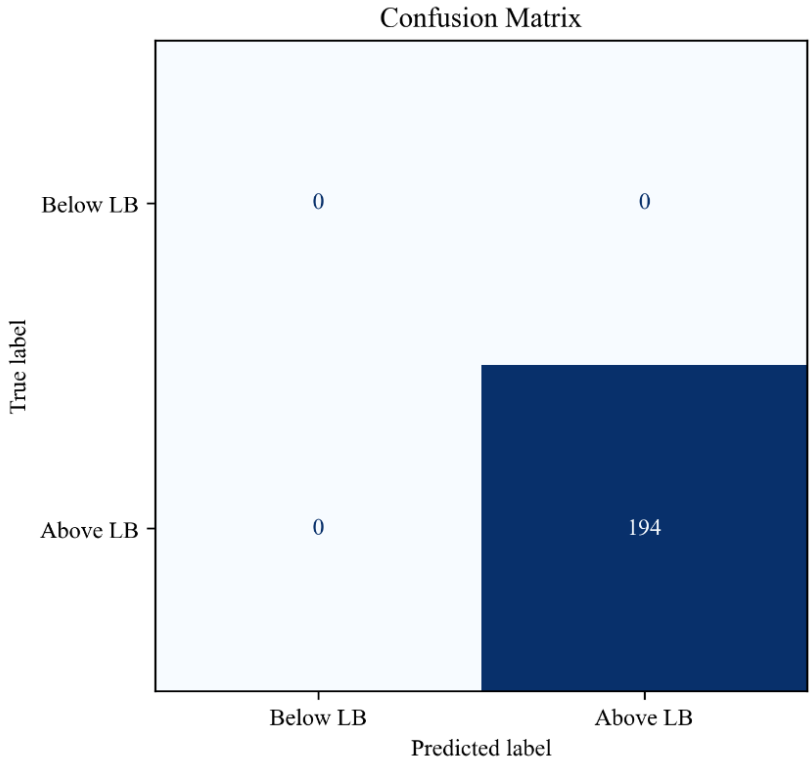


圖 5.7 EBM 模型混淆矩陣

分類評估指標整理於表 5.3，包含 Accuracy、Precision、Recall 與 F1-score 等，此結果說明 EBM 模型在此分類架構下具有良好的準確性與可信度。然而，由於資料中不包含非液化樣本，此一分類問題本質上為單一類別分類，其指標結果仍需搭備更多元的資料進行分析，才能發揮二元分類問題中各項指標的最大應用價值，後續將於第六章進一步討論其侷限性與應用建議。

doi:10.6342/NTU202502436

表 5.3 EBM 模型分類結果之各項指標

Index	Value
True Positive, TP	194
True Negative, TN	0
False Positive, FP	0
False Negative, FN	0
Accuracy	1
Precision	1
Recall	1
F1 score	1

5.2.3 解釋性分析

為發掘 EBM 模型特有的解釋性，針對 EBM 模型預測結果進行進一步的分析，欲說明前述五項特徵對 CSR 預測值的影響外，相較於前章所呈現之結果，本節額外加入 SHAP (Shapley Additive explanation summary Plot)，並搭配 EBM 模型內建之單變數貢獻函數圖 (Contribution plots)，一起說明模型對各特徵學習到的預測行為。

圖 5.8 為 EBM 模型之 SHAP 總結圖，橫軸為 SHAP 值，代表該特徵對模型預測的貢獻程度，正值表示增加預測值，負值則為降低；縱軸為各特徵名稱，點的顏色則代表該特徵值的高低，紅色表示高值，藍色為低值。

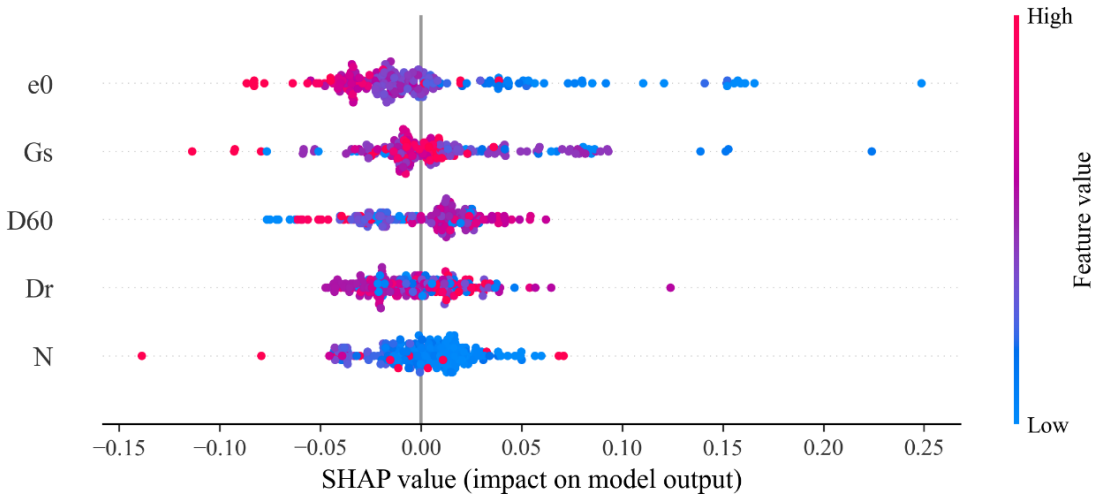


圖 5.8 EBM 模型 SHAP 總結圖

doi:10.6342/NTU202502436

以初始孔隙比 e_0 為例，其 SHAP 分佈顯示低值 (藍色) 多對模型輸出具有正貢獻 (SHAP 值為正)，表示初始孔隙比越低，模型傾向預測出較高之 CSR；反之，當 e_0 越高 (紅色)，其貢獻值多數集中在接近 0 或偏負區域，顯示高初始孔隙比會使模型傾向預測出較低之 CSR，此一觀察結果與土壤力學的基本觀念相符，即初始孔隙比較大的試體，顆粒排列越疏鬆，試體抗剪能力越低。

其他特徵亦可以用來說明模型解釋性。G_s 的高值多分佈於負的 SHAP 值等於 0 與負的區域，表示有負的貢獻，而低值多分佈於 SHAP 值等於 0 的右側，僅少部分有負的貢獻。顆粒比重 G_s 為反映土壤礦物組成的基本物理性質，雖不直接影響有效應力或目標預測值 CSR，但其在土壤中通常具穩定分布，值多介於 2.5 至 2.8 之間，變異性小，且在本研究中缺失率為 0%，可能使得模型充分學習其與目標預測值的關聯性。由於 Gs 可視為一種穩定變數，有助於模型區分不同樣本特徵，因此在預測中展現出較高的重要性。D_{60} 的 SHAP 值呈現雙向分佈，高值點大多落在 SHAP 的正號側，部分低值也有輕微正貢獻，此特徵與粒徑分佈之均勻程度有關，顯示對模型預測 CSR 有非線性的貢獻，可能需要搭配其他與粒徑相關的參數才能完成更全面的解釋。相對密度 D_r 作為土壤鬆緊程度的綜合性指標，直接關聯於抗液化強度高，其值多集中在 SHAP 值等於 0 的右側，說明相對密度高的試體有助於提高 CSR，低值分佈主要接近 0 或微偏負，相當於相對密度與 CSR 成正相關。最後是循環剪切次數 N，高值多分佈於負的 SHAP 值區，低值則多為正貢獻，此結果與抗液化曲線呈反比的關係相符，在試體強度相同的狀況下，循環剪切次數越多，所輸入的剪應力可以越小，反之，試體所受較少的循環剪切次數時，相對需要輸入較大的剪應力才會導致試體液化。

除 SHAP 總結圖外，EBM 模型亦提供每個特徵對預測結果之單變數貢獻函數圖 (Contribution plot)，可視覺化不同特徵的樹值範圍中，對模型輸出的影響趨勢。圖 5.9 至圖 5.13 分別呈現五項所選特徵 (e_0、G_s、D_{60}、D_r、N) 之貢獻函數圖，橫軸為特徵值的大小，縱軸為該特徵對預測 CSR 的貢獻分數 (Contribution

doi:10.6342/NTU202502436

Score)，藍色陰影區域代表模型對貢獻的不確定性區間 (Uncertainty Band)。

單變數貢獻函數圖中所呈現的藍色陰影區，即為不確定性區間 (Uncertainty Band)，亦可視為模型在該區段之信心水準。當不確定性區間較窄時，表示模型對該區段的學習結果相對穩定且有信心，具有較高的可信度；反之，若區間較寬，則代表該區間內樣本數較少，導致模型在該範圍內對資料分佈的學習結果不夠穩定，其預測貢獻可能有較大的波動性。以 D_{60} 為例，如圖 5.11，其在 0.4 至 0.6 區間的不確定性區域最窄，顯示模型對該數值範圍之貢獻趨勢較有信心，亦是資料樣本較密集的區段；在較高值 (>0.7) 與較低值 (<0.3) 的區域，則可觀察到不確定性範圍明顯擴大，顯示模型在此資料稀疏區段的預測結果較不穩定。

以圖 5.9 的初始孔隙比 e_0 為例，其在 0.35 至 0.45 間具有明顯正貢獻，隨數值上升其貢獻逐漸趨近 0 或轉為負值，顯示初始孔隙比較小的試體具有較高的 CSR 預測值，與 SHAP 總結圖所呈現的趨勢一致，其影響土壤顆粒間接觸程度與孔隙水壓的發展，是預測液化行為的核心參數。在圖 5.10 中比重 G_s 的值顯示在 2.6 附近貢獻值達到峰值，之後於 2.62 至 2.94 的區間內貢獻值在接近零的附近來回震盪，約於 2.75 再急遽的下降，隨後隨著數值增加貢獻值趨緩穩定，此非線性變化可能與蒐集資料中的樣本分佈有關。

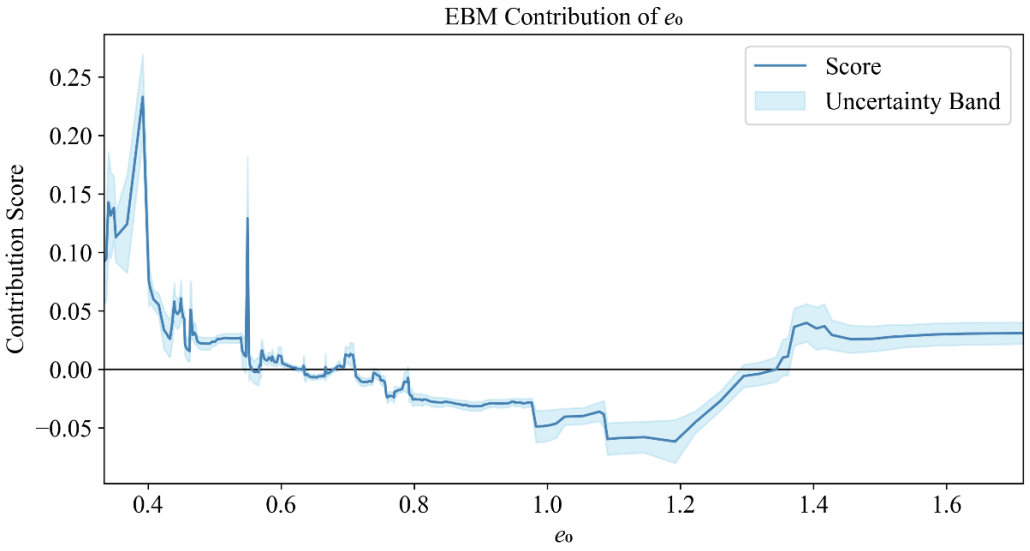


圖 5.9 EBM 模型中 e_0 對 CSR 的貢獻函數圖

doi:10.6342/NTU202502436

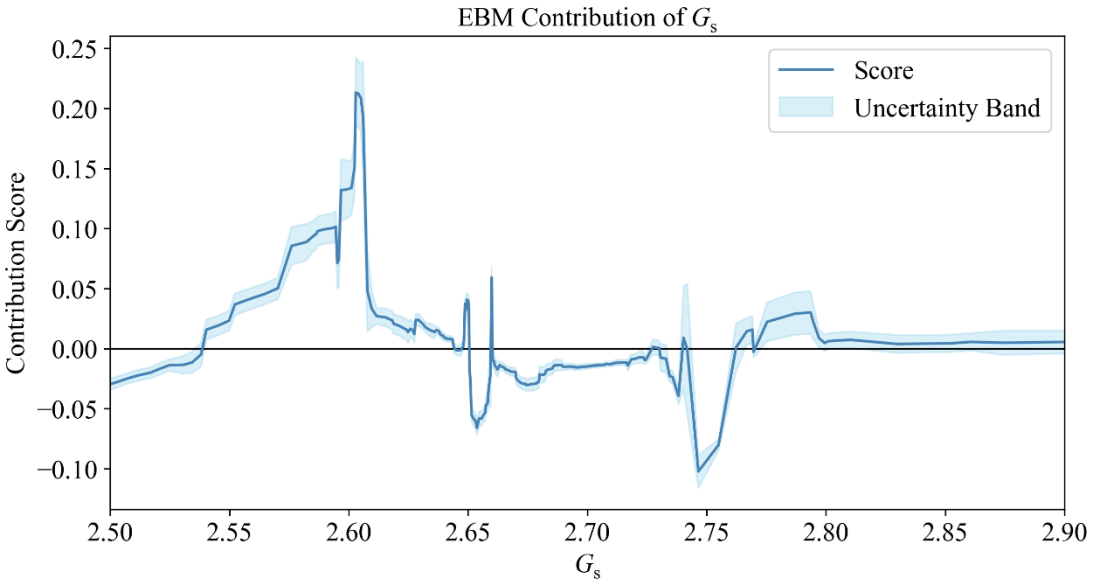


圖 5.10 EBM 模型中 G_s 對 CSR 的貢獻函數圖

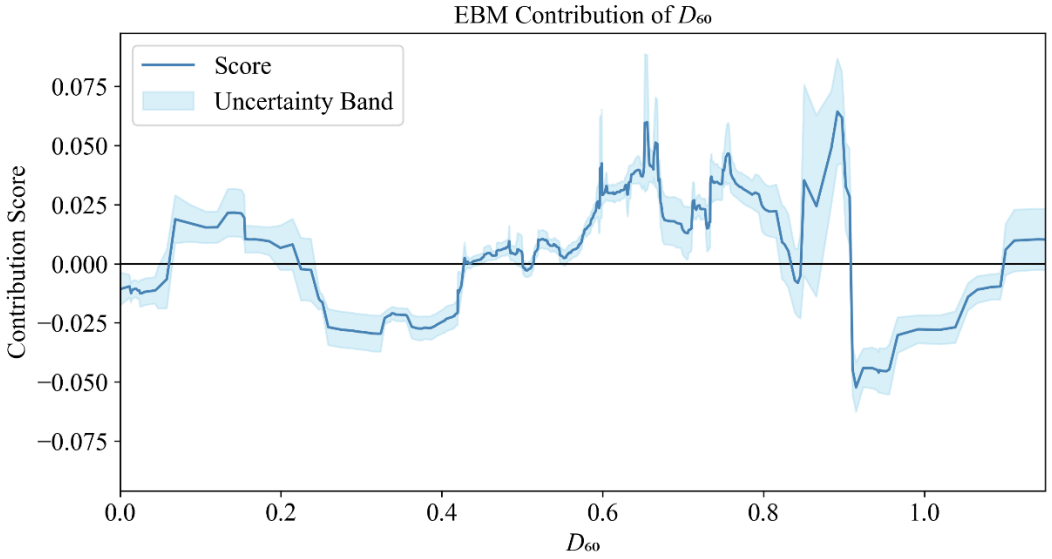


圖 5.11 EBM 模型中 D_{60} 對 CSR 的貢獻函數圖

通過百分比為 60%對應之土壤粒徑 D_{60} 之貢獻函數圖，如圖 5.11，其貢獻曲線呈現明顯的非線性分佈，約在 0.6 至 0.9 區間具有最大正貢獻，而在小於 0.4 和大於 0.9 的區間，貢獻值則為負值或趨近 0，這與 SHAP 圖中雙向分佈的觀察結果相符。圖 5.12 中相對密度 D_r 在約 60 以上具有相對大的正貢獻，小於 60 則貢獻值不穩定，小於 30 則為穩定負貢獻，顯示鬆散的試體對 CSR 預測具負面影響。

doi:10.6342/NTU202502436

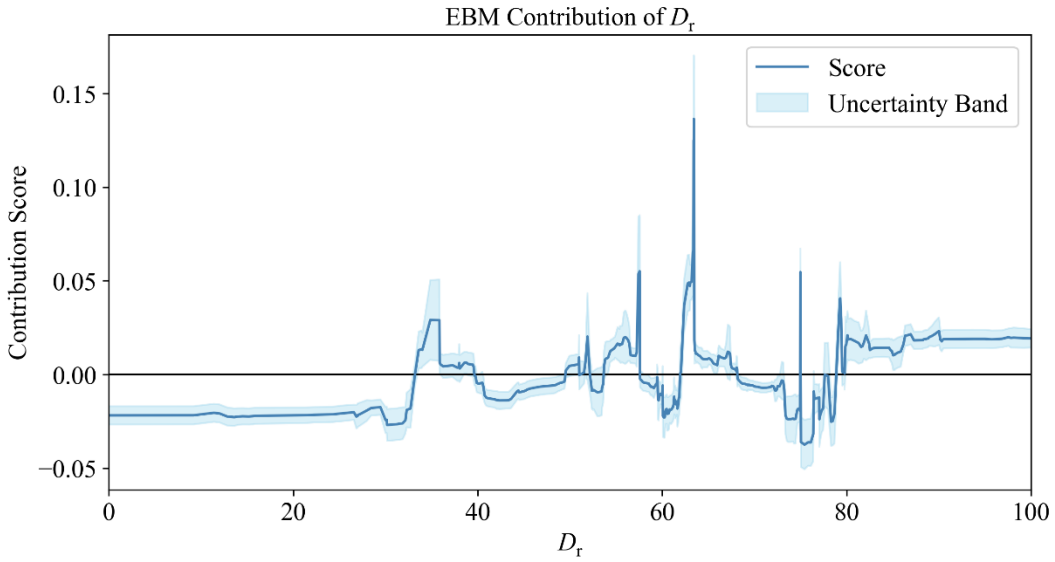


圖 5.12 EBM 模型中 D_r 對 CSR 的貢獻函數圖

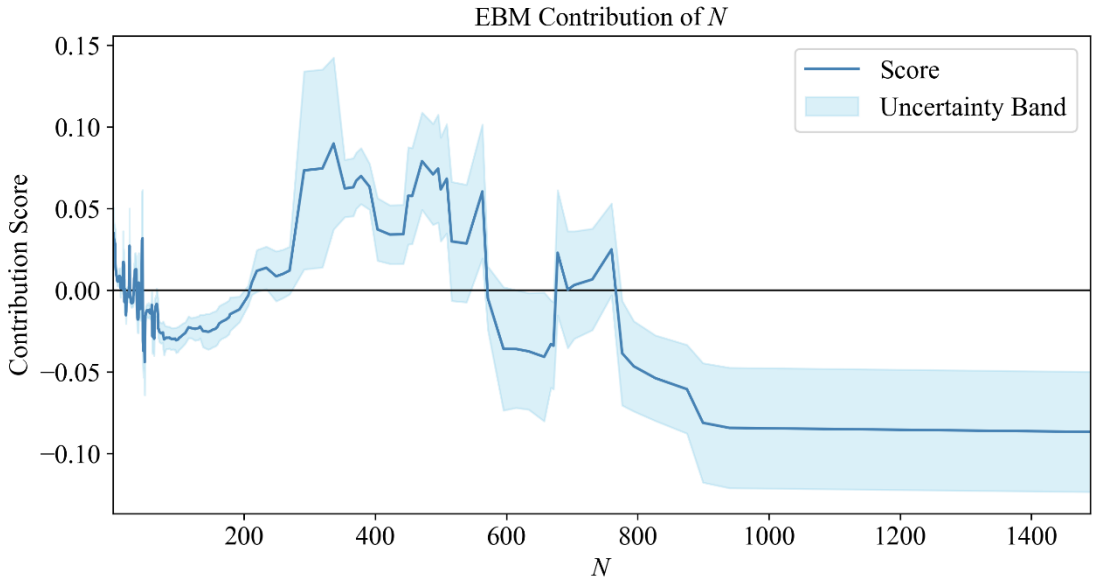


圖 5.13 EBM 模型中 N 對 CSR 的貢獻函數圖

最後，圖 5.13 為循環剪切次數 N 之貢獻函數圖，其在約 250 至 600 的區間顯示為正貢獻，而當 N 數值超過約 800 時，貢獻分數則顯著下降並趨於穩定負值。此趨勢與 SHAP 總結圖中觀察結果一致，高 N 值具有負貢獻，對應動態三軸試驗中「剪切次數越多，所需 CSR 越低」的典型反比關係。至於中間區段貢獻為正的情形，推測可能為相對貢獻的結果，在中等 N 值範圍內，模型預測的 CSR 相較於更低的 N 值略高，因而產生相對正貢獻。此外，EBM 所呈現的貢獻

doi:10.6342/NTU202502436

為「相對於基準值之貢獻分數」，不代表絕對關係，同時，此區段的正貢獻亦可能來自訓練資料分布或模型所學得之非線性關係結果，合理反映了試驗參數特性與資料中的區間行為變化。

整體而言，EBM 模型所提供之單變數貢獻函數圖與 SHAP 總結圖皆呈現出相似的特徵影響，能交叉驗證模型之學習結果，強化模型解釋力的同時，也提升了 EBM 模型在預測行為上的透明度與可信度。

