*Journal of Science and Engineering Technology, Vol. 21, No. 1, pp.27~36 (2025)*

# 結合正則化特徵工程與集成堆疊式學習準確率之研究

邱紹豐* 盧亭瑄 陳楷智
大葉大學 資訊工程學系
515006 彰化縣大村鄉學府路 168 號
*schiou@mail.dyu.edu.tw

## 摘 要

機器學習技術在各個領域的應用和成就，充分展示了其巨大的潛力和廣泛的應用前景。其中集成式學習是一種分階段應用不同機器學習演算法來訓練模型的技術，透過不同演算法間彼此校正誤差，可以降低模型的偏見並提高模型預測的準確率。集成學習中各個機器學習演算法獨自以資料集訓練資料，因此訓練模型的計算成本常常會較其他機器學習的方法高。如果能夠降低訓練資料量，且不會影響資料集中各特徵的代表性，就可以在不影響模型品質的前提之下，有效降低訓練的成本。而在機器學習領域中常用的特徵工程技術，則藉由原始數據中提取出對模型有意義的特徵，且不會降低整體資料的完整性。特徵工程應用在模型訓練的資料前處理中，可以提高模型的泛化能力、降低過擬合風險，並且有效降低模型訓練時間。本研究以特徵工程中特徵擷取法的三種方法，過濾法、包裝法、以及嵌入法等，以訓練資料比較彼此之間的準確率，作為集成式學習資料前處理的方法。實驗中也與特徵擷取常用的主成分分析方法比較，以驗證本研究所選擇的特徵工程方法具較高的優勢。在集成學習則採用集成堆疊式的學習方法，藉由其分階段整合不同學習器的預測結果，提升模型的品質。此外，為避免模型偏見過高的問題，本研究中以多樣性的觀點作為設計的重點，也就是提高學習器間的異質性，更提高模型的預測能力。為驗證所提出的架構，本研究採用 UCI 資料庫的威斯康辛乳癌資料集進行實驗，整合特徵工程與集成堆疊學習演算法作為訓練的方法，並與其他集成學習的方法在精確率、召回率、F1 成績、與準確率比較，結果均顯示以本研究所設計的多樣性集成堆疊方法有較佳的成績，提供一種具備更高預測率的學習架構。

**關鍵詞：**集成式學習，特徵工程，多樣性架構，主成分分析

# Improving Model Efficiency by Combining Feature Engineering and Ensemble Stacking Learning

**SHAO-FONG CHIOU*, TING-HSUAN LU and TONY CHEN**
*Department of Computer Science and Information Engineering, Da-Yeh University*
*No.168, University Rd., Dacun, Changhua 515006, Taiwan, R. O. C.*
**schiou@mail.dyu.edu.tw*

## ABSTRACT

Despite their potential, machine learning models are susceptible to overfitting and are expensive to train, and ensemble learning and feature engineering can be used to improve generalizability and reduce training costs. Ensemble learning, where various models are independently trained on a dataset, improves accuracy but increases computational cost. Feature engineering, where meaningful features are extracted from the training data, can be used to reduce the data volume and thus training costs for ensemble models without compromising model quality. In this study, several feature engineering methods, namely filter methods, wrapper methods, and embedded methods, were evaluated for their effectiveness as preprocessing methods for ensemble learning. These methods were evaluated against commonly used feature extraction methods, such as principal component analysis. The best-performing ensemble learning method that was then adopted in the proposed method was ensemble stacking learning, in which the predictions of each learner are integrated in stages to improve model quality. To reduce model bias, heterogeneity among learners was emphasized to further improve predictive capability. The proposed framework was validated experimentally on the Wisconsin Breast Cancer Dataset. The proposed diverse ensemble stacking method outperformed its state-of-the-art counterparts in precision, recall, F1 score, and accuracy.