若以廣義加法模型 (GAM) 的架構進行深入探討，輸出變數 y (即本研究中之目標參數 CSR) 可視為多組特徵貢獻函數 $f_j(x_j)$ 的加總結果。此節所挑選的五個特徵 (初始孔隙比 e_0、比重 G_s、通過百分比為 60%對應之土壤粒徑 D_{60}、相對密度 D_r、循環剪切次數 N) 即為模型輸入特徵 x_j，分別對應圖 5.9 至圖 5.13 所示之五組貢獻函數圖，也就是模型學習得之非線性函數f_j。

此加法架構使得模型在應用層面上具有彈性與操作性。未來若已知某試體樣本在上述五個特徵的數值，便可依據本研究所呈現之貢獻函數圖，查得各特徵於該值下對 CSR 的貢獻程度，逐項加總後即得最終之 CSR 預測值。此架構不僅有助於實務中對於目標變數的快速預測，亦可提供各特徵貢獻之解釋性與不確定性，為本研究在模型解釋與應用結合上提供示範，並展現其重要之實務價值。

doi:10.6342/NTU202502436

5.3 BNN 模型分析

本節將延續前述 EBM 模型之分析脈絡，進一步探討以深度學習方法中貝葉斯神經網路 (Bayesian Neural Network, BNN) 之應用性。相對於傳統人工神經網路 (ANN)，BNN 為更穩定的深度學習模型，具備不確定性推論的能力，能於預測過程中提供 CSR 預測值之信賴區間。由於 CSR 為受多個試驗因素影響之參數，其預測結果將具有一定變異性，BNN 於此類迴歸分析中具有應用潛力。本節將從迴歸預測結果、分類表現與不確定性分析三方面說明 BNN 模型之表現。

以所有特徵建立之初始 BNN 模型，其迴歸預測如表 5.4 所示，可見 BNN 初始模型相較於前節所建立之 EBM 模型表現相當。

表 5.4 BNN 初始模型之預測分數結果表

Index	Training score	Testing score
R^2	0.8933	0.7916
MAE	0.0212	0.0366
MSE	0.0013	0.0033
RMSE	0.0362	0.0576

圖 5.14 為依據模型訓練結果所得之特徵重要性排序圖，與先前模型所繪製的排序圖不同之處，在於每一條藍色橫條後方附有一條黑色橫線，代表該特徵重要性在不同抽樣下之標準差 (standard deviation)，反映模型對該特徵評估結果的變異程度。由於 BNN 模型內部包含機率分佈與不確定性推論機制，在每次訓練過程中會對模型參數進行隨機抽樣，使同一特徵在多次訓練中其重要性可能有所不同。透過平均值與標準差的同時呈現，可有效反映模型對各特徵影響力的穩定性與信心程度。其中，若黑線較短，表示該特徵在多次抽樣中評估結果較一致，模型對其重要性判斷具有較高的信心；若黑線較長，則表示該特徵的重要性變異較大，模型對其評估結果較不穩定，可能與樣本分佈不均獲特徵間存在交互作用有關。

doi:10.6342/NTU202502436

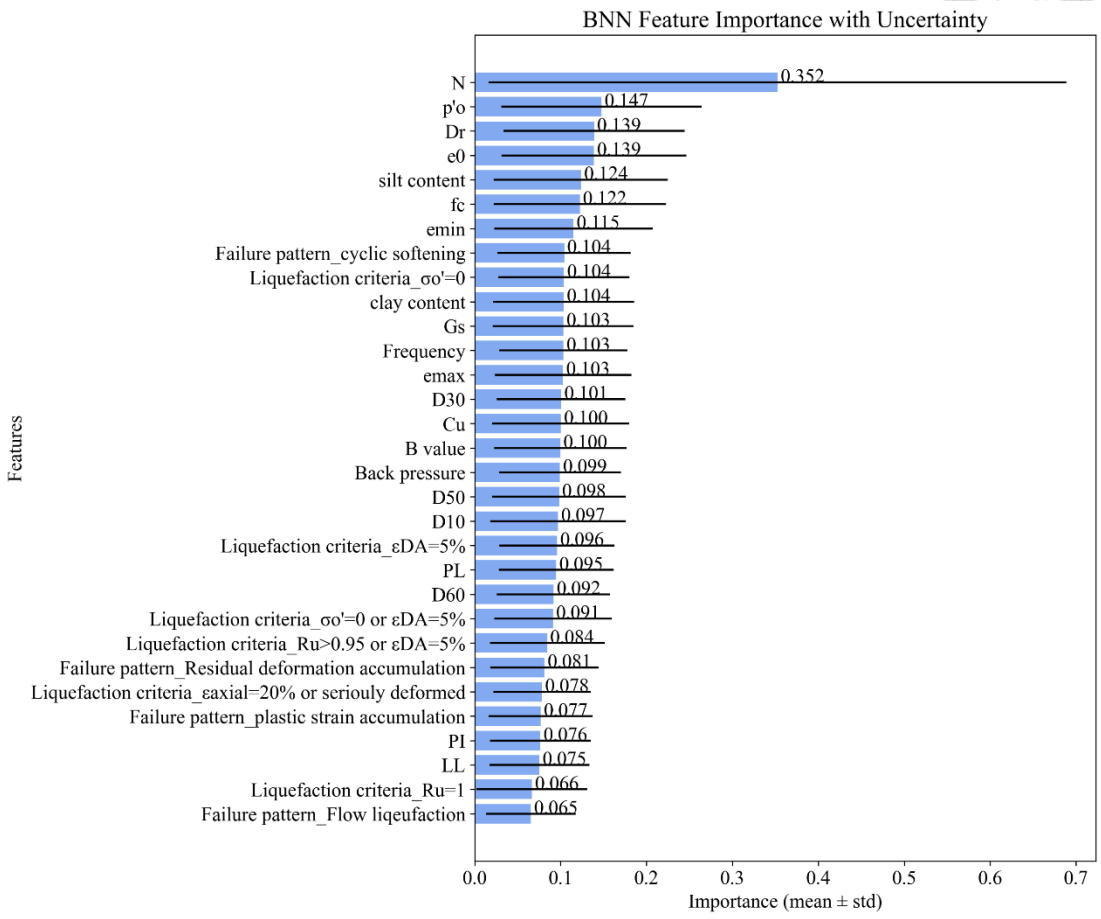


圖 5.14 BNN 初始模型之特徵重要性圖

此種可量化特徵不確定性的視覺化方式，正是 BNN 模型相較於其他深度學習架構的一項優勢，有助於提升模型解釋性與後續決策的透明度。

根據排序結果，BNN 模型中前五項最具影響力之特徵依序為循環剪切次數 (N)、相對密度 (D_r)、試驗圍壓 (p'_0)、初始孔隙比 (e_0) 與粉土含量 (silt content)，其中有三項 (e_0、N、D_r) 與前節之 EBM 模型結果相同，表示這些特徵具備對於不同模型的穩定性及關鍵性，對於 CSR 預測具有高參考價值。

doi:10.6342/NTU202502436

5.3.1 回歸預測結果

接續採用前述五項特徵建立新的 BNN 模型，其迴歸預測結果如表 5.5 所示。可見訓練集之 R^2 分數維持在 0.8 左右，然而測試集之 R^2 分數下降至 0.5657，明顯低於前節之 EBM 模型，顯示 BNN 模型對完整特徵的依賴性更高，特徵篩選後之預測能力下降，推測其非線性結構需要更多資訊才能夠支撐並有效學習。

表 5.5 BNN 模型之預測分數結果表

Index	Training score	Testing score
R^2	0.7801	0.5657
MAE	0.0372	0.0593
MSE	0.0027	0.0069
RMSE	0.0520	0.0831

圖 5.15 為該 BNN 模型於訓練集與測試集的散佈圖與殘差圖。從散佈圖 (圖 5.15(a)) 可觀察預測值與實際值整體趨勢分佈接近對角線，但在測試集 (圖 5.15(b)) 中可見預測偏差較大的情形。殘差圖 (圖 5.15 (c)、(d)) 則進一步顯示在中高 CSR 值區間的預測誤差分佈較為分散，表示模型在特定範圍內的擬合能力仍有進步空間。

圖 5.16 為 BNN 模型所計算之特徵重要性排序圖。可見循環剪切次數 (N) 仍為最具影響力之特徵，其次依序為粉土含量 (silt content)、初始孔隙比 (e_0)、相對密度 (D_r) 與試驗圍壓 (p'_0)。整體排序與先前 EBM 模型結果有所差異，顯示兩者於特徵影響評估方式上，於架構與學習機制存在的本質差異。

doi:10.6342/NTU202502436

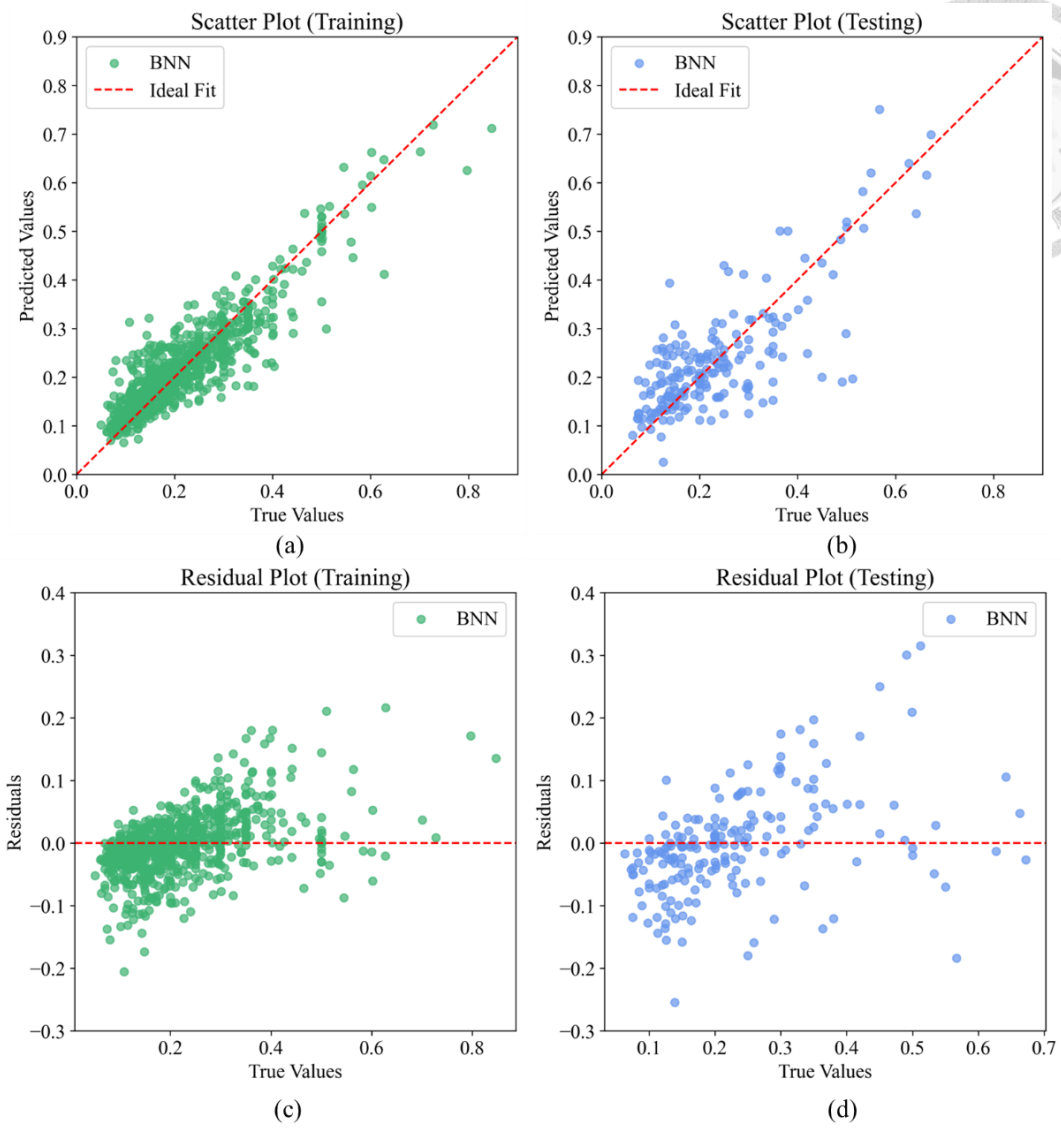


圖 5.15 BNN 模型於訓練與測試集之散佈與殘差圖

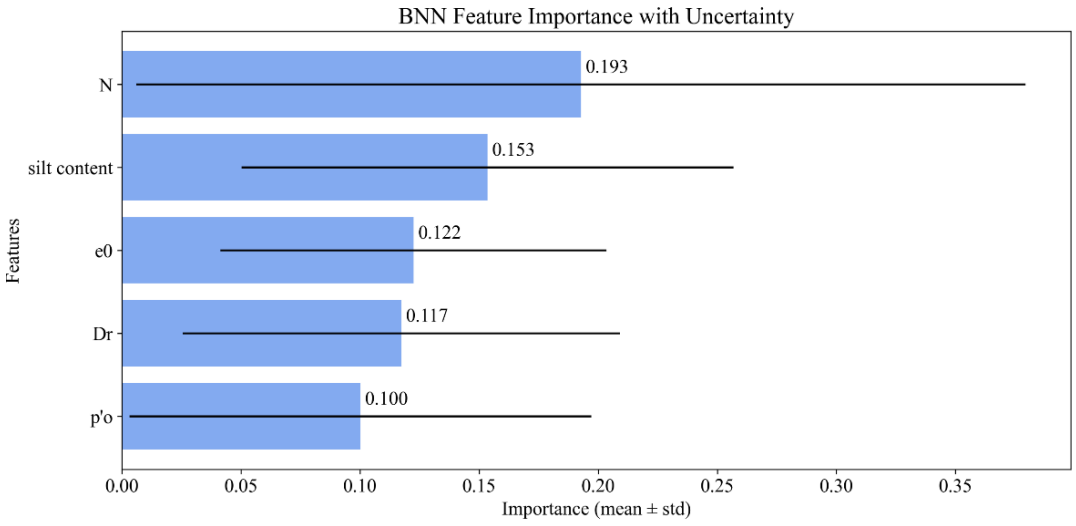


圖 5.16 BNN 模型之重要性排序圖

doi:10.6342/NTU202502436

為進一步檢視 BNN 模型於訓練歷程中的表現穩定性與收斂情形，本研究紀錄每一訓練週期 (epoch) 訓練資料 (Train) 與測試資料 (Test) 之誤差指標變化，包括損失函數 (Loss)、平均絕對誤差 (Mean Absolute Error, MAE) 與決定係數 (R^2)，將三者一同整理於圖 5.17。

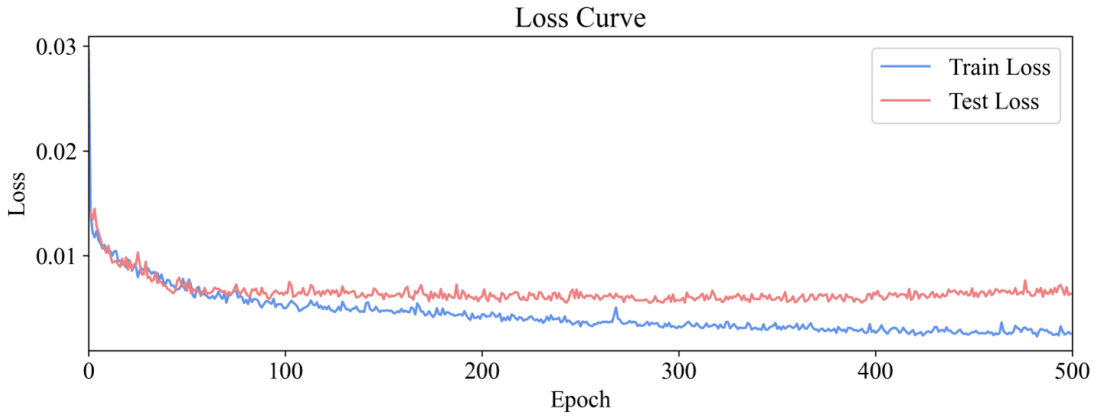
此種指標變化曲線可以用來觀察模型學習過程中的三種典型現象: 過擬合 (overfitting)、欠擬合 (underfitting) 與過度震盪 (instability)。若模型出現過擬合，通常表現為訓練誤差持續下降，但測試誤差在中期後反轉上升，表示模型雖能擬合訓練資料但無法泛化至沒見過的樣本；若模型欠擬合 (underfitting)，則訓練與測試誤差皆偏高，且曲線無明顯下降趨勢，可能代表模型容量不足或所選用的特徵無法提供足夠訊息，模型無法有效學習；若曲線於訓練過程中劇烈震盪，上下波動明顯，有可能是模型參數更新不穩定所致，原因可能是學習率設定過高、批次大小不適當，或模型架構不合適，此時應視情況調整後重新訓練。接下來將針對圖 5.17 中之三條曲線進行說明。

圖 5.17(a)為訓練與測試資料的損失函數變化圖。損失函數為深度學習模型訓練時的主要優化函數，本研究中以均方誤差 (MSE，於代碼中設為 `loss_fn = nn.MSELoss()`) 作為損失衡量指標。其數值代表預測值與真實質之平方誤差的平均值，值越小表示模型擬合得越好。圖中顯示，在訓練初期約前 100 輪，損失值由約 0.06 附近快速下降，隨後趨於穩定，表示模型成功學習資料特性並達到收斂，測試集的損失亦呈現類似趨勢，顯示模型並未出現明顯過擬合現象，足以說明 500 輪的訓練次數是足夠的；圖 5.17 (b) 為平均絕對誤差 (MAE) 變化趨勢，為另一衡量預測誤差的指標。與 MSE 相比，MAE 對異常值較不敏感，因此可用以輔助了解模型整體預測偏差的情形。由圖可見，訓練與測試的曲線均呈現下降並趨於穩定，亦顯示模型訓練過程整體穩定且預測偏差有逐步變小；圖 5.17(c) 中的決定係數 (R^2) 為衡量模型中預測資料變異的程度，數值越高表示模型擬合能力越佳。圖中顯示，訓練與測試之 R^2 分數於訓練初期迅速提升隨後趨於飽和，展現模型的預測能力逐漸建

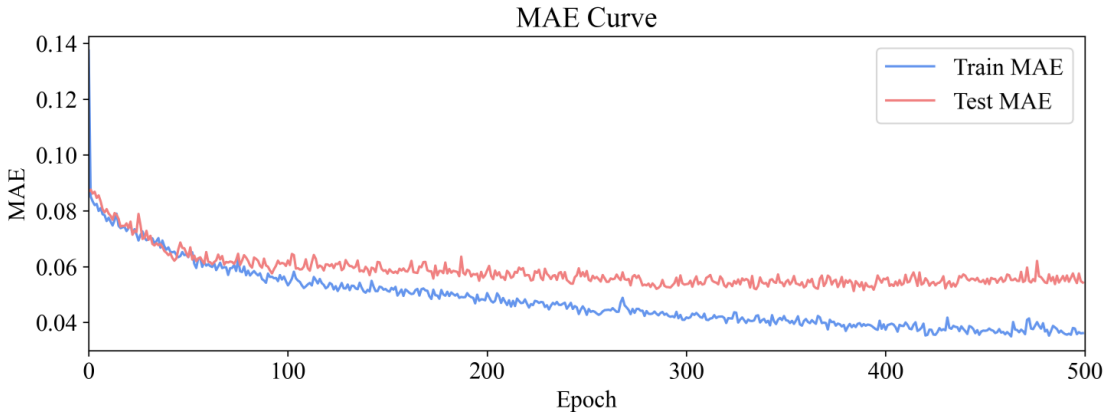
doi:10.6342/NTU202502436

立並趨於穩定，並未出現測試分數反轉下降之過擬合現象。

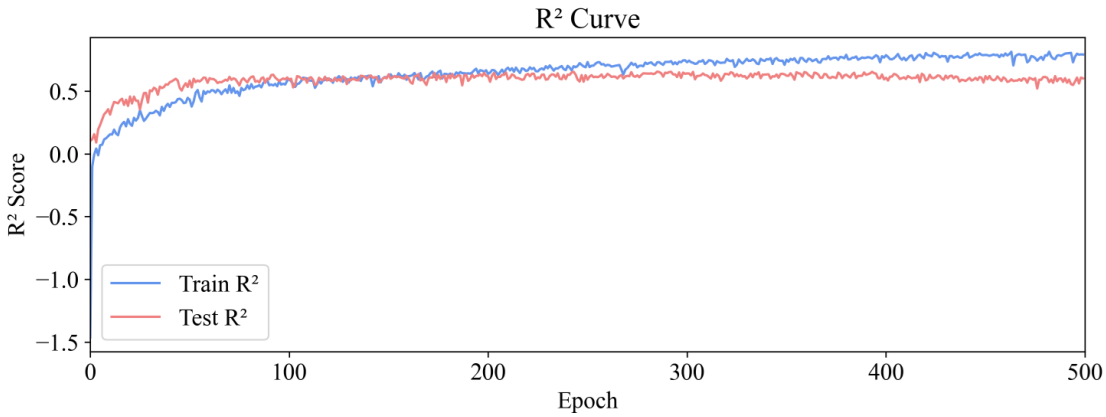
整體而言，三條訓練曲線顯示本研究所訓練之 BNN 模型可於 500 輪內達到收斂，且無過擬合、欠擬合或震盪情形，整體學習過程穩定，具備良好實用價值。



(a)



(b)



(c)

圖 5.17 BNN 模型於訓練迭代次數下之變化曲線圖

doi:10.6342/NTU202502436

5.3.2 模型分類結果

本節延續第 5.1 節所設定之 CSR-*N* 下限線作為分類標準，探討 BNN 模型對於液化與非液化樣本的判斷能力。分類標準同樣依據模型預測之 CSR 值與下限曲線進行比對，若預測值落於下限之上則視為液化樣本 (Above LB)，反之則為非液化樣本 (Below LB)。

由於 BNN 模型在訓練過程中需先對輸入特徵進行標準化處理，因此在模型預測階段亦須使用標準化後之樣本，這將導致若將標準化後的 *N* 值進行對數轉換，部分資料會因數值範圍問題而無法顯示於圖上。此小節分別呈現兩種圖以輔助分析。圖 5.18 為採對數尺度之 CSR-log(*N*) 關係圖，以利觀察低數值範圍內預測結果與下限線間之關係；圖 5.19 則使用未取對數之 *N* 值，以確保所有樣本皆能完整顯現，並標註預測錯誤之樣本。

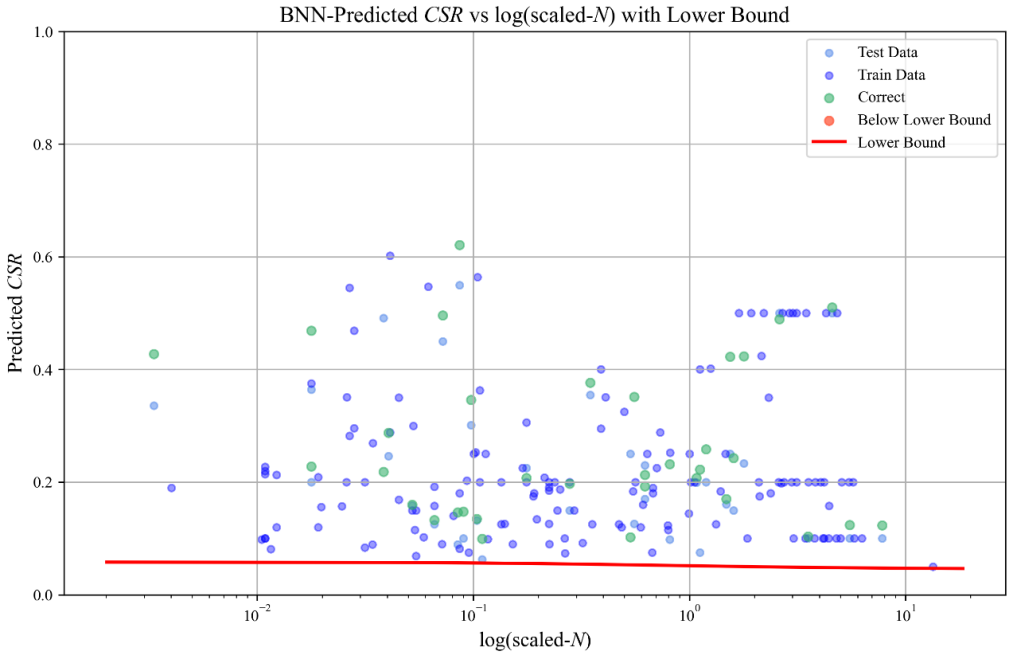


圖 5.18 BNN 模型預測 CSR 與 log(scaled-*N*)關係圖及下限線分類結果

由圖 5.19 可見，BNN 模型整體判斷表現良好，於所有測試樣本中只有兩筆樣本 (橘色點) 落於下限線之下，即被預測為非液化案例。相較於前節之 EBM 模型全數預測正確的表現，BNN 模型之分類準確性略有下降。

doi:10.6342/NTU202502436

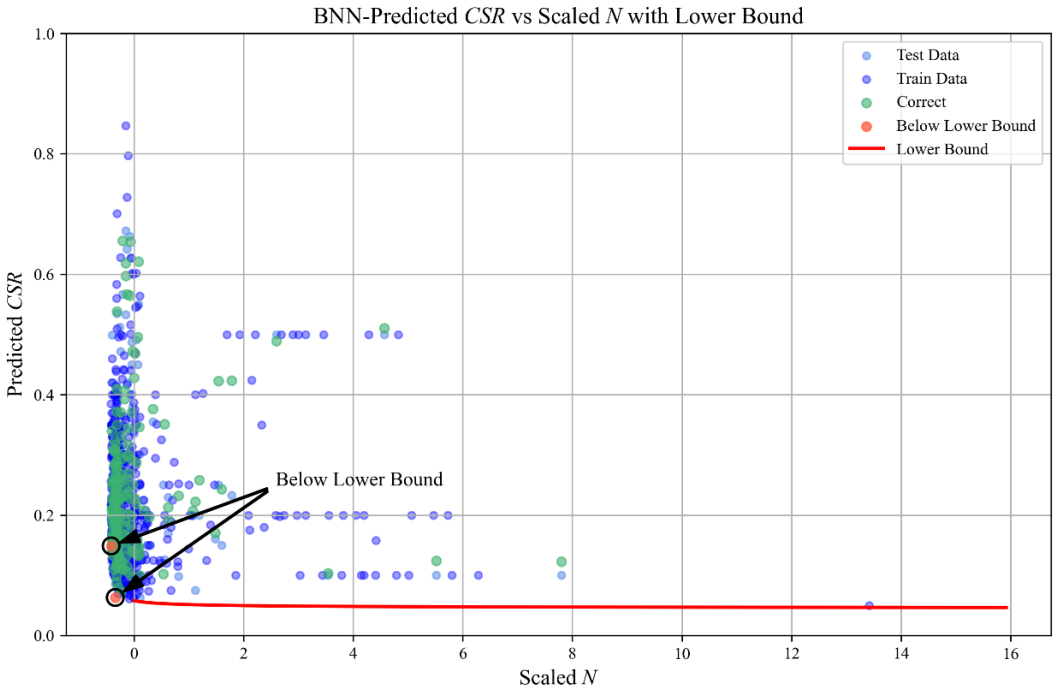


圖 5.19 BNN 模型預測 CSR 與 scaled-*N* 關係圖及下限線分類結果

圖 5.20 為 BNN 模型對應之混淆矩陣，橫軸為預測標籤，縱軸為真實標籤，依第 5.1 節所定義，CSR 高於下限曲線之資料點為正樣本 (Above LB)，低於下限者為負樣本 (Below LB)。

由圖可見，因本研究資料中無實際非液化樣本 (Below LB)，故混淆矩陣上半部皆為零。分類結果中，有 192 筆樣本被正確預測為液化 (TP)，2 筆樣本預測為非液化 (FN)，代表模型出現兩次漏判 (False Negative)。值得注意的是本章節的分類為一種後設分析方法，並非傳統定義下的二元分類，因此此兩點橘色的漏判點應理解為預測結果未能符合本研究所建立之分類準則。此結果可能反映模型在某些條件下對 CSR 預測的保守性或不確定性，亦可能揭示下界線本身在資料分布邊緣之適用性仍需進一步探討與精進。

doi:10.6342/NTU202502436

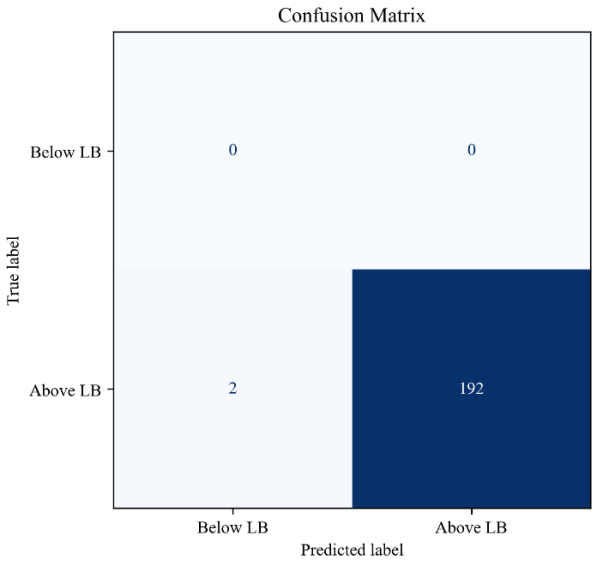


圖 5.20 BNN 模型混淆矩陣

分類評估指標整理於表 5.6。正確率 (Accuracy) 高達 98.7%，表示模型整體預測結果與實際情形相符；精準率 (Precision) 為 1.0，表示無誤判為液化的非液化樣本；召回率 (Recall) 為 0.9897，顯示模型能有效捕捉大部分液化樣本；F1 分數作為精準率與召回率的綜合評量，達到 0.9948，與前節 EBM 模型達成 100% 正確率的表現相比較差，但整體表現仍可視為優良。

表 5.6 BNN 模型分類結果之各項指標

Index	Value
True Positive, TP	192
True Negative, TN	0
False Positive, FP	0
False Negative, FN	2
Accuracy	0.987
Precision	1
Recall	0.9897
F1 score	0.9948

整理而言，BNN 模型在本研究設定下有優秀的分類能力，但考量本研究之資料集並不包含真實之非液化樣本，導致模型對該類別之識別能力無從認識與學習，其限制與應對方法將於第六章進行討論。

doi:10.6342/NTU202502436

5.3.3 不確定性分析

本節旨在呈現 BNN 模型於預測結果中所提供的不確定性，進一步增加模型的解釋性與實用性。如圖 5.21 為 BNN 模型在測試資料上預測之 CSR 結果與對應之不確定性區間 (uncertainty interval) 圖。

圖中藍色虛線為真實值 (True)，橘線為模型預測的平均值 (Predicted Mean)，橘色與淺橘色陰影區域分別代表 ± 1 標準差 ($\pm 1\sigma$) 與 ± 2 標準差 ($\pm 2\sigma$) 所形成的預測區間。此不確定性區間是利用模型內部參數的機率分佈推導而得，可反映模型對於每一筆樣本預測結果的信心程度。

由圖中可觀察到多數真實值均落於 $\pm 2\sigma$ 的區間內，顯示模型所給定的預測區間合理且可信賴，同時代表模型整體對不確定性有很好的掌握能力。另一方面，部分樣本於 $\pm 1\sigma$ 區間內即能涵蓋真實值，表示模型對於這些樣本的更具信心。針對其他不落在預測區間內的樣本，則可能表示該區段內資料具有較高的變異性，或是模型對該特徵條件下的資料學習效果不足，可作為後續強化資料品質或模型泛化能力的參考。

整體而言，BNN 模型所提供之預測不確定性範圍，除了有助於評估預測結果的信賴程度，亦可作為後續應用中進行判斷與決策的基礎。

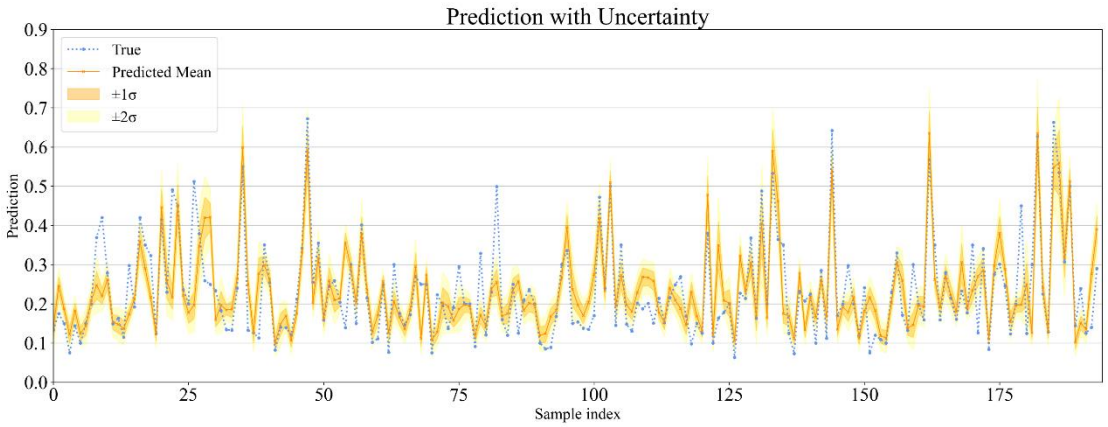


圖 5.21 BNN 模型預測結果與對應之不確定性區間

doi:10.6342/NTU202502436

第六章 結論與建議

6.1 結論

本研究針對土壤動態三軸試驗資料來預測液化時循環剪切次數，整理來自 32 篇學術期刊文獻，共蒐集近千筆動態三軸試驗數據，建立迴歸預測的機器學習架構，並進行分析結果比較。研究以循環剪應力比 (Cyclic Stress Ratio, CSR) 作為主要迴歸目標，並依據 CSR 與循環剪切次數 (N) 之關係建立分類標準，判斷試體是否液化，進一步提升模型結果的可讀性與應用價值。

研究採用多種資料前處理方式，包括缺失值填補策略 (如填入 NaN 或 0、KNN 填補與 MICE 填補)、類別型欄位之 One-Hot 編碼，以及數值型欄位之標準化處理，以確保資料品質與模型訓練穩定性。下一步使用七種監督式學習模型進行迴歸分析與效能評估，其中包含五種傳統機器學習方法，分別為隨機森林 (Random Forest, RF)、分類增強模型 (Categorical Boosting, CatBoost)、極端梯度提升機 (Extreme Gradient Boosting, XGBoost)、可解釋增強模型 (Explainable Boosting Machine, EBM) 與支持向量機 (Support Vector Machine, SVM)，以及兩種深度學習模型，為人工神經網路 (Artificial Neural Network, ANN) 與貝葉斯神經網路 (Bayesian Neural Network, BNN)，共七種模型。

研究前半段聚焦於不同資料處理方式下，各模型於迴歸分析中的預測準確度與穩定性，系統性的評估不同超參數設定、特徵組合與補值方法的影響。研究後半段則選擇具應用潛力的兩種模型進行深入分析。EBM 模型具備良好之特徵可解釋性，而 BNN 模型則可以提供預測信賴區間與不確定性分析。本研究亦針對變數篩選方式、SHAP 值視覺化、特徵貢獻圖與預測信賴區間等面向，提供詳細說明與分析。

doi:10.6342/NTU202502436

面對上述研究設計與模型分析結果，以下整理本研究所呈現之成果與其價值:

1. 手動調參優於使用 Grid search 調參，特別適用於深度學習模型

   相較於 Grid search 所依賴的固定參數網格組合，手動調參可根據模型行為做調整及設定，也能因應資料特性與模型架構的不同，找出相對穩定的最佳參數組合。特別是在 ANN 與 BNN 等深度學習模型中，Grid search 雖在訓練集表現良好，但常見測試集表現明顯下降，反映其易受超參數組合影響而發生過擬合。

2. 主要特徵即可支撐模型預測能力，次要特徵須審慎納入

   即便只使用主要特徵進行建模，多數模型已可達到不錯的預測準確度，顯示這些主要變數在預測中具高度代表性。雖部分模型在加入次要特徵後表現略有提升，但差異有限，且部分模型 (如 ANN、SVM) 因次要特徵造成雜訊干擾，導致測試表現下滑。

3. MICE 填補為整體最穩定之補值方式，KNN 相對來說易導致模型失準

   比較三種補值方式後發現，MICE (Multiple Imputation by Chained Equations) 能提供大多數模型的穩定表現，填補 NaN 或 0 為次之，最後是 KNN 填補方式，測試集與訓練集的分數有較大的差距，且於非 Boosting 類的模型來說表現降下頗多。

4. CatBoost、XGBoost 與 Random Forest 為七種模型中表現最好的三個模型

   本研究的樣本數相較於典型大數據應用較有限，像 CatBoost、XGBoost 與 Random Forest 等基於決策樹架構之模型，能在近千筆樣本下展現良好且穩定的預測表現。其原因在於這類模型對資料量的需求相對較低，且對於特徵尺度差異、缺失值及填補方式的精確性容忍度較高。相對而言，深度學習模型（如 ANN 與 BNN）雖理論上具備更強的學習能力，但其訓練過程仰賴大量資料以支持模型架構並培養其泛化能力。因此，在資料量有限且樣本品質不一的情況下，選用容錯力較高且效能穩定的樹模型，反而能獲得更佳的預測表現。

doi:10.6342/NTU202502436

5. 特徵重要性分析歸納出本研究的核心、次核心與次級變數

核心變數包含初始孔隙比 (e_0)、循環剪切次數 (N)、細粒料含量 (f_c) 與液化判斷標準 (Liquefaction criteria)；次核心變數包含平均粒徑 (C_u)、相對密度 (D_r)、試驗圍壓 (p'_0) 與反水壓 (Back pressure)；次級變數包含破壞模式 (Failure pattern) 與剪切頻率 (Frequency)，表示它們在預測 CSR 中的影響力階層，亦可視為具代表性與工程意義的關鍵特徵，其中，核心變數對液化潛勢之尤具參考價值，在未來模型建構上可作為優先考慮對象。雖然特徵的缺失率可能影響模型對其重要性的評估，但本研究的核心特徵多具低缺失率且重要性排序穩定，顯示缺失率對本研究特徵重要性的影響不大。

6. 透過比較不同模型的訓練時間，可了解模型的資源消耗與模型複雜度

訓練時間長短反映出模型架構的規模、超參數量及計算方式，如 Random Forest 與 SVM 等模型能於數秒至數分鐘內完成訓練，而 BNN 與 EBM 模型在特定設定下訓練時間可達數分鐘甚至數小時，此比較可作為模型選擇的依據，幫助使用者在準確度與時間成本上取得平衡。

7. 將迴歸問題轉成二元分類問題，EBM 與 BNN 仍有好的分類能力

透過建立 CSR 與 log(N) 關係為基礎的下限曲線，作為判斷液化與非液化樣本之二元分類標準，使原本迴歸問題具備分類判斷功能。於此分類架構下，EBM 與 BNN 模型均展現出良好的辨識準確度，成功辨識液化與非液化樣本。

8. EBM 模型可透過貢獻函數圖對特徵值進行加總，以獲得目標預測值

EBM 為廣義加法模型 (GAM) 的延伸形式，模型預測值為各輸入特徵貢獻的總和。透過該模型所提供之貢獻函數圖，使用者可查詢每一特徵值對預測結果的貢獻程度，進而逐項加總以獲得目標預測值。此步驟不僅透明且易於操作，亦為此研究之價值所在。

9. BNN 模型提供的預測信賴區間，知其對預測結果有好的掌握能力

BNN 模型在預測過程中能同時產出平均值與信賴區間，透過預測值的正負 1 與 2 個標準差 (σ) 的範圍提示預測結果的不確定性，多數真實值均落於

doi:10.6342/NTU202502436

「$\pm 2\sigma$」區間內，顯示預測區間整體合理且可信；部分樣本甚至在「$\pm 1\sigma$」區間內即能涵蓋真實值，代表模型對這些預測值具有高度信心。

從資料前處理、特徵選擇、模型比較、視覺化分析到應用延伸，本研究提出一套完整可操作的分析流程。未來若有不同的土壤試驗資料集或其他欲關注試驗參數之預測與分類需求，均可參考本研究架構進行延伸應用。

6.2 建議

根據本研究之研究設計、模型分析結果與應用延伸，以下針對資料與模型兩大面向提出幾點建議，以提升土壤液化預測模型的實用性與擴展性。

資料面

1. 增加資料多樣性以提升模型泛化能力

本研究之資料雖涵蓋多筆來自不同文獻之試驗數據，但集中於標準單元試體條件，建議可蒐集包含不同土壤類型 (如含礫石等) 與試體控制變數 (如試體尺寸與滲透係數等)，或是與初始空隙比或相對密度等參數具相關性的單位重資料，亦可作為補充特徵進行驗證，將有助於近一步了解其對模型預測效果的影響，同時以強化模型對動態三軸試驗行為的學習範疇。此外，為提升分類模型的資料平衡性，可加入未發生液化之樣本作為負類資料，避免模型僅學習液化案例，導致分類結果過度偏向「會液化」，或是嘗試設計如「以 BNN 模型漏判樣本為參考，再搭配正確預測樣本進行抽樣」的方法進行再訓練和分類，將有助於檢視模型對不同樣本的辨識能力與穩定性。

doi:10.6342/NTU202502436

2. 處以不平衡資料以提升模型實用性

本研究在分類分析中呈現明顯的類別不平衡問題，缺乏實際負累樣本，可參考 Jas and Dodagoudar (2023) 使用的 SMOTE (Synthetic Minority Over-sampling Technique) 進行少數類別樣本的合成擴增，在平衡資料後進行解釋性分析，觀察各特徵在模型中所扮演的角色，以確保平衡資料不會犧牲模型的可解釋性。另可利用如 BNN 模型的預測錯誤結果作為潛在非液化樣本，進行標記與再訓練，作為分類模型建構的替代策略。

3. 引入非正常樣本以強化模型判別能力

可嘗試導入「非正常樣本 (Anomaly Detection)」或「異常點分類 (outlier classification)」的概念進行模型建構。此類方法不需明確的負類標籤，而是透過學習正常樣本，辨識與其顯著不同的輸入條件，特別是當無法取得非液化資料時，此方法可作為分類模型的輔助策略，亦可結合模型預測誤差分析，例如以 BNN 模型所產生之預測不確定性較高或分類錯誤之樣本，作為潛在異常點進行標記與學習，以提升模型對潛在特殊樣本的敏感度與辨識能力。

4. 加入機制導向之特徵

許多相關研究傾向關注液化或破壞機制的過程及理解，其中所使用到的參數 (如孔隙水壓比 (R_u) 或試體應變 (ε_v) 等) 與其變化過程，亦可成為模型輸入的特徵之一。建議可嘗試將這類過程性參數做量化，例如計算 R_u 隨時間變化之斜率 (液化速率)，或 R_u 曲線下方的積分面積 (累積孔隙水壓發展量)，作為具有物理意義的衍生特徵，將有助於提升模型對於液化行為的辨識能力。