***Key Words*****:** Ensemble Learning, Feature Engineering, Diversity Structure, Principal Component Analysis

## 一、前言

在這個一個資訊量急遽擴增的時代，大量資料被蒐集和儲存在各種龐大的數據庫中。雖然各項資料與數據能透過各種資料處理的技術，能有更精細的歸納與整理，但是這也衍生了一個挑戰，即如何在海量資料中挖掘出重要的資訊。隨著機器學習技術的出現提供了有力的幫助，它能夠將龐大的資料轉化為實用的知識，並且發現資料中的模式和趨勢，從中提取出有價值的訊息，成為解決各種複雜問題的主要工具。機器學習技術已應用在不同領域，如在醫療領域中機器學習被廣泛應用於疾病診斷、治療方案優化和患者預後預測等方面。透過分析大量的醫學影像數據，機器學習模型能夠準確識別早期癌症病變，從而提高診斷的準確性和及時性。在各種機器學習的技術中，集成式學習（Ensemble Learning）在多個研究中已經被證實其卓越的性能和穩定性，相較於其他機器學習的方法具有明顯的優勢。

集成學習透過結合多個基礎學習器的預測結果，能夠提高模型的泛化能力和預測精度。在學者 Fitriyani 等人的研究中顯示，在三種集成學習，分別為袋裝法（Bagging）、提升法（Boosting）、以及堆疊法（Stacking），以堆疊法透過二層學習器的架構，能提供較高的準確率 [5]。因為在模型建立過程中，資料集需經過二層不同的學習器訓練，因此訓練的成本會比其他學習的技術高。此外，訓練資料集中的特徵數量也會因為資料集的特性也會有不同。例如在常用的鳶尾花（Iris）的資料集僅有 4 個特徵，但是其他資料集，如威斯康辛大學乳癌資料集有 30 個特徵屬性 [3]，或是電力負載的資料集則有 140256 個特徵屬性 [4]。因此，如果資料集在訓練之前能夠減少特徵數量，對模型的訓練成本可以明顯的降低。

特徵工程（Feature Engineering）是一種由原始數據中提取對機器學習有意義特徵的技術，能夠提高模型的泛化能力、降低過擬合風險，並縮短訓練時間。在特徵工程中，主要分為特徵轉換、特徵擷取以及特徵選擇等三類。三類中以特徵選擇透過選擇原始特徵集中最具代表性的特徵，從而減少數據維度，可以減少資料的失真，因此在本研究中以特徵選擇做為特徵工程實驗的技術。

在本研究所設計的方法整合了特徵工程與集成堆疊的學習方法，透過特徵工程降低訓練資料量來減少集成堆疊學習所衍生的計算成本。同時在設計集成堆疊學習方法中，提高基礎模型之間的差異程度，也就是提升基礎模型的多樣性，有助於模型預測的準確率並減少最終結果的分類錯誤。希望經由計算整合不同特徵工程與集成學習的技術，找出一組最佳組合，能夠達到最高的預測準確率、召回率、精確率等指標。本論文的架構如下：第二章為在相關研究領域的文獻，介紹在特徵工程、集成學習等領域具代表性的研究成果，以探討本研究所採用的技術。第三章介紹在本研究中所設計的特徵工程與集成堆疊學習整合的架構，並說明組成選擇組

成元件的理由。第四章羅列以威斯康辛大學乳癌診斷資料集做為訓練資料所建立的模型，與未使用特徵工程的集成式學習各項指標的比較結果。最後第五章討論實驗結果的意義以及本實驗限制，並討論未來可能發展的方向。

## 二、文獻探討

特徵工程的技術在很多的研究中都顯示能夠降低資料集中的特徵數量，進而減少模型訓練的時間。而堆疊集成式的學習方式，則提供一整合多種學習的模式，盼藉由各種模型間的互補能夠提升預測的準確率。在本章中探討相關的研究成果，作為本研究發展的基礎。

### （一）特徵工程研究探討

特徵工程（Feature Engineering）在機器學習中是一項相當重要的技術，其目的是指從原始數據中提取出對機器學習模型有意義的特徵，以提升機器學習模型的性能。特徵工程是預處理的重要步驟，會直接影響到模型的效能和預測準確性。有效的特徵工程能夠提高模型的泛化（generalization）能力、降低過擬合（overfitting）風險，並能夠縮短訓練時間。隨著資料集中特徵數量快速的增長，特徵選擇的技術應用在機器學習和模式識別等領域，可以更凸顯此技術的優勢。學者 Chandrashekar 等人指出特徵選擇演算法之間的比較只能使用單一資料集來完成，因為每個底層演算法對於不同的資料會有不同的行為 [2]。這表示沒有一種特徵選擇演算法能夠通用於特定領域的全部資料集，因此當需要為資料集進行特徵工程時，必須實際測試該資料集於每一種特徵算法下的表現。

特徵工程可以大致分為三大類：特徵創建（Feature Creation）、特徵轉換（Feature Transformation）和特徵降維（Feature Dimensionality Reduction）等，而各分類則衍生出不同的技術種類，如圖 1 所示。