5. 擴展資料分佈有助於降低貢獻函數之不確定性

若資料涵蓋更完整的數值範圍與分佈型態，貢獻函數圖將能更細緻的呈現特徵值對預測結果的非線性關係，也能使不確定性區間 (Uncertainty band) 縮小，建議可特別針對信心區間較大的變數範圍做蒐集，以提升模型的泛用能力。

doi:10.6342/NTU202502436

6. CatBoost 模型對類別型變數處理策略

本研究使用 CatBoost 模型時，直接將類別變數輸入模型，未進行獨熱編碼處理，此為 CatBoost 模型的一項優勢，能有效處理類別型特徵，避免高維資料之相關問題，建議後續可針對相同資料集進行獨熱編碼後的模型訓練與比較，分析類別變數轉換方式對模型效能、訓練時間與特徵重要性之影響，作為特徵處理方法選擇之參考。

7. 無填補策略之可行性

本研究採用多種缺失值填補方式進行模型訓練與評估，建議可進一步比較「不進行缺失值填補」與「進行不同填補方式」下模型表現與特徵重要性排序的影響，特別是對於具備內建處理缺失能力之模型 (如 CatBoost) 而言，無填補策略亦可能具備一定的適用性與效益。

8. 資料填補的適用性

由於本研究所蒐集之資料來源涵蓋不同國家及地區，其試驗條件、樣本特性和地質背景皆可能存在差異，若直接對所有資料套用統一的補植策略 (如 KNN 或 MICE)，便可能產生以某地區樣本特徵補值其他地區樣本缺失的情形，進而混淆地區間的本質差異，降低模型對區域性特徵的辨識能力。因此可考慮依據地區來源進行分組補值，或是將「地區」做為模型特徵納入建模，以避免補值過程中地區資訊的遺失。

9. 強化超參數的調整

在實務應用中，測試集的表現常作為模型泛化能力與實用價值的關鍵指標，若能透過更細緻的超參數搜尋方式來提升測試集之預測準確度，將有助於建立更穩定且可信的模型表現，建議可針對測試集的分數進行分析，尤其當測試分數高於訓練分數時，代表模型成功從特徵中學習到具代表性的規則，而非僅僅記憶訓練資料，是模型訓練成功的重要指標之一。

doi:10.6342/NTU202502436

10. 分類標準之適用性

於第五章所建立之 CSR-log(*N*) 下限線為針對近 1000 筆動態三軸試驗資料所試誤擬合出的包絡線，具備資料集的代表性，並可作為模型預測結果轉換為分類標籤的依據，但該分類方式之判定仍有其侷限，並非適用於所有資料集，其原因主要是來自樣本分佈與多變數液化機制，實務上亦可能觀察到部分樣本雖落於下限線之下，卻在試驗中仍發生液化的情形。建議未來若資料組成或來源不同，仍須依據實際分佈狀況重新定義分類標準，並可搭配不確定性分析，辨識模型在界線附近預測結果的信心程度。本研究所建立之分類方式可作為後續研究參考架構，但不應作為唯一通用標準，需視個別資料特性調整分類依據與判定方式。

11. 以多為特徵建立分類包絡面

目前之分類判斷標準為以 CSR-log(*N*) 雙變數關係所建構之下限線。建議可嘗試引入第三維特徵 (如初始孔隙比或相對密度等)，建立三維包絡面 (Lower Bound Surface)，以提升分類邊界在多樣特徵條件下的代表性與靈活性，並提升模型在判斷上的說服力。

模型面

1. 使用跨模型融合機制以提升預測穩定性

本研究所探討之單一模型雖已涵蓋傳統與深度學習架構，建議可嘗試引入集成學習 (ensemble learning) 方式，如模型加權 (weighted averaging)、堆疊 (stacking) 或多模型平均 (model averaging) 等，以融合各模型的預測優勢，提升模型準確度與穩健性。

2. 重新評估特徵組合以提升模型效能

建議可根據特徵重要性排序結果，刪除在多數模型中影響力較低之變數，重新訓練並進行模型表現比較，評估簡化模型在預測準確度、運算效率與解

doi:10.6342/NTU202502436

釋性之綜合表現，將有助於建立更精簡且具工程實用性之模型架構。

3. 提升對模型結構與參數控制的理解

由於深度模型之行為與調參過程較複雜，其預測結果之穩定性較不如傳統機器學習模型。建議針對模型架構進行更細緻的調整與測試，以提升模型之可控性與訓練穩定性。

doi:10.6342/NTU202502436

參考文獻

Arif, A., Zhang, C., Sajib, M. H., Uddin, M. N., Habibullah, M., Feng, R., Feng, M., Rahman, M. S., & Zhang, Y. (2025). Rock Slope Stability Prediction: A Review of Machine Learning Techniques. *Geotechnical and Geological Engineering*, *43*(3), 124. https://doi.org/10.1007/s10706-025-03091-5

Ashfaque, J., & Iqbal, A. (2019). Introduction to Support Vector Machines and Kernel Methods.

Balamurugan, N. M., Kannadasan, R., Alsharif, M. H., & Uthansakul, P. (2022). A Novel Forward-Propagation Workflow Assessment Method for Malicious Packet Detection. *Sensors*, *22*(11), 4167. https://www.mdpi.com/1424-8220/22/11/4167

Belete, D., & D H, M. (2021). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, *44*, 1-12. https://doi.org/10.1080/1206212X.2021.1974663

Belov, D., & Armstrong, R. (2011). Distributions of the Kullback-Leibler divergence with applications. *The British journal of mathematical and statistical psychology*, *64*, 291-309. https://doi.org/10.1348/000711010X522227

Chen, R., Zhang, P., Wu, H., Wang, Z., & Zhong, Z. (2019). Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Frontiers of Structural and Civil Engineering*, *13*(6), 1363-1378. https://doi.org/10.1007/s11709-019-0561-3

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.1145/2939672.2939785

doi:10.6342/NTU202502436

Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Chwała, M., Kok-Kwang, P., Marco, U., Jie, Z., Limin, Z., & and Ching, J. (2023). Time capsule for geotechnical risk and reliability. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *17*(3), 439-466. https://doi.org/10.1080/17499518.2022.2136717

Demir, S., & Sahin, E. K. (2022). Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data. *Soil Dynamics and Earthquake Engineering*, *154*, 107130. https://doi.org/https://doi.org/10.1016/j.soildyn.2021.107130

Faska, Z., Lahbib, K., Haddouch, K., & el Akkad, N. (2023). A robust and consistent stack generalized ensemble-learning framework for image segmentation. *Journal of Engineering and Applied Science*. https://doi.org/10.1186/s44147-023-00226-4

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, *7*(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Häse, F., Fdez. Galván, I., Aspuru-Guzik, A., Lindh, R., & Vacher, M. (2019). How machine learning can assist the interpretation of ab initio molecular dynamics simulations and conceptual understanding of chemistry [10.1039/C8SC04516J]. *Chemical Science*, *10*(8), 2298-2307. https://doi.org/10.1039/C8SC04516J

Hernández, M., Ramon-Julvez, U., Vilades, E., Cordón Ciordia, B., Mayordomo, E., & Garcia-Martin, E. (2023). *Explainable artificial intelligence toward usable and*

doi:10.6342/NTU202502436

trustworthy computer-aided early diagnosis of multiple sclerosis from Optical Coherence Tomography. https://doi.org/10.48550/arXiv.2302.06613

Hsiao, C.-H., Chen, A., Ge, L., & Yeh, F.-H. (2022). Performance of artificial neural network and convolutional neural network on slope failure prediction using data from the random finite element method. *Acta Geotechnica*. https://doi.org/10.1007/s11440-022-01520-w

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, *574*, 1029-1041. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.04.085

Iqbal, S., Qureshi, A., Khursheed, K., Alhussein, M., Haider, S., & Rida, I. (2023). AMIAC: adaptive medical image analyzes and classification, a robust self-learning framework. *Neural Computing and Applications*, 1-29. https://doi.org/10.1007/s00521-023-09209-1

Ishihara, K., Troncoso, J., Kawase, Y., & Takahashi, Y. (1980). Cyclic Strength Characteristics of Tailings Materials. *Soils and Foundations*, *20*(4), 127-142. https://doi.org/https://doi.org/10.3208/sandf1972.20.4_127

Jafari-Marandi, R. (2021). Supervised or unsupervised learning? Investigating the role of pattern recognition assumptions in the success of binary predictive prescriptions. *Neurocomputing*, *434*, 165-193. https://doi.org/https://doi.org/10.1016/j.neucom.2020.12.063

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695. https://doi.org/10.1007/s12525-021-00475-2

Jas, K., & Dodagoudar, G. R. (2023a). Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dynamics*

doi:10.6342/NTU202502436

and Earthquake Engineering, *165*, 107662. https://doi.org/https://doi.org/10.1016/j.soildyn.2022.107662

Jas, K., & Dodagoudar, G. R. (2023b). Liquefaction Potential Assessment of Soils Using Machine Learning Techniques: A State-of-the-Art Review from 1994–2021. *International Journal of Geomechanics*, *23*(7), 03123002. https://doi.org/doi:10.1061/IJGNAI.GMENG-7788

Khan, M. Y., Qayoom, A., Nizami, M., Siddiqui, M. S., Wasi, S., & Syed, K.-U.-R. R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity*. https://doi.org/10.1155/2021/2553199

Kimura, N., Yoshinaga, I., Sekijima, K., Azechi, I., & Baba, D. (2019). Convolutional Neural Network Coupled with a Transfer-Learning Approach for Time-Series Flood Predictions. *Water*, *12*, 96. https://doi.org/10.3390/w12010096

Kose, T., Özgür, S., Cosgun, E., Keskinoğlu, A., & Keskinoğlu, P. (2020). Effect of Missing Data Imputation on Deep Learning Prediction Performance for Vesicoureteral Reflux and Recurrent Urinary Tract Infection Clinical Study. *BioMed Research International*, *2020*, 1-15. https://doi.org/10.1155/2020/1895076

Liu, C., & Macedo, J. (2024). Machine learning-based models for estimating liquefaction-induced building settlements. *Soil Dynamics and Earthquake Engineering*, *182*, 108673. https://doi.org/https://doi.org/10.1016/j.soildyn.2024.108673

Mahajan, A., Das, S., Su, W., & Bui, V.-H. (2024). Bayesian Neural Network-Based Approach for Probabilistic Prediction of Building Energy Demands. https://doi.org/10.20944/preprints202409.2211.v1

doi:10.6342/NTU202502436

Manjrekar, O., & Duduković, M. (2019). Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique. *Chemical Engineering Science: X*, *2*, 100023. https://doi.org/10.1016/j.cesx.2019.100023

Maxwell, A., Sharma, M., & Donaldson, K. (2021). Explainable Boosting Machines for Slope Failure Spatial Predictive Modeling. *Remote Sensing*, *13*, 4991. https://doi.org/10.3390/rs13244991

Mitelman, A., Yang, B., Urlainis, A., & Elmo, D. (2023). Coupling Geotechnical Numerical Analysis with Machine Learning for Observational Method Projects. *Geosciences*, *13*(7), 196. https://www.mdpi.com/2076-3263/13/7/196

Monego, V. S., Anochi, J. A., & de Campos Velho, H. F. (2022). South America Seasonal Precipitation Prediction by Gradient-Boosting Machine-Learning Approach. *Atmosphere*, *13*(2), 243. https://www.mdpi.com/2073-4433/13/2/243

Nguegnang, G. M., Rauhut, H., & Terstiege, U. (2024). Convergence of gradient descent for learning linear neural networks. *Advances in Continuous and Discrete Models*, *2024*(1), 23. https://doi.org/10.1186/s13662-023-03797-x

Obermeier, S. F. (1996). Use of liquefaction-induced features for paleoseismic analysis — An overview of how seismic liquefaction features can be distinguished from other features and how their regional distribution and properties of source sediment can be used to infer the location and strength of Holocene paleo-earthquakes. *Engineering Geology*, *44*(1), 1-76. https://doi.org/https://doi.org/10.1016/S0013-7952(96)00040-3

Öztornaci, B., Ata, B., & Kartal, S. (2024). Analysing Household Food Consumption in Turkey Using Machine Learning Techniques. *Agris on-line Papers in*

doi:10.6342/NTU202502436

Economics and Informatics, *16*, 97-105. https://doi.org/10.7160/aol.2024.160207

Pan, K., Cai, Y. Q., Yang, Z. X., & Pan, X. D. (2019). Liquefaction of sand under monotonic and cyclic shear conditions: Impact of drained preloading history. *Soil Dynamics and Earthquake Engineering*, *126*, 105775. https://doi.org/https://doi.org/10.1016/j.soildyn.2019.105775

Pan, K., Zhou, G. Y., Yang, Z. X., & Cai, Y. Q. (2020). Comparison of cyclic liquefaction behavior of clean and silty sands considering static shear effect. *Soil Dynamics and Earthquake Engineering*, *139*, 106338. https://doi.org/https://doi.org/10.1016/j.soildyn.2020.106338

Papadopoulou, A., & Tika, T. (2008). The Effect of Fines on Critical State and Liquefaction Resistance Characteristics of Non-Plastic Silty Sands. *Soils and Foundations*, *48*(5), 713-725. https://doi.org/https://doi.org/10.3208/sandf.48.713

Park, S.-S., & Kim, Y.-S. (2013). Liquefaction Resistance of Sands Containing Plastic Fines with Different Plasticity. *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(5), 825-830. https://doi.org/doi:10.1061/(ASCE)GT.1943-5606.0000806

Prakash, S., & Sandoval, J. A. (1992). Liquefaction of low plasticity silts. *Soil Dynamics and Earthquake Engineering*, *11*(7), 373-379. https://doi.org/https://doi.org/10.1016/0267-7261(92)90001-T

Prokhorenkova, L. G., Gleb; Vorobev, Aleksandr; Dorogush, Anna Veronika; Gulin, Andrey. (2017). CatBoost: Unbiased Boosting with Categorical Features. *arXiv*(arXiv:1706.09516). https://arxiv.org/abs/1706.09516

doi:10.6342/NTU202502436

Ragb, H., Ali, R., Jera, E., & Buaossa, N. (2021). Convolutional neural network based on transfer learning for breast cancer screening. https://doi.org/10.48550/arXiv.2112.11629

Rani, V., Nabi, S. T., Kumar, M., Mittal, A., & Kumar, K. (2023). Self-supervised Learning: A Succinct Review. *Archives of Computational Methods in Engineering*, *30*(4), 2761-2775. https://doi.org/10.1007/s11831-023-09884-2

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536. https://doi.org/10.1038/323533a0

Seed, H. B. (1976). *Evaluation of soil liquefaction potantial for level ground during earthquakes*.

Seed, H. B., & Idriss, I. M. (1971). Simplified Procedure for Evaluating Soil Liquefaction Potential. *Journal of the Soil Mechanics and Foundations Division*, *97*(9), 1249-1273. https://doi.org/doi:10.1061/JSFEAQ.0001662

Shakir Mohamed, M. R., Michael Figurnov, Anfriy Mnih. (2020). Monte Carlo gradient estimation in machine learning. *Macjine Learning Research*, 1-63.

Stamatopoulos, C. A. (2010). An experimental study of the liquefaction strength of silty sands in terms of the state parameter. *Soil Dynamics and Earthquake Engineering*, *30*(8), 662-678. https://doi.org/https://doi.org/10.1016/j.soildyn.2010.02.008

Sujon, K. M., Hassan, R. B., Towshi, Z. T., Othman, M. A., Samad, M. A., & Choi, K. (2024). When to Use Standardization and Normalization: Empirical Evidence From Machine Learning Models and XAI. *IEEE Access*, *12*, 135300-135314. https://doi.org/10.1109/ACCESS.2024.3462434

Tonyalı, İ., Akbas, S. O., Beyaz, T., Kayabalı, K., & Gokceoglu, C. (2024). Case study of a foundation failure induced by cyclic softening of clay during the 2023

doi:10.6342/NTU202502436

Kahramanmaraş earthquakes. *Engineering Geology*, *332*, 107477. https://doi.org/https://doi.org/10.1016/j.enggeo.2024.107477

Tsai, P. H., Lee, D. H., Kung, G. T.-C., & Hsu, C. H. (2010). Effect of content and plasticity of fines on liquefaction behaviour of soils. *Quarterly Journal of Engineering Geology and Hydrogeology - Q J ENG GEOL HYDROGEOL*, *43*, 95-106. https://doi.org/10.1144/1470-9236/08-019

Tsukamoto, Y., & Ishihara, K. (2010). Analysis on Settlement of Soil Deposits Following Liquefaction During Earthquakes. *Soils and Foundations*, *50*(3), 399-411. https://doi.org/https://doi.org/10.3208/sandf.50.399

Tsukamoto, Y., Ishihara, K., & Sawada, S. (2004). Settlement of Silty Sand Deposits Following Liquefaction During Earthquakes. *Soils and Foundations*, *44*(5), 135-148. https://doi.org/https://doi.org/10.3208/sandf.44.5_135

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373-440. https://doi.org/10.1007/s10994-019-05855-6

Wahab, S., Salami, B. A., AlAteah, A. H., Al-Tholaia, M. M. H., & Alahmari, T. S. (2024). Exploring the interrelationships between composition, rheology, and compressive strength of self-compacting concrete: An exploration of explainable boosting algorithms. *Case Studies in Construction Materials*, *20*, e03084. https://doi.org/https://doi.org/10.1016/j.cscm.2024.e03084

Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, *136*, 190-197. https://doi.org/https://doi.org/10.1016/j.patrec.2020.05.035

doi:10.6342/NTU202502436

Wei, X., Yang, Z., & Yang, J. (2023). Cyclic failure characteristics of silty sands with the presence of initial shear stress. *Soil Dynamics and Earthquake Engineering*, *171*, 107909. https://doi.org/https://doi.org/10.1016/j.soildyn.2023.107909

Wu, S., Zhang, J.-M., & Wang, R. (2021). Machine learning method for CPTu based 3D stratification of New Zealand geotechnical database sites. *Advanced Engineering Informatics*, *50*, 101397. https://doi.org/https://doi.org/10.1016/j.aei.2021.101397

Xinghuo, Y., Efe, M. O., & Kaynak, O. (2002). A general backpropagation algorithm for feedforward neural networks learning. *IEEE Transactions on Neural Networks*, *13*(1), 251-254. https://doi.org/10.1109/72.977323

Xu, L., Chen, Y., Zuo, L., Sun, M., & Li, W. (2024). An experimental investigation on undrained cyclic behaviour of a saturated intact loess. *Soil Dynamics and Earthquake Engineering*, *181*, 108668. https://doi.org/https://doi.org/10.1016/j.soildyn.2024.108668

Xueqian, N., Ye, B., Zhang, F., & Feng, X. (2021). Influence of Specimen Preparation on the Liquefaction Behaviors of Sand and Its Mesoscopic Explanation. *Journal of Geotechnical and Geoenvironmental Engineering*, *147*, 04020161. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002456

Youd, T. L., Idriss, I. M., Andrus, R. D., Arango, I., Castro, G., Christian, J. T., Dobry, R., Finn, W. D. L., Harder, L. F., Hynes, M. E., Ishihara, K., Koester, J. P., Liao, S. S. C., Marcuson, W. F., Martin, G. R., Mitchell, J. K., Moriwaki, Y., Power, M. S., Robertson, P. K.,…Stokoe, K. H. (2001). Liquefaction Resistance of Soils: Summary Report from the 1996 NCEER and 1998 NCEER/NSF Workshops on Evaluation of Liquefaction Resistance of Soils. *Journal of Geotechnical and Geoenvironmental Engineering*, *127*(10), 817-833. https://doi.org/doi:10.1061/(ASCE)1090-0241(2001)127:10(817)

doi:10.6342/NTU202502436

Zhang, S. (2012). Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, *85*(11), 2541-2552. https://doi.org/https://doi.org/10.1016/j.jss.2012.05.073

Zhang, W., Gu, X., Hong, L., Han, L., & Wang, L. (2023). Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges. *Applied Soft Computing*, *136*, 110066. https://doi.org/https://doi.org/10.1016/j.asoc.2023.110066

Zhang, Z. (2016). Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of Translational Medicine*, *4*(2), 30. https://atm.amegroups.org/article/view/8847

Zhang, Z., Ding, S., & Sun, Y. (2021). MBSVR: Multiple birth support vector regression. Information Sciences, 552, 65-79. https://doi.org/https://doi.org/10.1016/j.ins.2020.11.033

Zhao, T., Shen, F., & Xu, L. (2024). Review and comparison of machine learning methods in developing optimal models for predicting geotechnical properties with consideration of feature selection. *Soils and Foundations*, *64*(6), 101523. https://doi.org/https://doi.org/10.1016/j.sandf.2024.101523

Zhou, G. Y., Pan, K., & Yang, Z. X. (2023). Energy-based assessment of cyclic liquefaction behavior of clean and silty sand under sustained initial stress conditions. *Soil Dynamics and Earthquake Engineering*, *164*, 107609. https://doi.org/https://doi.org/10.1016/j.soildyn.2022.107609

doi:10.6342/NTU202502436

附錄 A 超參數影響分析與模型效能趨勢

表 A.1 對應圖 4.2 之 CatBoost 各超參數設定

圖 4.2	depth	iterations	learning_rate	L2_leaf_reg	random_strength	border_count	Training score	Testing score
(a)	5	500	0.05	3	5	200	0.796	0.633
	10	500	0.05	3	5	200	0.853	0.638
	16	500	0.05	3	5	200	0.879	0.669
(b)	10	300	0.05	3	5	200	0.838	0.631
	10	500	0.05	3	5	200	0.853	0.638
	10	1000	0.05	3	5	200	0.874	0.651
(c)	10	500	0.01	3	5	200	0.657	0.554
	10	500	0.05	3	5	200	0.853	0.638
	10	500	0.1	3	5	200	0.877	0.670
(d)	10	500	0.05	1	5	200	0.851	0.662
	10	500	0.05	3	5	200	0.853	0.638
	10	500	0.05	5	5	200	0.848	0.639
(e)	10	300	0.05	3	1	200	0.797	0.620
	10	300	0.05	3	3	200	0.797	0.629
	10	500	0.05	3	5	200	0.853	0.638
(f)	10	500	0.05	3	5	32	0.824	0.640
	10	500	0.05	3	5	200	0.853	0.638
	10	500	0.05	3	5	255	0.861	0.660

doi:10.6342/NTU202502436

表 A.2 對應圖 4.3 之 SVM 各超參數設定

圖 4.3	degree	coef0	c	epsilon	gamma	Training score	Testing score
(a)	2	1	1	0.1	scale	0.265	0.212
	3	1	1	0.1	scale	0.360	0.316
	4	1	1	0.1	scale	0.407	0.345
(b)	4	0	1	0.1	scale	0.340	0.300
	4	0.1	1	0.1	scale	0.355	0.318
	4	10	1	0.1	scale	0.513	0.424
(c)	4	10	1	0.1	scale	0.513	0.424
	4	10	10	0.1	scale	0.595	0.417
(d)	4	10	1	0.1	scale	0.513	0.424
	4	10	1	0.01	scale	0.546	0.408
	4	10	1	0.001	scale	0.546	0.414
(e)	3	1	100	0.01	scale	0.822	0.684
	3	1	100	0.01	auto	0.810	0.697
	3	1	100	0.01	0.01	0.673	0.586
	3	1	100	0.01	0.1	0.818	0.682

doi:10.6342/NTU202502436

表 A.3 對應圖 4.4 之 XGBoost 各超參數設定

圖 4.4	max_depth	min_child_weight	learning_rate	n_estimators	gamma	reg_alpha	reg_lambda	subsample	colsample_bytree	Training score	Testing score
(a)	3	5	0.05	1200	0	0.1	3	0.8	0.8	0.938	0.757
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	8	5	0.05	1200	0	0.1	3	0.8	0.8	0.982	0.772
(b)	5	2	0.05	1200	0	0.1	3	0.8	0.8	0.979	0.775
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	10	0.05	1200	0	0.1	3	0.8	0.8	0.955	0.757
(c)	5	5	0.01	1200	0	0.1	3	0.8	0.8	0.913	0.720
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.1	1200	0	0.1	3	0.8	0.8	0.983	0.777
(d)	5	5	0.05	1000	0	0.1	3	0.8	0.8	0.902	0.712
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1500	0	0.1	3	0.8	0.8	0.977	0.774
(e)	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1200	0.1	0.1	3	0.8	0.8	0.567	0.483
	5	5	0.05	1200	0.5	0.1	3	0.8	0.8	0.219	0.199
(f)	5	5	0.05	1200	0	0	3	0.8	0.8	0.925	0.716
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1200	0	0.5	3	0.8	0.8	0.847	0.681
(g)	5	5	0.05	1200	0	0.1	1	0.8	0.8	0.923	0.724
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1200	0	0.1	5	0.8	0.8	0.905	0.716
(h)	5	5	0.05	1200	0	0.1	3	0.7	0.8	0.906	0.718
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1200	0	0.1	3	1	0.8	0.913	0.723
(i)	5	5	0.05	1200	0	0.1	3	0.8	0.7	0.909	0.713
	5	5	0.05	1200	0	0.1	3	0.8	0.8	0.973	0.771
	5	5	0.05	1200	0	0.1	3	0.8	1	0.921	0.727

doi:10.6342/NTU202502436

表 A.4 對應圖 4.5 之 EBM 各超參數設定

圖 4.5	max_bins	max_interaction_bins	interactions	learning_rate	max_rounds	early_stopping_rounds	min_samples_leaf	max_leaves	Training score	Testing score
	128	32	10	0.01	5000	50	20	3	0.830	0.679
(a)	256	32	10	0.01	5000	50	20	3	0.868	0.680
	512	32	10	0.01	5000	50	20	3	0.831	0.673
	256	16	10	0.01	5000	50	20	3	0.812	0.664
(b)	256	32	10	0.01	5000	50	20	3	0.868	0.680
	256	64	10	0.01	5000	50	20	3	0.852	0.682
	256	32	10	0.01	5000	50	20	3	0.868	0.680
(c)	256	32	20	0.01	5000	50	20	3	0.864	0.684
	256	32	30	0.01	5000	50	20	3	0.868	0.693
	256	32	10	0.005	5000	50	20	3	0.859	0.706
(d)	256	32	10	0.01	5000	50	20	3	0.868	0.680
	256	32	10	0.05	5000	50	20	3	0.819	0.665
	256	32	10	0.01	5000	50	20	3	0.868	0.680
(e)	256	32	10	0.01	8000	50	20	3	0.835	0.680
	256	32	10	0.01	10000	50	20	3	0.835	0.680
	256	32	10	0.01	5000	50	20	3	0.868	0.680
(f)	256	32	10	0.01	5000	80	20	3	0.838	0.679
	256	32	10	0.01	5000	100	20	3	0.839	0.680
	256	32	10	0.01	5000	50	20	3	0.868	0.680
(g)	256	32	10	0.01	5000	50	30	3	0.798	0.638
	256	32	10	0.01	5000	50	50	3	0.762	0.594
	256	32	10	0.01	5000	50	20	3	0.868	0.680
(h)	256	32	10	0.01	5000	50	20	5	0.849	0.695
	256	32	10	0.01	5000	50	20	8	0.865	0.696

doi:10.6342/NTU202502436

表 A.5 對應圖 4.6 之 BNN 各超參數設定

圖 4.6	hdden_layers	use_dropout	activation_fn	learning_rate	epochs	optimizer	Training score	Testing score
(a)	[256,128,128,64,32,16]	[True]	nn.Sigmoid	0.001	500	Adam	0.761	0.721
	[256,128,128,64,32,16]	[True]	nn.Tanh	0.001	500	Adam	0.892	0.720
	[256,128,128,64,32,16]	[True]	nn.ReLU	0.001	500	Adam	0.880	0.747
	[256,128,128,64,32,16]	[True]	nn.Softplus	0.001	500	Adam	0.824	0.714
	[256,128,128,64,32,16]	[True]	nn.LeakyReLU (0.01)	0.001	500	Adam	0.896	0.782
(b)	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	500	Adam	0.924	0.782
	[256,128,128,64,32,16]	[True]	nn.GELU	0.01	500	Adam	-0.028	-0.041
	[256,128,128,64,32,16]	[True]	nn.GELU	0.1	500	Adam	-0.054	-0.069
(c)	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	300	Adam	0.895	0.780
	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	500	Adam	0.924	0.782
	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	1000	Adam	0.954	0.770
(d)	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	500	Adam	0.924	0.782
	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	500	SGD	-0.008	-0.035
	[256,128,128,64,32,16]	[True]	nn.GELU	0.001	500	RMSprop	0.914	0.785

doi:10.6342/NTU202502436

表 A.6 對應圖 4.7 之 ANN 各超參數設定-1

ANN	Sequential							history					
圖 4.7	hdden_layers	dropout_rates	use_batch_norm	loss	optimizer	activations	metrics	validation_split	epochs	batch_size	verbose	Training score	Testing score
(a)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	100	32	1	0.556	0.350
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	200	32	1	0.692	0.474
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	250	32	1	0.667	0.561
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	275	32	1	0.633	0.453
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	32	1	0.745	0.642
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	500	32	1	0.741	0.596
(b)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	16	1	0.655	0.554
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	32	1	0.636	0.497
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	64	1	0.788	0.551
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	128	1	0.778	0.445
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	0.1	300	256	1	0.863	0.547
(c)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.636	0.497
	[256, 128, 128, 64, 32, 16]	[0.001, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.715	0.547
	[256, 128, 128, 64, 32, 16]	[0.0001, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.769	0.621
(f)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.640	0.457
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	False	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.598	0.306
(g)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.636	0.497
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	huber	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.797	0.595
(h)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.640	0.457
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	SGD (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.256	0.058
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	RMSprop (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.366	0.209
(i)	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	sigmoid	[mae, r2_metric]	32	300	100	1	0.636	0.497
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	tanh	[mae, r2_metric]	32	300	100	1	0.434	0.324
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	relu	[mae, r2_metric]	32	300	100	1	0.789	0.666
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	leaky_relu	[mae, r2_metric]	32	300	100	1	0.626	0.520
	[256, 128, 128, 64, 32, 16]	[0.01, ...]	True	mae	Adam (learning_rate=0.002)	linear	[mae, r2_metric]	32	300	100	1	-0.105	-0.104

doi:10.6342/NTU202502436

表 A.7 對應圖 4.7 之 ANN 各超參數設定-2

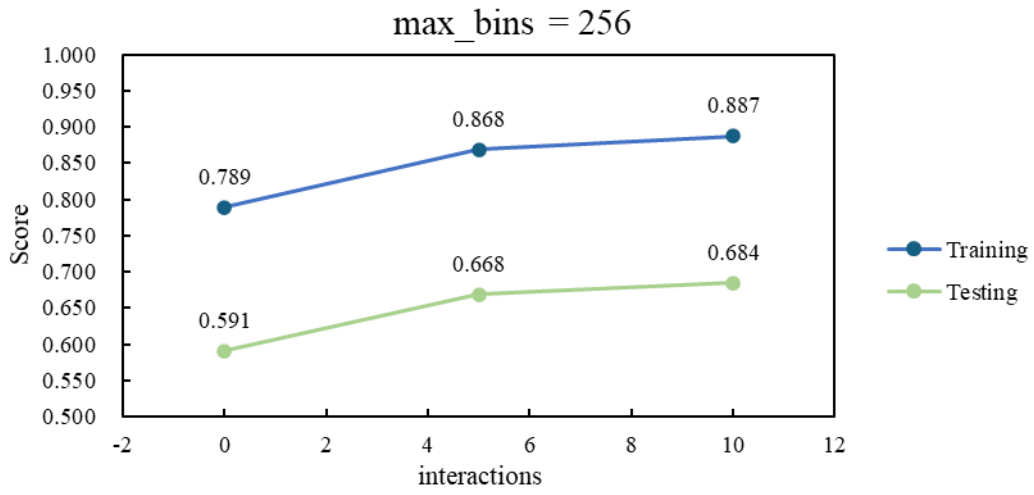
圖 4.7	units	num_layers	activation	BatchNormalization	Dropout	optimizer	learning_rate	loss
	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse
(d)	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse
	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse
	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse
(e)	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse
	32~256	3~5	[relu,tanh,selu]	True	0.1~0.3	Adam	0.00005 ~ 0.001（log scale）	mse

	metrics	objective	factor	max_epochs	directory	project_name	Training csore	Testing score
	['mae']	'val_mae'	3	100	'tuner_dir'	'csr_ann_script'	0.307	0.175
(d)	['mae']	'val_mae'	4	100	'tuner_dir'	'csr_ann_script'	0.379	0.079
	['mae']	'val_mae'	5	100	'tuner_dir'	'csr_ann_script'	0.322	0.162
	['mae']	'val_mae'	3	100	'tuner_dir'	'csr_ann_script'	0.307	0.175
(e)	['mae']	'val_mae'	3	300	'tuner_dir'	'csr_ann_script'	0.405	0.057
	['mae']	'val_mae'	3	500	'tuner_dir'	'csr_ann_script'	0.249	0.108

doi:10.6342/NTU202502436

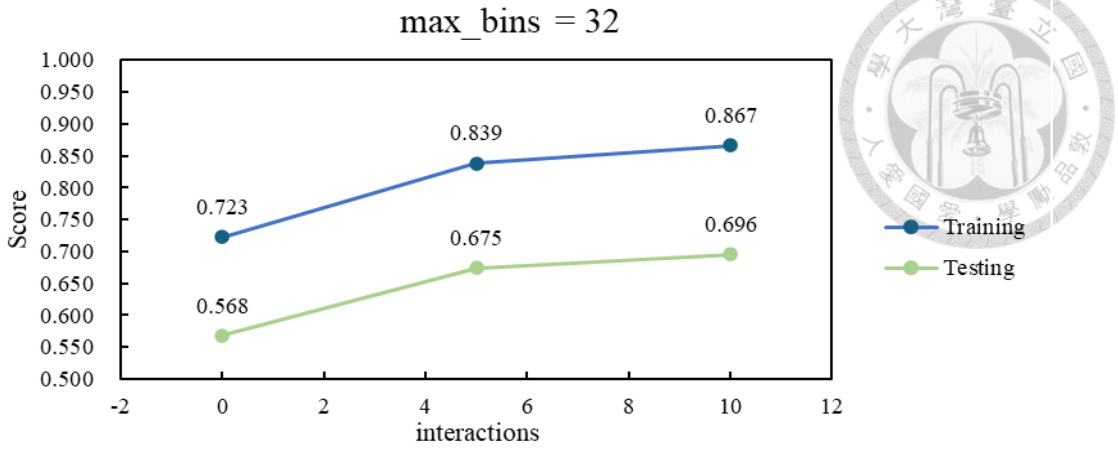
表 A.8 EBM 模型使用所有特徵並填以 NaN 之各參數設定

圖 A.1	(a)			(b)			(c)			(d)		
max_bins	256	256	256	32	32	32	128	128	128	256	256	256
max_interaction_bins	32	32	32	32	32	32	32	32	32	32	32	32
interactions	0	5	10	0	5	10	5	10	20	5	10	20
learning_rate	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
max_rounds	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
early_stopping_rounds	50	50	50	50	50	50	50	50	50	50	50	50
min_samples_leaf	20	20	20	20	20	20	20	20	20	20	20	20
max_leaves	3	3	3	3	3	3	3	3	3	3	3	3
Training score	0.7887	0.8684	0.887	0.7227	0.8385	0.8666	0.8661	0.8752	0.8957	0.8684	0.887	0.9003
Testing score	0.5908	0.6684	0.6842	0.568	0.6746	0.6959	0.6639	0.6744	0.6911	0.6684	0.6842	0.6963

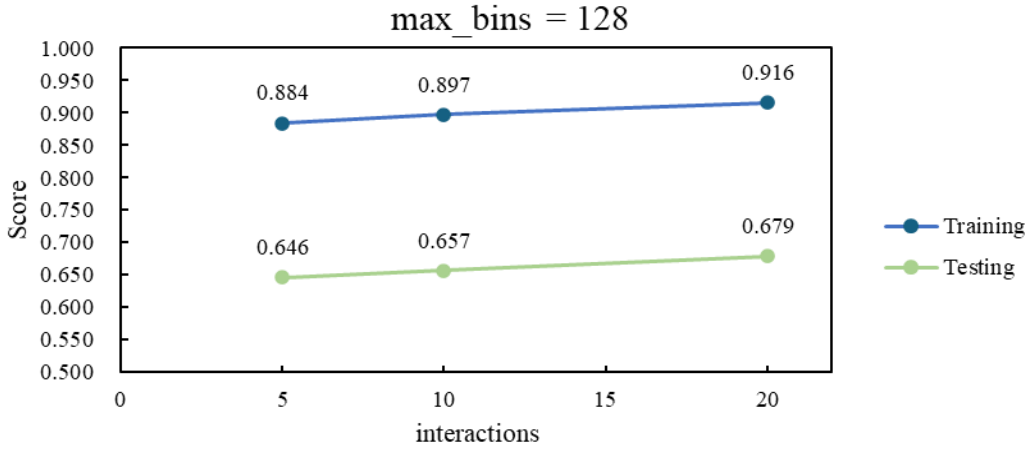


(a)

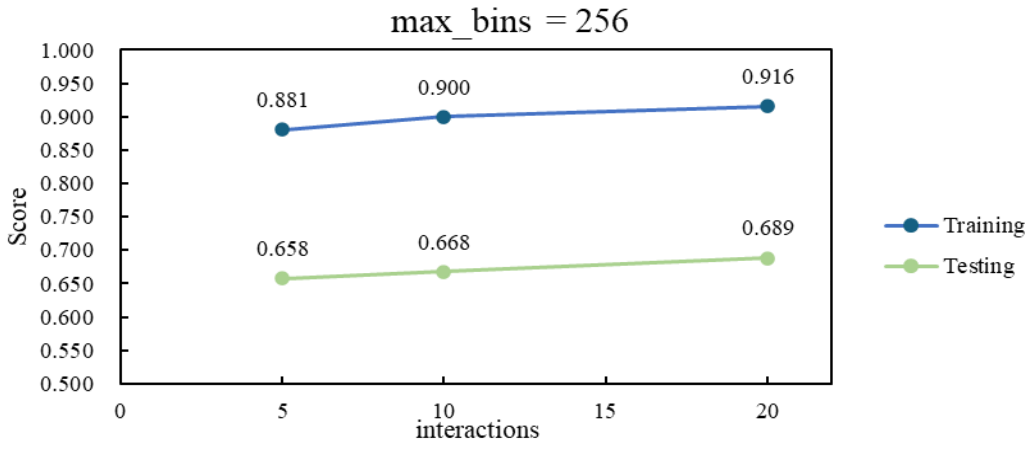
doi:10.6342/NTU202502436



(b)



(c)



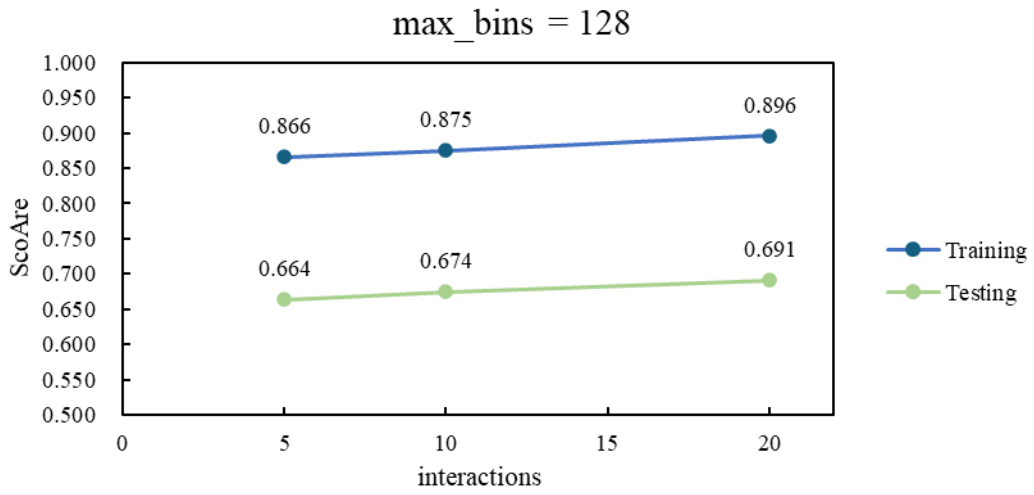
(d)

圖 A.1 EBM 模型不同 interactions 設定之結果比較 (所有特徵，NaN 填補) (a) 於 max_bins = 256 下 (b) 於 max_bins = 32 下 (c) 於 max_bins = 128 下 (d) 於 max_bins = 256 下

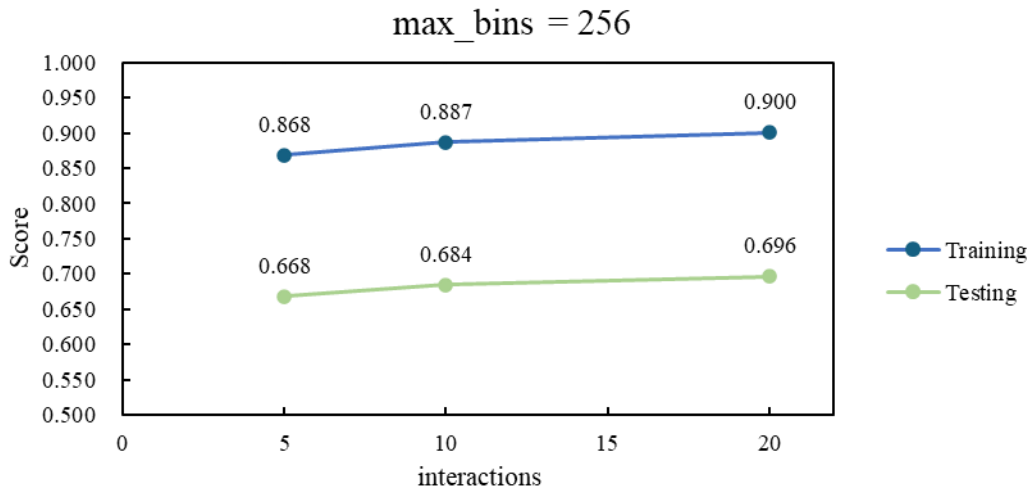
doi:10.6342/NTU202502436

表 A.9 EBM 模型使用所有特徵並填以 KNN 填補之各參數設定

圖 A.2	(a)			(b)		
max_bins	128	128	128	256	256	256
max_interaction_bins	32	32	32	32	32	32
interactions	5	10	20	5	10	20
learning_rate	0.01	0.01	0.01	0.01	0.01	0.01
max_rounds	5000	5000	5000	5000	5000	5000
early_stopping_rounds	50	50	50	50	50	50
min_samples_leaf	20	20	20	20	20	20
max_leaves	3	3	3	3	3	3
Training score	0.8842	0.8969	0.9161	0.8809	0.8996	0.9156
Testing score	0.6458	0.6568	0.6785	0.6581	0.6675	0.6886