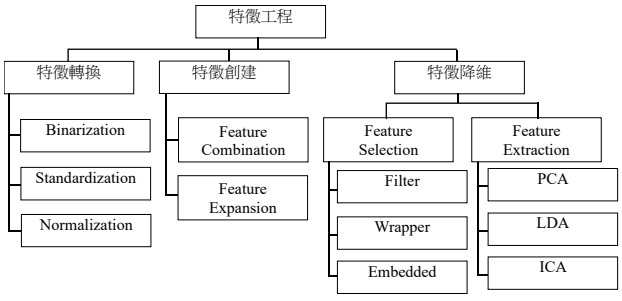



**圖 1. 特徵工程分類**

1. 特徵轉換

特徵轉換（Feature Transformation）是指將原始特徵進行數據轉換，以改變特徵的分佈或尺度，使其更適合的機器學習算法。特徵轉換的目的在於避免資料集中某些特徵領域(Domain)遠大其他特徵領域，造成在機器學習演算法中特徵權重不均的問題。特徵轉換的方法包含以下三種：標準化（Standardization）、正規化（Normalization）、以及特徵二值化（Binarization）。

標準化常用的技術為 Z-Score 標準化，其特色是根據每個數值特徵的均值和標準差對其進行縮放，使得所有特徵在同一尺度上進行比較。正規化方法較常用的方法是最小-最大縮放(Min-Max Scaling)，其特色是將數值特徵縮放到特定範圍內（通常是 0 到 1 之間），助於消除不同特徵之間的比例差異，使它們在同一範圍內進行比較。而特徵二值化是將連續數值特徵轉換為二進制形式，根據設置的閾值將其劃分為 0（小於或等於閾值）或 1（大於閾值）。

2. 特徵創建

特徵創建（Feature Creation）的原理是以領域知識（Domain Knowledge）創建新特徵的方法，過程透過使用原始數據或現有特徵來創建新的特徵，以捕捉更多的信息、提高模型性能或提供更好的數據表示。創建新特徵的過程包含了轉換、組合、提取或衍生，以創建預測能力更強的特徵。常用的兩種方法包含了特徵組合（Feature Combination），透過組合多個現有特徵來創建新特徵，例如，將身高和體重組合計算成 BMI 指數。另外則為特徵拓展（Feature Expansion）方法，對現有特徵進行多項式擴展或交互項擴展，以捕捉更高階的訊息和創造特徵間的交互作用。

3. 特徵降維

特徵降維（Feature Dimensionality Reduction）是指透過選擇或抽取最具信息量的特徵，以減少數據的維度。特徵降維的優點在於可以降低模型的複雜度，提升訓練速度並減少過擬合風險，因此往往是運用在特徵工程中的關鍵步驟。特徵降維常用的方法包含了過濾法（Filter Methods）、包裝法（Wrapper Methods）、以及嵌入法（Embedded Methods）等。過濾法通常不以機器學習算法，而是透過統計測量或其他評估指標來評估該特徵是否能夠幫助預測，然後選擇或移除特定的特徵。包裝法以所有特徵子集訓練並建構模型，再使用驗證資料集為每一種模型評分，以評分最高的模型所使用的特徵作為最終模型。這種方法的計算成本相對較大，但是能

為特定類型的模型找到性能最好的特徵集。嵌入法將特徵選擇嵌入到機器學習模型的訓練過程中，也就是嵌入法在模型的訓練過程中自動地選擇最重要的特徵，常見做法分為基於樹（Tree）模型與基於正則化（Regularization）的嵌入法。

**（二）集成學習研究探討**

集成學習（Ensemble Learning）是一種透過結合多個獨立的機器學習模型的機器學習方法，這種方法通常能夠產生比單個模型更好的預測性能。第一層這些被結合的模型又稱為基學習器（Base Learner）或基礎模型（Base Model），在組合成第二層單個模型後可以降低偏差和方差，從而提高預測準確性 [13, 15]。配合交叉驗證，有助於防止模型在第一層分類器和第二層分類器中使用相同的訓練集，可以有效防止過擬合的發生 [3]。常被應用於集成學習的方法，包括了裝袋法（Bagging）、提升法（Boosting）與堆疊法（Stacking）[1, 10]。

1. 裝袋法

裝袋法（Bagging）的技術是透過隨機抽樣產生多個子樣本集，這些子樣本集則用來分別訓練不同的基學習器。每個基學習器獨立訓練後，它們的預測結果會透過如投票、共識、或是取平均值進行組合，以獲得最終的預測結果。裝袋法可以有效地減少模型的方差，從而提升模型的穩定性和預測精度，常見的隨機森林（Random Forest）就是裝袋法的一種經典演算法。

2. 提升法

提升法（Boosting）透過依次訓練一系列的基學習器，每個基學習器都根據前一個學習器的表現來調整數據權重，使得後一個學習器更加關注前一個學習器分類錯誤的樣本，最終這些基學習器的預測結果會被線性組合以得到最終的預測結果。提升法可以有效地減少模型的偏差，從而提高預測的準確性。

3. 堆疊法

堆疊法（Stacking）透過訓練多層模型來提高預測性能，其中第一層由多個基學習器組成，這些基學習器對原始數據進行訓練並生成預測結果，並作為第二層學習器的輸入。第二層學習器通常被稱為元學習器（meta-learner）或元模型（meta-model），它會對這些輸入進行訓練並產生最終的預測結果。元模型通常使用 Logistic Regression（LR）作為第二層分類器，因為他能夠提供良好的分類效果 [5]。本研究採用的學習方法為堆疊法，其架構如圖 2 所示。

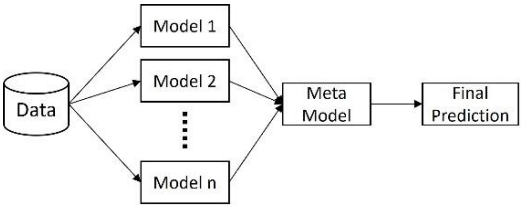



**圖 2. 堆疊法架構示意圖**

**(三) 集成學習多樣性研究**

在許多研究中發現，集成學習效能與預測準確度主要受到兩個因素的影響：基礎模型的準確性，以及基礎模型的多樣性(Diversity) [8, 12]。在基礎模型準確性對整體集成學習預測結果的影響，學者 Masisi 等人在 2008 年的研究中表示，如果各基礎模型的準確率偏低，會導致最終組合而成的集成學習模型準確率也會相應下降 [8]。基礎模型多樣性指的是各基礎模型之間的差異程度，也就是在相同數據輸入時，能夠有不同的預測結果。在集成學習中如果基礎模型之間有不同的預測結果，則表示該集成學習的模型具備有較高的多樣性，如此可以降低最終結果的分類錯誤 [6]。例如兩個基本模型，分別是決策樹（Decision Tree, DT），另一個是支持向量機（Support Vector Machine, SVM），決策樹傾向於將資料分成多個區域，每個區域做出不同的預測；而 SVM 則更注重在資料的分界線上找到最大間隔。如果將這兩個模型結合在一起，它們的不同看法可以彼此補充，使得集成模型能夠更全面地理解資料並做出更準確的預測。反之，如果所有基礎模型對某個數據的預測結果完全一致(無論是正確還是錯誤)，則表示該集成模型的多樣性較低。雖然已經有多項研究提出了不同多樣性的測量方法，但迄今為止尚未有一統一的測量標準 [9]。

## 三、研究方法

在前述相關研究中，多位學者指出透過集成學習所訓練的模型，在準確度上優於其他建立模型的方法。為進一步提升集成學習的成效，本研究的目的在探討特徵工程對集成學習的效能與準確度的影響。在實驗的設計中以不同的特徵工程搭配集成學習中的堆疊法，再依最終預測的結果評估每一種特徵工程與堆疊法組合的成效。因為特徵工程的種類眾多，因此在配合集成堆疊演算法訓練模型之前，會先進行各種特徵工程之間的比較，並且觀察在使用特徵工程之前資料的標準化對最後的預測性能的影響，實驗架構如圖 3 所示。實驗

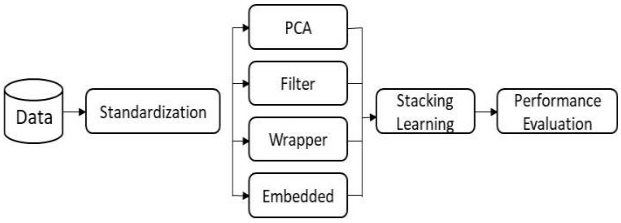



**圖 3. 研究架構圖**

的過程首先透過交叉比對有無標準化與不同特徵工程，包含了過濾法、包裝法、以及嵌入法等，並與常用的 PCA（Principal Component Analysis，主成分分析）比較、觀察預測性能在何種情況下表現最佳，最後分析特徵工程與具多樣性的集成學習，是否能夠提模型預測性能。