(a)



(b)

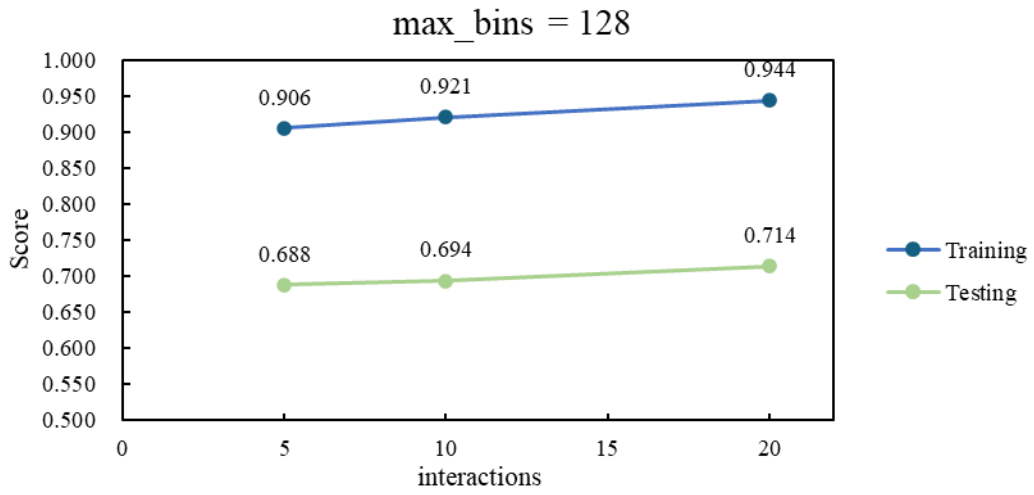
圖 A.2 EBM 模型不同 interactions 設定之結果比較 (所有特徵，KNN 填補)

(a) 於 max_bins = 128 下 (b) 於 max_bins = 256 下

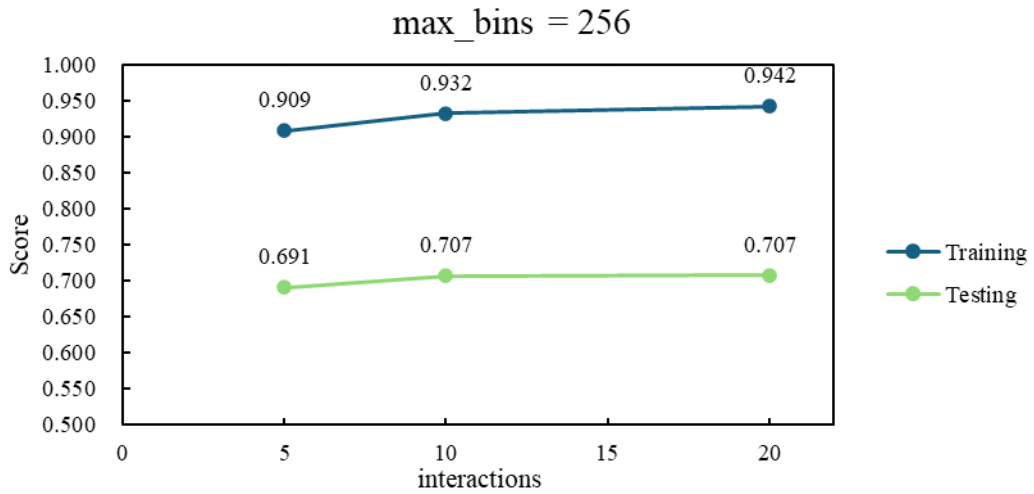
doi:10.6342/NTU202502436

表 A.10 EBM 模型使用所有特徵並填以 MICE 填補之各參數設定

圖 A.3	(a)			(b)		
max_bins	128	128	128	256	256	256
max_interaction_bins	32	32	32	32	32	32
interactions	5	10	20	5	10	20
learning_rate	0.01	0.01	0.01	0.01	0.01	0.01
max_rounds	5000	5000	5000	5000	5000	5000
early_stopping_rounds	50	50	50	50	50	50
min_samples_leaf	20	20	20	20	20	20
max_leaves	3	3	3	3	3	3
Training score	0.9056	0.9207	0.9439	0.9086	0.9319	0.9419
Testing score	0.6877	0.6936	0.714	0.6905	0.7069	0.7072



(a)



(b)

圖 A.3 EBM 模型不同 interactions 設定之結果比較 (所有特徵，MICE 填補)

(a) 於 max_bins = 128 下 (b) 於 max_bins = 256 下

doi:10.6342/NTU202502436

表 A.11 EBM 模型各特徵選擇及填補方式下之 GridSearch 設定

圖 A.4

Approach	Primary + NaN + K-Fold	All + NaN + K-Fold	All + KNN + K-Fold	All + MICE + K-Fold
max_bins	256	256	256	256
max_interaction_bins	30	30	30	30
interactions	10	10	10	10
learning_rate	0.01	0.01	0.01	0.01
max_rounds	5000	5000	5000	5000
early_stopping_rounds	50	50	50	50
min_samples_leaf	20	20	30	20
max_leaves	8	5	5	8
Training score	0.8662	0.8677	0.8681	0.9268
Testing score	0.6859	0.6886	0.6738	0.693

doi:10.6342/NTU202502436

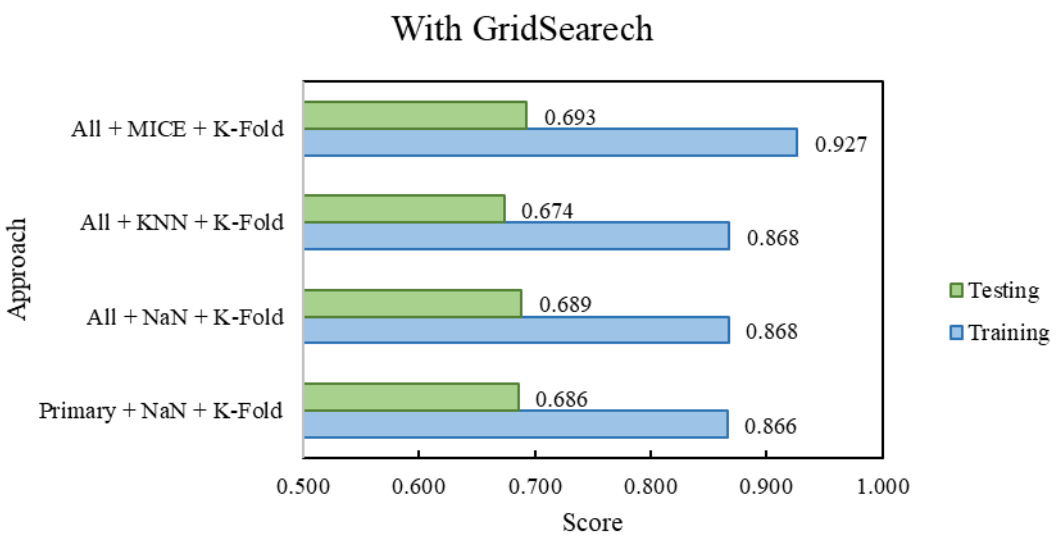


圖 A.4 EBM 模型各特徵選擇及填補方式下之 GridSearch 結果比較

doi:10.6342/NTU202502436

附錄 B 最佳模型參數設定總覽

表 B.1 各機器學習模型於不同條件下之 R^2、MAE、MSE、RMSE 分數表

Model	Features involved	Imputation	Parameter selection	R^2 score		MAE		MSE		RMSE		Time (s)	File name
				Training	Testing	Training	Testing	Training	Testing	Training	Testing		
Random Forest	Primary	NaN	Manual tuning	0.819	0.689	0.035	0.055	0.002	0.006	0.047	0.074	0.4	0311_4
CatBoost				0.898	0.688	0.025	0.048	0.001	0.005	0.037	0.068	958.0	0319_12
XGBoost				0.967	0.733	0.014	0.044	0.000	0.004	0.021	0.063	0.0	0325_12
EBM		———		0.912	0.681	0.023	0.048	0.001	0.005	0.034	0.069	36.8	0328_1
SVM_rbf				0.908	0.755	0.021	0.041	0.001	0.004	0.035	0.061	0.0	0401_4
SVM_poly		"0"		0.892	0.711	0.023	0.043	0.001	0.004	0.038	0.066	7.8	0402_6
ANN				0.745	0.642	0.038	0.054	0.003	0.005	0.058	0.073	32.8	0408_7
BNN				0.918	0.751	0.021	0.039	0.001	0.004	0.033	0.061	125.1	0419_5
Random Forest	Primary	NaN	Grid search tuning	0.949	0.799	0.017	0.043	0.001	0.004	0.025	0.060	33.0	0311_6
CatBoost				0.885	0.886	0.027	0.026	0.002	0.001	0.040	0.036	3240.0	0320
XGBoost				0.925	0.721	0.023	0.046	0.001	0.004	0.031	0.065	1856.4	0325_11
EBM		———		0.866	0.686	0.028	0.051	0.002	0.005	0.042	0.069	1437.6	0329
SVM_rbf				0.862	0.731	0.027	0.043	0.002	0.004	0.043	0.064	0.3	0401_6
SVM_poly		"0"		0.419	0.336	0.067	0.077	0.008	0.010	0.088	0.100	4.1	0403_3
ANN				0.702	0.502	0.043	0.060	0.004	0.008	0.063	0.086	1.3	0409
BNN				0.780	0.698	0.033	0.046	0.003	0.005	0.054	0.067	10732.4	0420

doi:10.6342/NTU202502436

Model	Features involved	Imputation	Parameter selection	R^2 score		MAE		MSE		RMSE		Time (s)	File name
				Training	Testing	Training	Testing	Training	Testing	Training	Testing		
Random Forest	All	NaN	Manual tuning	0.880	0.622	0.015	0.046	0.001	0.005	0.023	0.069	0.4	0313
CatBoost				0.934	0.702	0.018	0.045	0.001	0.005	0.030	0.067	1480.9	0320_2
XGBoost				0.983	0.777	0.010	0.039	0.000	0.003	0.015	0.058	0.8	0327_19
EBM		———		0.900	0.696	0.022	0.048	0.001	0.005	0.036	0.067	11.1	0329_27
SVM_rbf				0.942	0.657	0.016	0.042	0.001	0.005	0.028	0.072	0.8	0428_15
SVM_poly		"0"		0.929	0.731	0.018	0.042	0.001	0.004	0.031	0.064	6.6	0428_14
ANN				0.555	0.253	0.051	0.073	0.006	0.011	0.077	0.106	26.3	0409_12
BNN				0.924	0.782	0.020	0.038	0.001	0.003	0.032	0.057	62.8	0420_34
Random Forest	All	NaN	Grid search tuning	0.961	0.690	0.015	0.046	0.001	0.005	0.023	0.068	180.0	0314_2
CatBoost				0.901	0.899	0.024	0.024	0.001	0.001	0.037	0.034	22.3	0321
XGBoost				0.951	0.754	0.015	0.040	0.001	0.004	0.026	0.061	2.9	0326_12
EBM		———		0.868	0.689	0.026	0.048	0.002	0.005	0.042	0.068	2212.6	0329_7
SVM_rbf				0.912	0.766	0.017	0.040	0.001	0.004	0.034	0.059	4.1	0424
SVM_poly		"0"		0.765	0.700	0.032	0.045	0.003	0.005	0.056	0.067	34.5	0424_1
ANN				0.502	0.427	0.048	0.063	0.007	0.009	0.081	0.093	9.9	0410
BNN				0.756	0.640	0.033	0.050	0.003	0.005	0.057	0.073	162.0	0420_41

doi:10.6342/NTU202502436

Model	Features involved	Imputation	Parameter selection	R^2 score		MAE		MSE		RMSE		Time (s)	File name
				Training	Testing	Training	Testing	Training	Testing	Training	Testing		
Random Forest	All	KNN	Manual tuning	0.911	0.597	0.022	0.053	0.001	0.006	0.034	0.078	1.5	0315_3
CatBoost				0.927	0.648	0.020	0.050	0.001	0.005	0.031	0.073	65.8	0321_3
XGBoost				0.983	0.710	0.009	0.044	0.000	0.004	0.015	0.066	0.8	0327_2
EBM				0.916	0.689	0.020	0.049	0.001	0.005	0.033	0.068	16.8	0329_13
SVM_rbf				0.510	0.314	0.049	0.070	0.007	0.010	0.080	0.101	0.1	0402_13
SVM_poly				0.595	0.417	0.055	0.068	0.005	0.009	0.073	0.093	20.6	0404_23
ANN				0.630	0.457	0.045	0.063	0.005	0.008	0.069	0.090	34.6	0410_11
BNN				0.702	0.251	0.036	0.067	0.004	0.011	0.063	0.106	245.7	0421_2
Random Forest	All	KNN	Grid search tuning	0.954	0.678	0.016	0.049	0.001	0.005	0.025	0.069	292.5	0317_7
CatBoost				0.904	0.907	0.024	0.021	0.001	0.001	0.037	0.033	386.7	0321_6
XGBoost				0.947	0.684	0.015	0.047	0.001	0.005	0.026	0.069	3.3	0327_6
EBM				0.868	0.674	0.025	0.050	0.002	0.005	0.042	0.070	2529.1	0329_14
SVM_rbf				0.419	0.336	0.067	0.077	0.008	0.010	0.088	0.100	0.3	0403_1
SVM_poly				0.419	0.336	0.067	0.077	0.008	0.010	0.088	0.100	4.1	0403_3
ANN				0.435	0.104	0.062	0.080	0.007	0.013	0.086	0.116	10.9	0411_2
BNN				0.622	0.127	0.038	0.070	0.005	0.013	0.071	0.114	11893.6	0420_15

doi:10.6342/NTU202502436

Model	Features involved	Imputation	Parameter selection	R^2 score		MAE		MSE		RMSE		Time (s)	File name
				Training	Testing	Training	Testing	Training	Testing	Training	Testing		
Random Forest	All	MICE	Manual tuning	0.931	0.708	0.019	0.046	0.001	0.004	0.030	0.066	1.6	0428_1
CatBoost				0.927	0.715	0.019	0.043	0.001	0.004	0.031	0.065	74.3	0321_8
XGBoost				0.987	0.757	0.008	0.038	0.000	0.004	0.013	0.060	1.0	0327_10
EBM				0.944	0.714	0.022	0.046	0.001	0.005	0.035	0.068	9.5	0329_18
SVM_rbf				0.997	0.302	0.002	0.072	0.000	0.010	0.006	0.102	0.0	0403_8
SVM_poly				0.867	0.769	0.027	0.042	0.002	0.004	0.042	0.059	4.0	0415_16
ANN				0.852	0.624	0.027	0.051	0.002	0.006	0.044	0.075	26.3	0411_13
BNN				0.864	0.678	0.026	0.048	0.002	0.005	0.042	0.069	150.1	0420_19
Random Forest	All	MICE	Grid search tuning	0.967	0.724	0.014	0.044	0.000	0.004	0.021	0.064	387.3	0317_11
CatBoost				0.912	0.897	0.022	0.022	0.001	0.001	0.035	0.034	409.9	0321_12
XGBoost				0.952	0.736	0.013	0.039	0.001	0.004	0.025	0.063	3.1	0327_14
EBM				0.927	0.693	0.018	0.044	0.001	0.005	0.031	0.068	2093.5	0329_21
SVM_rbf				0.991	0.352	0.003	0.070	0.000	0.010	0.011	0.099	3.2	0403_5
SVM_poly				0.415	0.369	0.058	0.068	0.008	0.009	0.088	0.097	79.3	0404_27
ANN				0.369	0.175	0.065	0.079	0.008	0.012	0.091	0.111	1272.0	0412
BNN				0.617	0.118	0.037	0.070	0.005	0.013	0.071	0.115	8948.1	0420_28

doi:10.6342/NTU202502436

表 B.2 各機器學習模型於不同條件下之最佳超參數組合表-1

Random Froest	(a)	(b)	(c)	(d)
n_estimators	500	500	500	500
max_depth	10	None	10	20
min_samples_split	5	2	5	2
min_samples_leaf	1	1	1	1
max_features	squrt	None	squrt	None
random_state	42	42	42	42
cv	5	5	5	10
file name	0311_4	0311_6	0313	0314_2
	(e)	(f)	(g)	(h)
n_estimators	450	350	450	550
max_depth	None	None	None	20
min_samples_split	5	2	5	2
min_samples_leaf	2	1	2	1
max_features	None	None	None	None
random_state	42	42	42	42
cv	5	10	5	10
file name	0315_3	0317_7	0428_1	0317_11

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

Cat Boost	(a)	(b)	(c)	(d)
iterations	500	1000	1000	500
depth	10	16	16	10
learning_rate	0.1	0.1	0.1	0.1
l2_leaf_reg	3	1	3	1
cat_features	categorical_features	categorical_features	categorical_features	categorical_features
verbose	100	100	100	100
random_strength	5	5	5	5
border_count	200	255	200	200
task_type	"GPU"	GPU	GPU	GPU
file name	0319_12	0320	0320_2	0321
	(e)	(f)	(g)	(h)
iterations	500	500	500	500
depth	12	12	12	12
learning_rate	0.1	0.1	0.1	0.1
l2_leaf_reg	3	3	3	3
cat_features	categorical_features	categorical_features	categorical_features	categorical_features
verbose	100	100	100	100
random_strength	5	5	5	5
border_count	200	200	200	200
task_type	GPU	GPU	GPU	GPU
file name	0321_3	0321_6	0321_8	0321_12

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

XGBoost	(a)	(b)	(c)	(d)
objective	'reg:squarederror'	'reg:squarederror'	'reg:squarederror'	'reg:squarederror'
learning_rate	0.05	0.05	0.1	0.05
n_estimators	1500	1500	1200	1500
max_depth	5	5	5	5
min_child_weight	5	5	5	5
gamma	0	0	0	0
reg_alpha	0.1	0.1	0.1	0.1
reg_lambda	3	3	3	3
subsample	0.8	0.8	0.8	0.8
colsample_bytree	0.8	0.8	0.8	0.8
random_state	42	42	42	42
enable_categorical	True	True	True	True
file name	0325_12	0325_11	0327_19	0326_12
	(e)	(f)	(g)	(h)
objective	'reg:squarederror'	'reg:squarederror'	'reg:squarederror'	'reg:squarederror'
learning_rate	0.05	0.05	0.05	0.05
n_estimators	1500	1500	1500	1500
max_depth	5	5	5	5
min_child_weight	5	5	5	5
gamma	0	0	0	0
reg_alpha	0.1	0.1	0.1	0.1
reg_lambda	3	3	3	3
subsample	0.8	0.8	0.8	0.8
colsample_bytree	0.8	0.8	0.8	0.8
random_state	42	42	42	42
enable_categorical	True	True	True	True
file name	0327_2	0327_6	0327_10	0327_14

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

EBM	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
max_bins	512	256	256	256	256	256	256	256
max_interaction_bins	32	30	32	30	32	30	32	30
interactions	20	10	20	10	20	10	5	10
learning_rate	0.005	0.01	0.01	0.01	0.01	0.01	0.01	0.01
max_rounds	8000	5000	5000	5000	5000	5000	5000	5000
early_stopping_rounds	200	50	50	50	50	50	50	50
min_samples_leaf	2	20	20	20	20	30	20	20
max_leaves	5	8	8	5	8	5	8	8
random_state	42	42	42	42	42	42	42	42
file name	0328_1	0329	0329_27	0329_7	0329_13	0329_14	0329_18	0329_21

SVM_rbf	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
kernel	rbf	rbf	rbf	rbf	rbf	rbf	rbf	rbf
c	100	10	1000	1000	100	10	10	1
epsilon	0.01	0.01	0.01	0.01	0.01	0.1	0.001	0.001
gamma	scale	scale	scale	scale	scale	scale	0.01	0.01
file name	0401_4	0401_6	0428_15	0424	0402_13	0403_1	0403_8	0403_5

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

SVM_poly	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
kernel	poly	poly	poly	poly	poly	poly	poly	poly
c	100	100	1000	10	10	10	10	1
epsilon	0.01	0.01	0.01	0.1	0.1	0.1	0.001	0.01
gamma	0.1	auto	scale	scale	scale	scale	0.01	scale
degree	4	5	3	3	4	3	3	4
coef0	1	0	0	0	10	0	0	1
file name	0402_6	0402_10	0428_14	0424_1	0404_23	0403_3	0415_16	0404_27

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

表 B.3 各機器學習模型於不同條件下之最佳超參數組合表-2

Model / Hyperparameter	(a)	(b)	(c)	(d)
ANN				
Sequential				
hidden_layers/units	[256, 128, 128, 64, 64, 32, 16]	32~256	[256, 128, 128, 64, 64, 32, 16]	32~256
activations	sigmoid	['relu', 'tanh']	sigmoid	['relu', 'tanh', 'selu'], 3~5 layers
dropout_rates	0.01	0.1, 0.3, step=0.1	0.01	0.1, 0.3, step=0.1
learning_rate	0.001	0.00005 ~ 0.001（log scale）	0.001	0.00005 ~ 0.001（log scale）
use_batch_norm	True	True	True	True
optimizer	Adam	Adam	Adam	Adam
loss_fn	mae	mse	mae	mae
metrics	['mae', r2_metric]	['mae', r2_metric]	['mae', r2_metric]	['mae', r2_metric]
batch_size	32	32	128	32
epochs	100	300	300	300
history				
validation_split	0.1	0.1	0.1	0.1
batch_size	32	32	128	32
epochs	300	300	300	300
verbose	1	1	1	1
file name	0408_7	0409	0409_12	0410
BNN				
hidden_layers	[256,128,128,64,32,16]	[256,128,128,64,32,16]	[256,128,128,64,32,16]	[256,128,64,32,16,8]
use_dropout	True	FALSE	True	FALSE
activation_fn	GELU	relu	GELU	sigmoid
learning_rate	0.001	0.001	0.001	0.001
optimizer_type	torch.optim.Adam	torch.optim.Adam	torch.optim.Adam	torch.optim.Adam
dropout_rate	0.1	0.1	0.1	0.1
input_dim	X_train.shape[1]	X_train.shape[1]	X_train.shape[1]	X_train.shape[1]
loss_fn	L1Loss()	L1Loss()	L1Loss()	L1Loss()
kl_weight	1.0 / len(X_train)	1.0 / len(X_train)	1.0 / len(X_train)	1.0 / len(X_train)
epochs	500	1000	500	1000
file name	0419_5	0420	0420_34	0420_41

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

Model / Hyperparameter	(e)	(f)	(g)	(h)
ANN				
Sequential				
hidden_layers/units	[256, 128, 128, 64, 64, 32, 16]	32~256	[256, 128, 128, 64, 64, 32, 16]	32~256
activations	sigmoid	['relu', 'tanh', 'selu'], 3~5 layers	sigmoid	['relu', 'tanh', 'selu'], 3~5 layers
dropout_rates	0.01	0.1, 0.3, step=0.1	0.01	0.1, 0.3, step=0.1
learning_rate	0.001	0.00005 ~ 0.001（log scale）	0.001	0.00005 ~ 0.001（log scale）
use_batch_norm	True	True	FALSE	True
optimizer	Adam	RMSprop	Adam	RMSprop
loss_fn	mae	mse	mae	mse
metrics	['mae', r2_metric]	['mae', r2_metric]	['mae', r2_metric]	['mae', r2_metric]
batch_size	32	32	32	32
epochs	300	300	300	300
history				
validation_split	0.1	0.1	0.1	0.1
batch_size	32	32	32	32
epochs	300	300	300	300
verbose	1	1	1	1
file name	0410_11	0411_2	0411_13	0412
BNN				
hidden_layers	[256,128,128,64,32,16]	[256,128,64,32,16,8]	[256,128,128,64,32,16]	[256,128,64,32,16,8]
use_dropout	True	FALSE	True	FALSE
activation_fn	LeakyReLU(0.01)	sigmoid	Softplus	sigmoid
learning_rate	0.001	0.001	0.001	0.001
optimizer_type	torch.optim.RMSprop	torch.optim.Adam	torch.optim.Adam	torch.optim.Adam
dropout_rate	0.1	0.1	0.1	0.1
input_dim	X_train.shape[1]	X_train.shape[1]	X_train.shape[1]	X_train.shape[1]
loss_fn	L1Loss()	L1Loss()	L1Loss()	L1Loss()
kl_weight	1.0 / len(X_train)	1.0 / len(X_train)	1.0 / len(X_train)	1.0 / len(X_train)
epochs	1000	1000	500	1000
file name	0421_2	0420_15	0420_19	0420_28