主成分分析（Principal Components Analysis, PCA）找到資料中變異量最大的方向，而這些方向最能代表資料的整體結構。因此 PCA 具有降維與資訊保留、去除相關性、數據壓縮、以及降噪與資料簡化等特性。PCA 能夠將高維度數據轉換到低維度空間，並以少數幾個主成分來描述數據的主要結構，這樣既減少了數據的維度，又保留了大部分有用資訊，是數據分析和機器學習中非常重要的工具。

**（一）特徵工程分析比較**

在這個部分的研究主要聚焦於提升集成模型的預測效能，因此可以先透過特徵工程提高訓練資料集品質，提升後續堆疊集成式模型的效能以及穩定性，再進行預測性能評估。根據文獻回顧的探討中，學者指出資料集的特性會影響不同特徵工程的效能，因此需透過實驗在相同資料集中，不同特徵工程與集成學習演算法組合的分類預測表現。在不同特徵工程分類所選擇的算法如下所列：

1. 過濾法：以皮爾森相關係數（Pearson Correlation Coefficient）作為代表算法，主要是此方法關注在特徵之間的相關性，可以用以汰除不需用為訓練資料的特徵。
2. 包裝法：選擇結合交叉驗證之遞迴特徵刪除法（RFECV），因為此算法結合了交叉驗證，具備有較好的堅固性（Robustness）。
3. 嵌入法：以基於正則化的羅吉斯迴歸（Logistic Regression）算法，因為正則化（Regularization）可以防止模型的過擬合。具相關研究顯示，L2 正則化提供了多種不同的求解器（Solver），且較 L1 正則化能保留較多的特徵，因此在嵌入法的實驗中採用 L2 正則化。
4. 除上述三種特徵選擇的方法，實驗中也以在特徵工程中特徵擷取技術常用的 PCA 演算法的結果，作為上述三種方法比較的基準，以驗證在特定資料集中效能較好的演算法。

為了解在資料集中特徵值範圍差距的大小是否會影響預測結果，以上四種特徵工程的方法均應用有標準化以及未標準化的資料集，並比較其間差異。標準化的方法是以均值為中心，按分量比例縮放至單位標準差（Variance），所謂單位標準差是一種數據縮放的方法如式(1)所示。透過此標準化方法，數據的平均值會變為 0，標準差會變為 1，使得數據符合標準常態分佈（Standard Normal Distribution），以便預測模型進行比較和分析。

$$x_i^{std} = \frac{x_i - \bar{x}}{\sigma(x)} \quad (1)$$

其中，$x$為資料的各種特徵數據，$x_i^{std}$為第 $i$ 筆標準化數據，$x^i$為第 $i$ 筆數據，$\bar{x}$是樣本平均值，$\sigma(x)$則是樣本標準差。

**（二）集成學習多樣性架構研究**

本研究採用集成學習的堆疊（Stacking）的方法，而依先前的研究顯示，堆疊法的成效依賴於基礎模型多樣性的程度。多樣性是在預測錯誤時，每個基礎模型盡可能擁有不同的預測結果。例如兩個基本模型為決策樹（DT），另一個是支持向量機（SVM），決策樹傾向於將資料分成多個區域，每個區域做出不同的預測；而 SVM 則更注重在資料的分界線上找到最大間隔，因此兩者所組成的基礎模型可以提高其多樣性。在本研究所設計的實驗中，提出一個具備多樣性基礎模型的堆疊架構，如表 1 所示。

SVM 是一種基於超平面決策邊界的分類器，其目標是找到能夠最大化類別之間邊界的超平面。這種方法使得 SVM 能夠處理複雜的非線性分類問題，並且對於高維度數據集也表現出色。這種獨特的歸納偏差使得 SVM 在處理不同類型的數據時具有獨特的性能。

**表 1. 基礎模型組成架構**

| | SVM | RF | NB |
|---|---|---|---|
| 原理 | 基於超平面決策邊界 | 基於決策樹 | 基於機率統計 |
| 優勢 | 對於非線性數據表現良好 | 泛化能力佳於大規模數據有良好表現 | 於小規模數據有良好表現 |