註: (a) ~ (h) 分別對照圖 4.8 至圖 4.11 之標號設定

doi:10.6342/NTU202502436

附錄 C 未填補資料下手動調參模型之 SHAP 圖

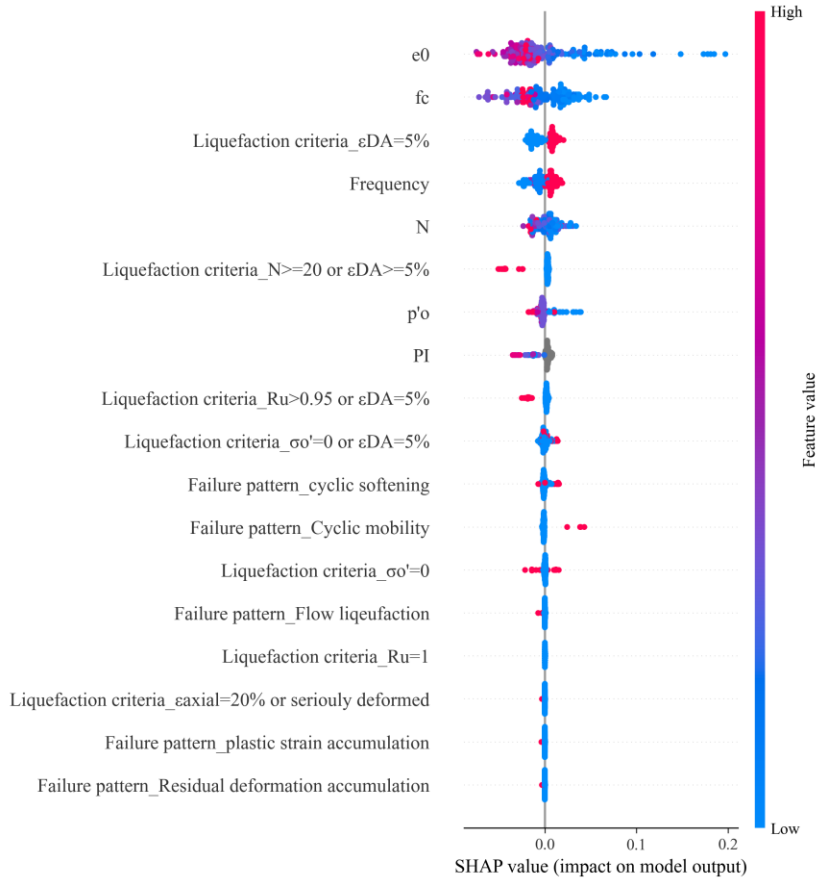


圖 C.1 包含主要特徵之 Random Forest 模型 SHAP 總結圖

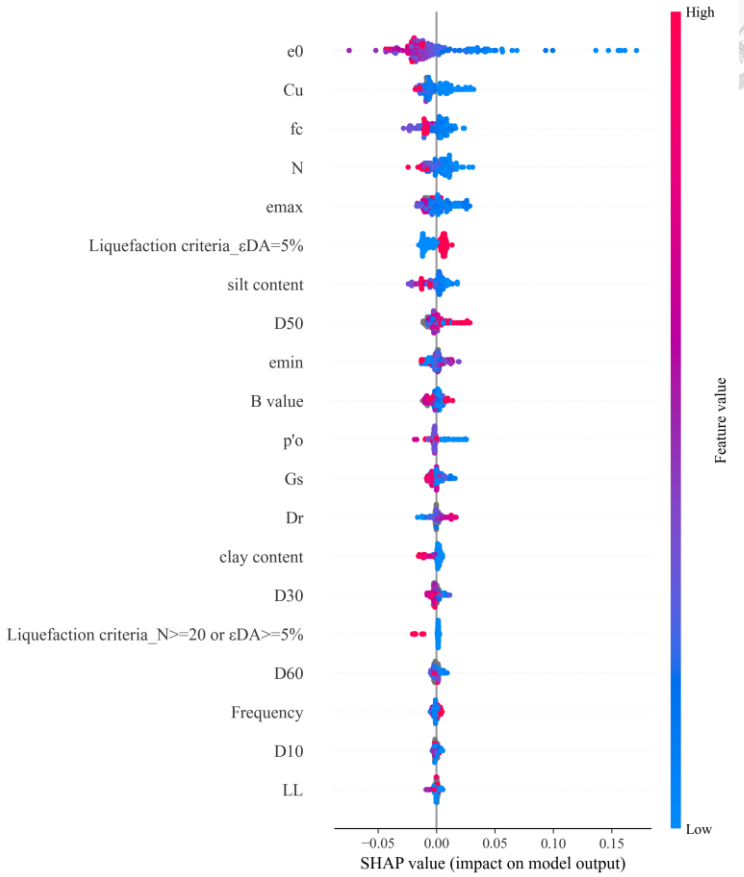


圖 C.2 包含所有特徵之 Random Forest 模型 SHAP 總結圖

doi:10.6342/NTU202502436

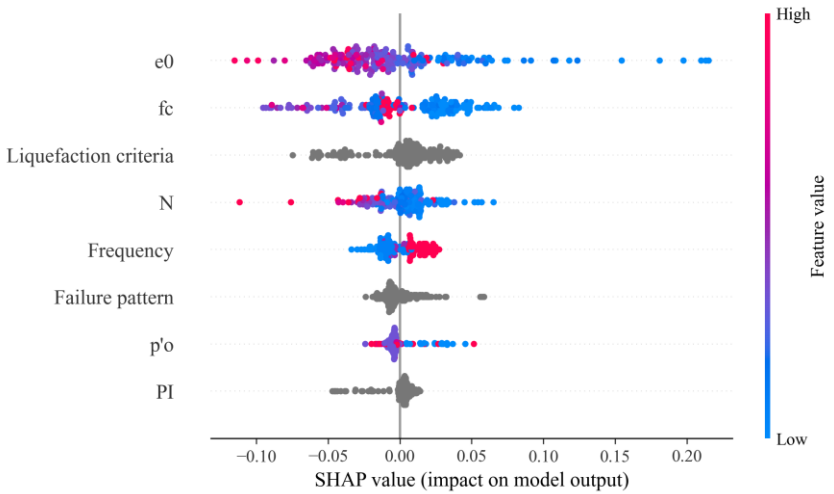


圖 C.3 包含主要特徵之 CatBoost 模型 SHAP 總結圖

註: CatBoost 可直接將類別型特徵輸入至模型，不需做獨熱編碼。

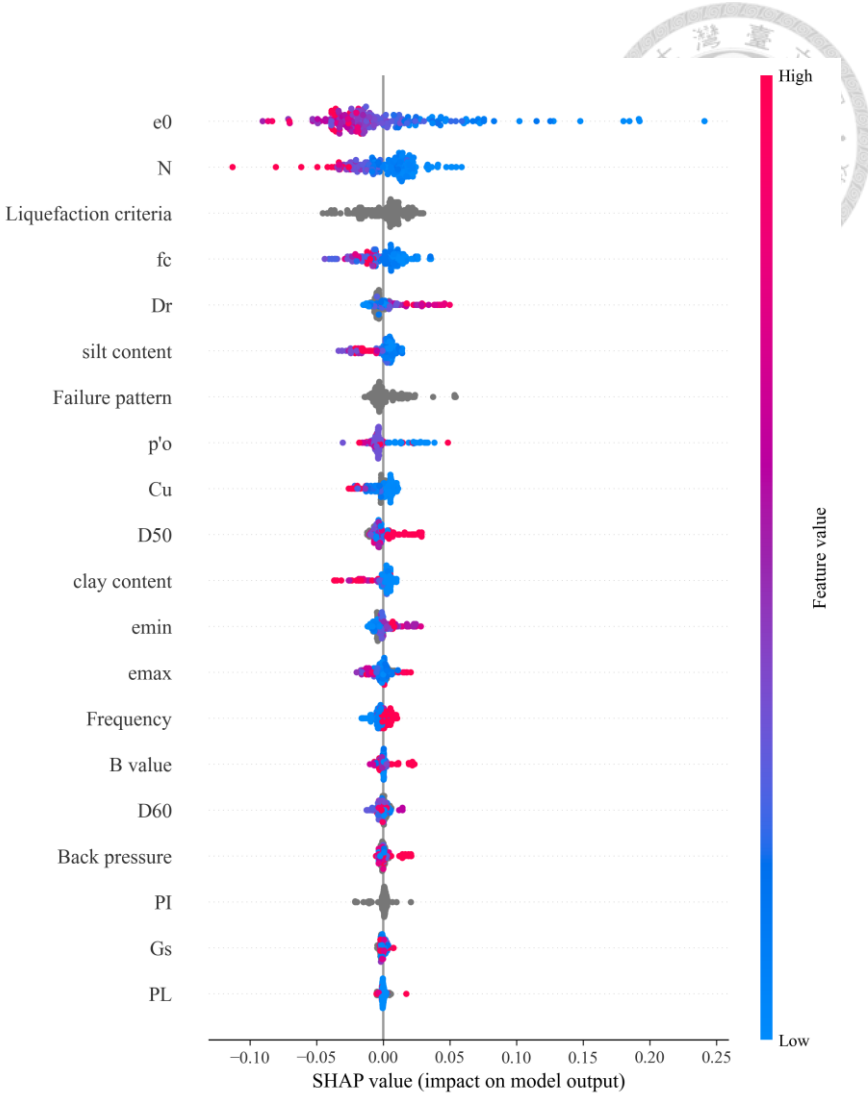


圖 C.4 包含所有特徵之 CatBoost 模型 SHAP 總結圖

doi:10.6342/NTU202502436

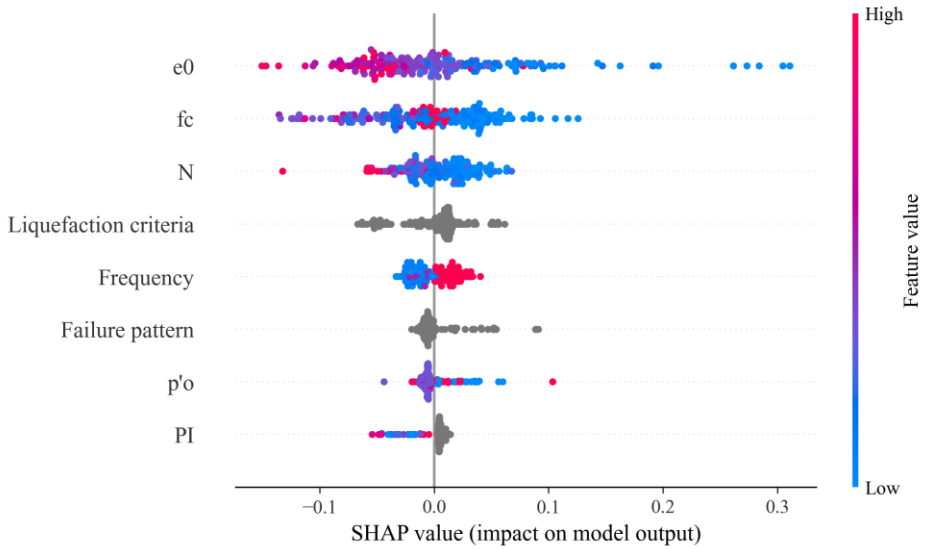


圖 C.5 包含主要特徵之 XGBoost 模型 SHAP 總結圖

註: 代碼中的 `enable_categorical=True` 可將獨熱編碼分出去的特徵做合併。

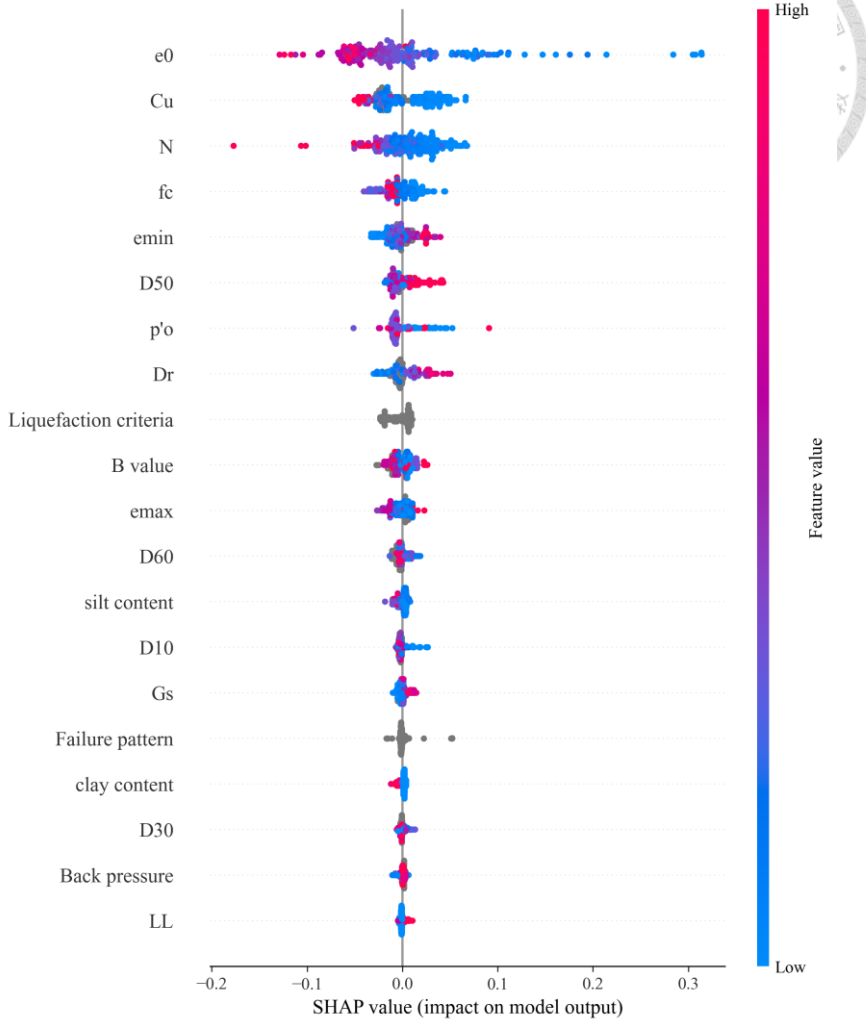


圖 C.6 包含所有特徵之 XGBoost 模型 SHAP 總結圖

doi:10.6342/NTU202502436

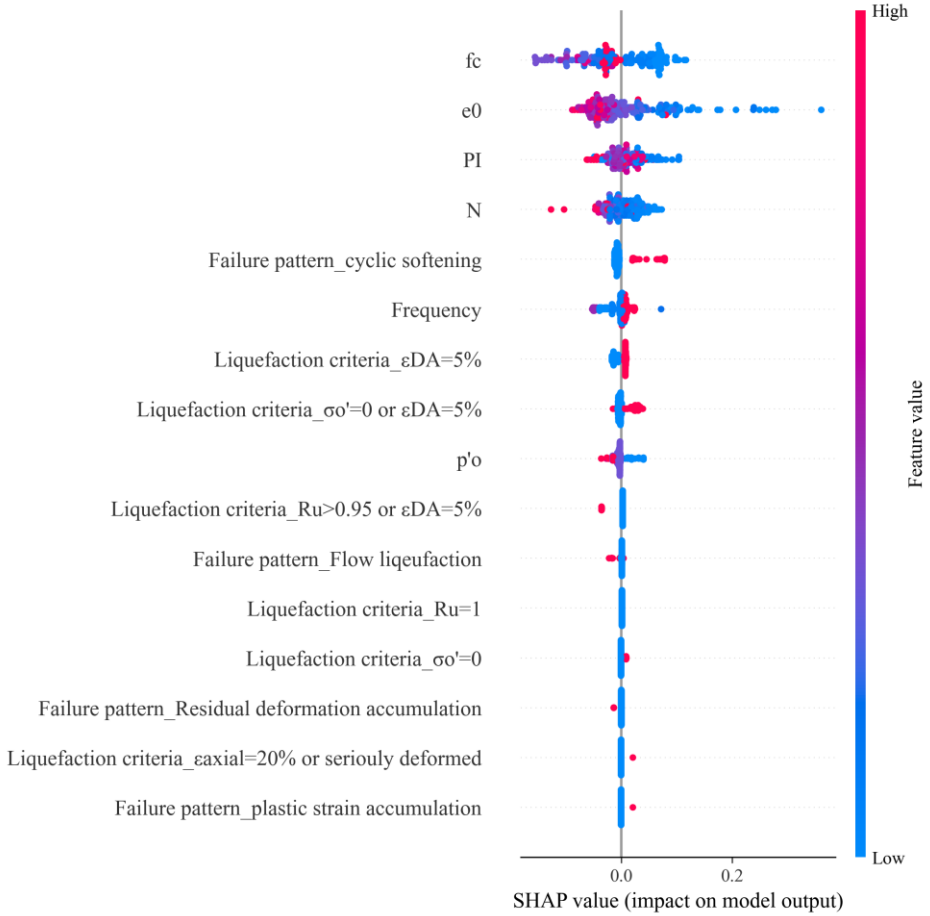


圖 C.7 包含主要特徵之 EBM 模型 SHAP 總結圖

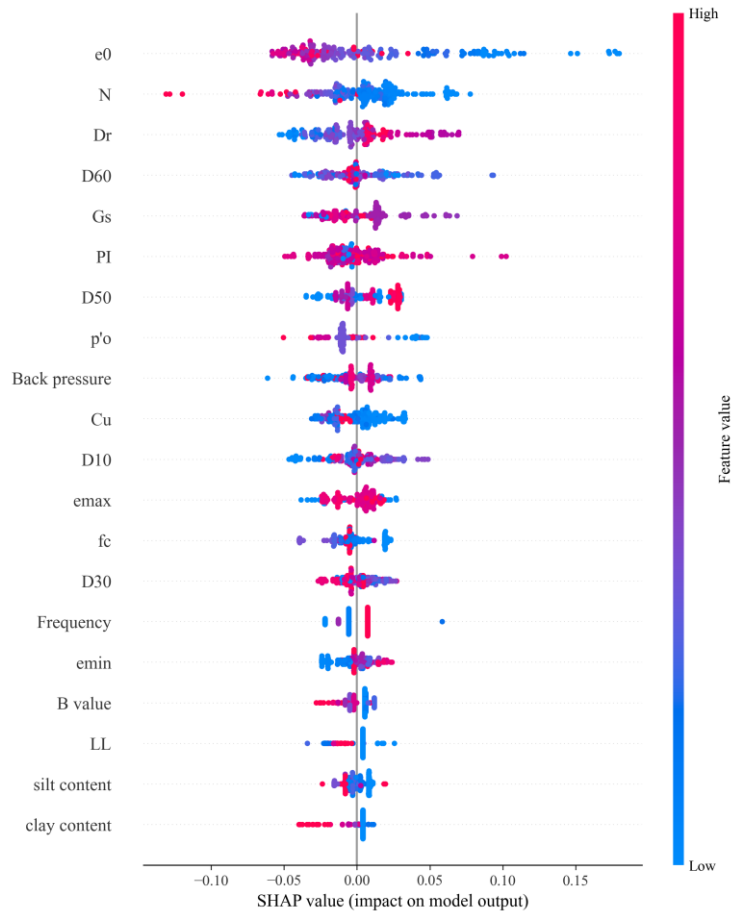


圖 C.8 包含所有特徵之 EBM 模型 SHAP 總結圖

doi:10.6342/NTU202502436

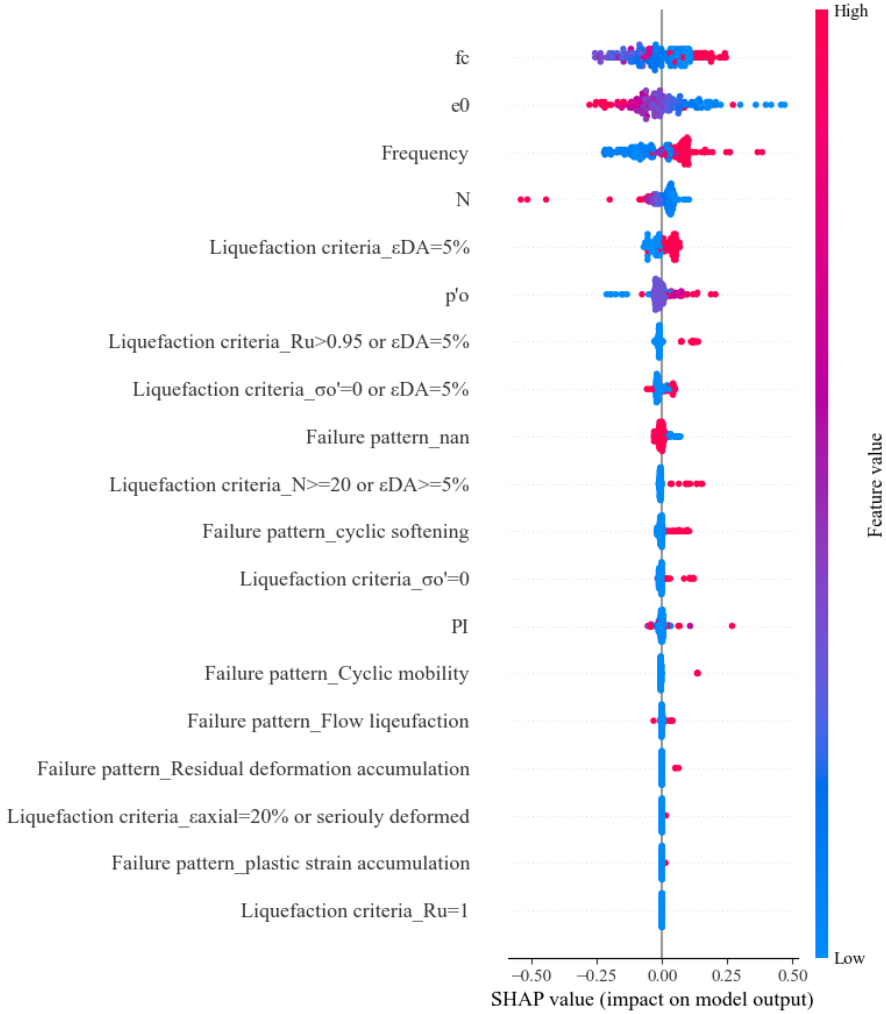


圖 C.9 包含主要特徵之 SVM (RBF) 模型 SHAP 總結圖

圖 C.10 包含所有特徵之 SVM (RBF) 模型 SHAP 總結圖

doi:10.6342/NTU202502436

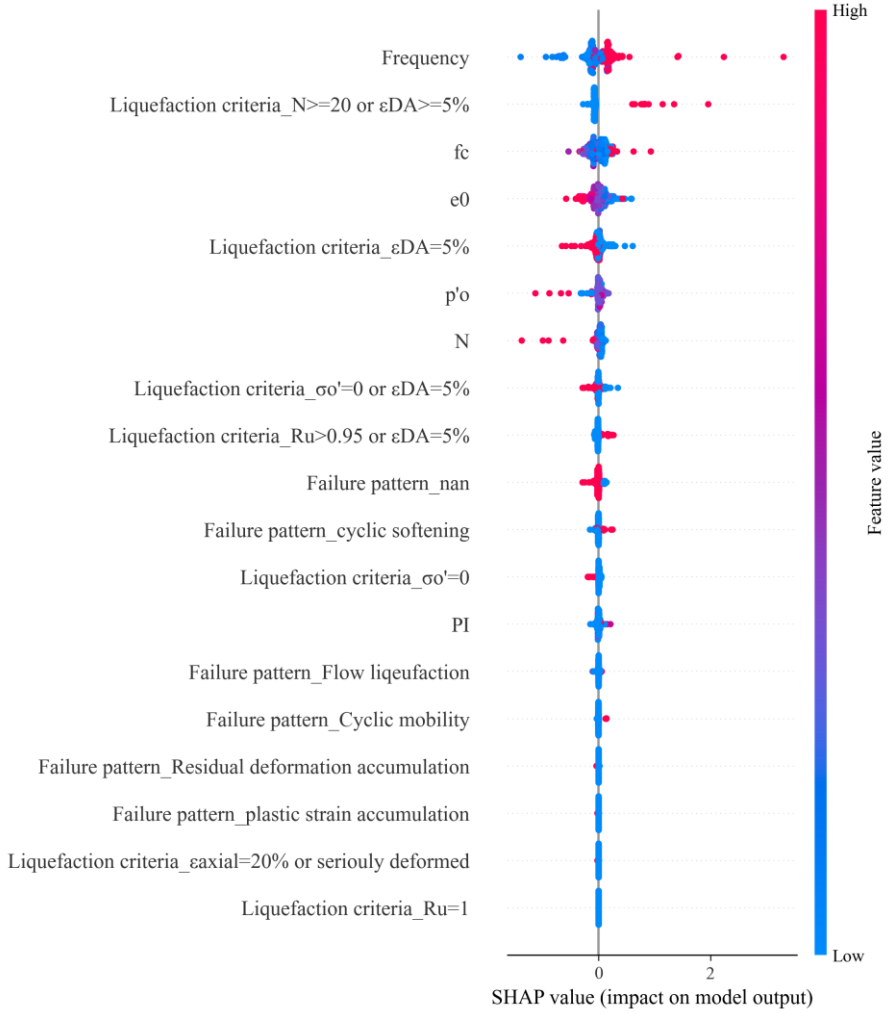


圖 C.11 包含主要特徵之 SVM (poly) 模型 SHAP 總結圖

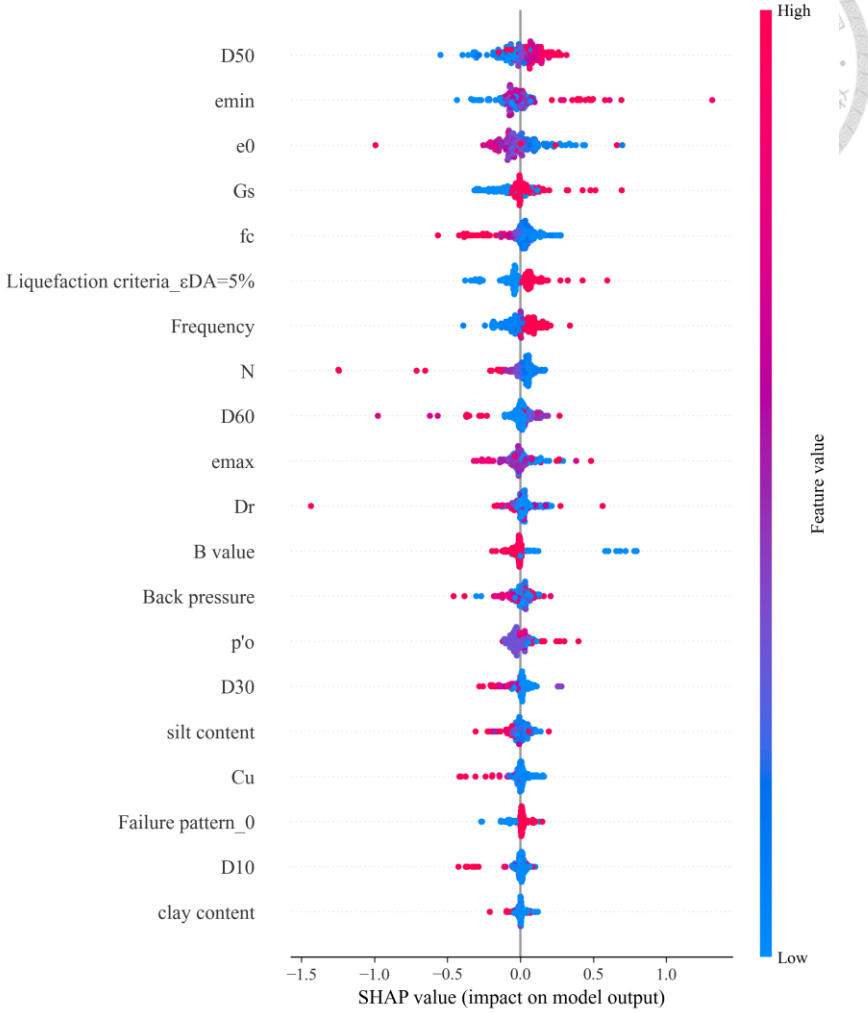


圖 C.12 包含所有特徵之 SVM (poly) 模型 SHAP 總結圖

doi:10.6342/NTU202502436

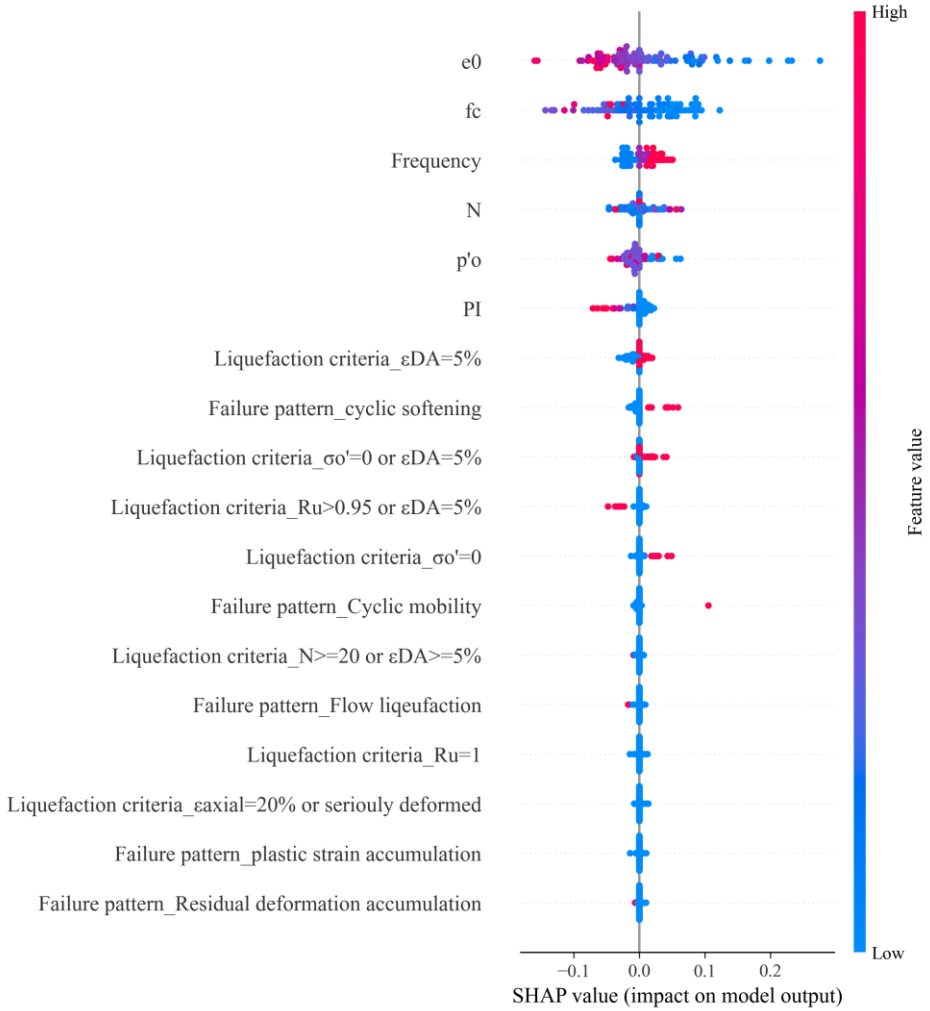


圖 C.13 包含主要特徵之 ANN 模型 SHAP 總結圖

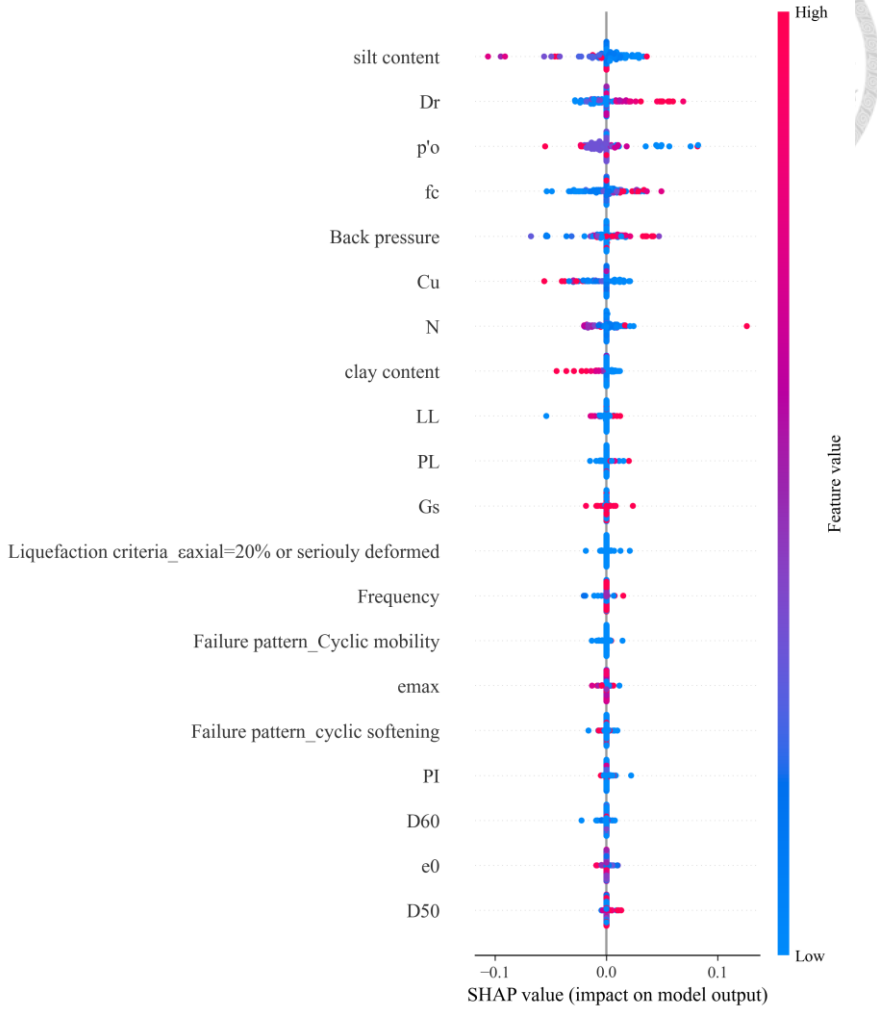


圖 C.14 包含所有特徵之 ANN 模型 SHAP 總結圖

doi:10.6342/NTU202502436

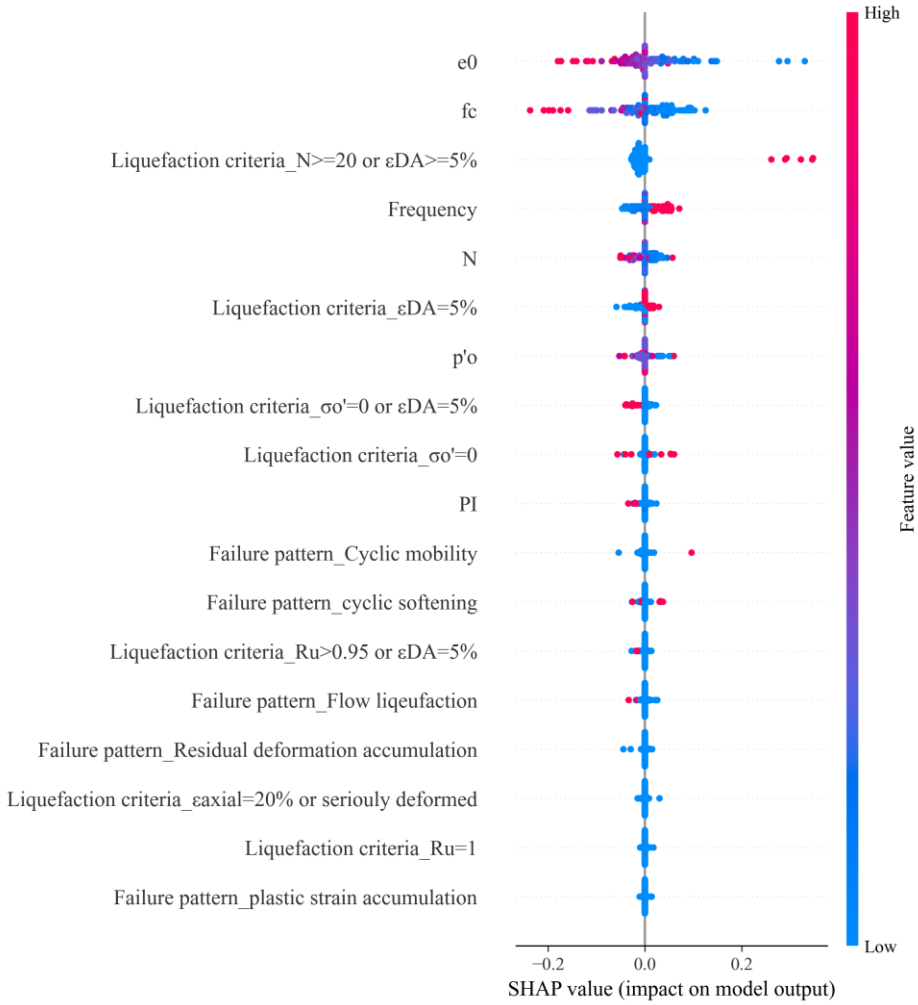


圖 C.15 包含主要特徵之 BNN 模型 SHAP 總結圖

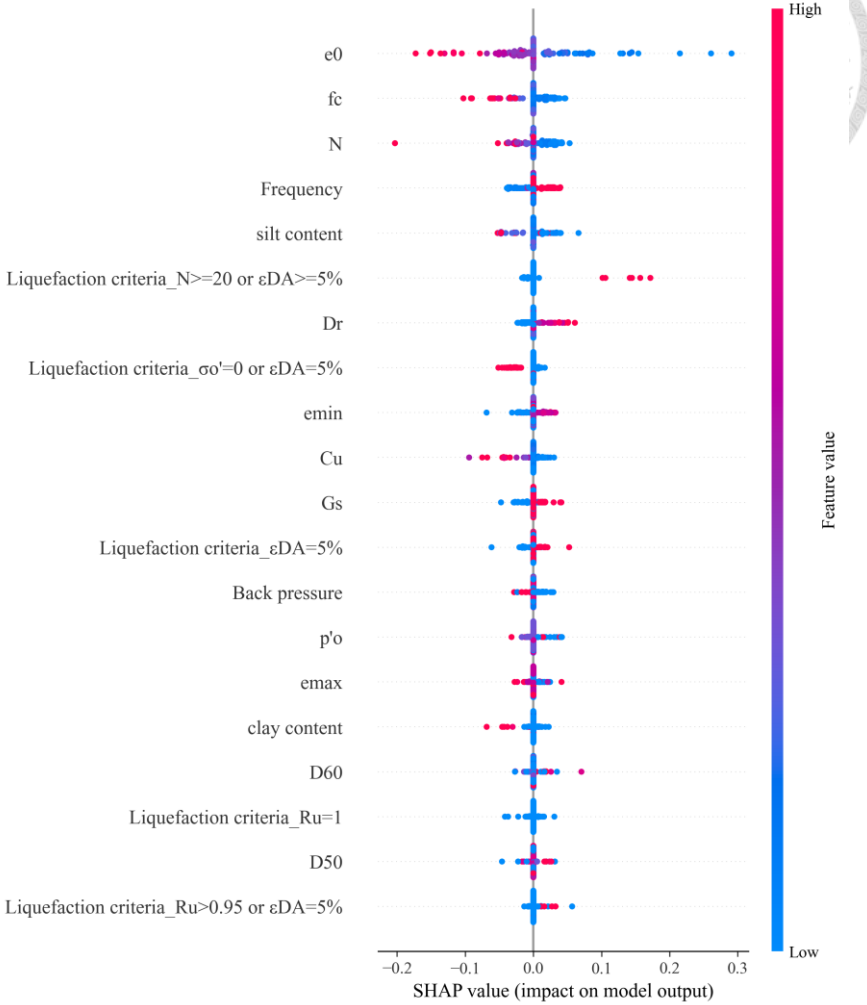


圖 C.16 包含所有特徵之 BNN 模型 SHAP 總結圖

doi:10.6342/NTU202502436