Random Forest（RF）是一種基於決策樹的集成學習方法，透過隨機抽樣和特徵選擇，建立多個決策樹並進行投票或平均來做出最終的預測。這種集成方式增加了模型的隨機性與多樣性，有助於降低過度擬合，同時也反映了對數據的不同解釋方式。

Naïve Bayes（NB）是基於機率統計的分類器，其核心是根據特徵之間的獨立性假設來計算後驗概率。這種獨特的設計使得 Naive Bayes 在處理低維度數據時效果顯著，同時也賦予了它與其他模型不同的歸納偏差。

羅吉斯迴歸（Logistic Regression, LR）是一種監督式分類學習的演算法，是統計學中對數機率模型。與線性迴歸（Linear Regression）類似之處，兩者都是用來評估變數之間的關聯，不同之處在線性迴歸應用在輸出連續數值的結果，而羅吉斯迴歸則輸出分類的結果。依相關研究的文獻顯示，羅吉斯迴歸（Logistic Regression, LR）做為第二層學習器，能夠提供良好的分類效果，參考 Fitriyani 等學者的研究成果[5]，本研究所設計的架構如圖 4 所示，包含交叉驗證的集成學習運作方式如下所示。

1. 將資料集分為訓練集 D 與測試集。
2. 將訓練集 D 劃分為 K 個不相交的子集，並訓練第一層分類器 K 次。
3. 將每次訓練完成的分類器應用於第 K 個子集，並將輸出蒐集作為第二層分類器的輸入特徵空間。
4. 使用蒐集的第一層分類器輸出訓練第二層分類器。
5. 整體訓練集重新訓練第一層分類器，以便使用所有數據集實例。
6. 第二層分類器與第一層分類器訓練完成，將測試集輸入第一層分類器。
7. 然後將第二層分類器應用於第一級分類器測試集輸出，以獲得最終預測結果。

**（三）評估指標**

本研究採用的評估指標為準確率（Accuracy）、精確率（Precision）、召回率（Recall）、F1 值（F1 score），這些指標主要基於混淆矩陣中的真陽性（True Positive, TP）、偽陽性（False Positive, FP）、真陰性（True Negative, TN）和偽陰性（False Negative, FN）四個數值相互計算而成。以下為各項評估指標計算公式：

1. 準確率（Accuracy）：代表著模型正確預測的樣本數佔總樣本數的比例，其值越高代表模型預測正確的能力越強，

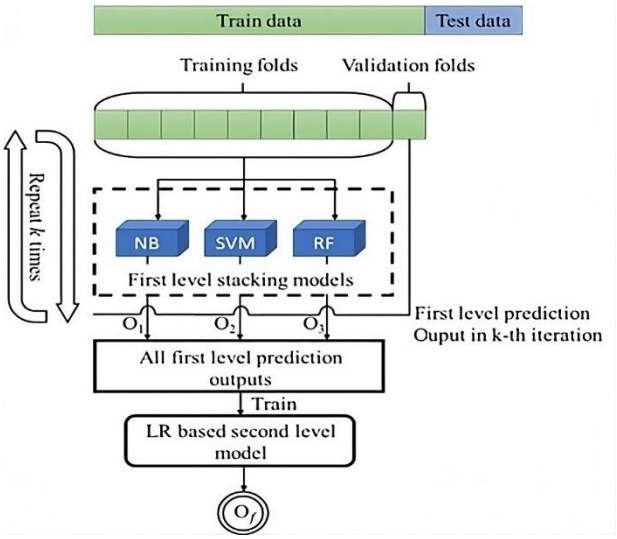



**圖 4. 集成學習模型架構圖 [7]**

其公式如式（2）所示。

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

1. 精確率(Precision)：在所有被模型判斷為正例的樣本中，實際上是正例的比例，Precision 越高，代表模型判斷為陽性的樣本中實際為陽性的比例越高，如式（3）所示。

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3. 召回率（Recall）：在所有實際為正例的樣本中，被模型正確判斷為正例的比例，Recall 值越高，代表模型能夠正確找到更多的陽性樣本，如式（4）所示。

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4. F1 值（F1 score）：是 Precision 和 Recall 的調和平均值，綜合考慮模型的精確率與召回率，其值越高代表模型在精確率和召回率之間取得了較好的平衡，如式(5)所示。

$$F1\ Score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \quad (5)$$

在實驗結果中會採用這四項指標進行最終評估，以 Accuracy 指標為主，觀察模型整體的預測正確率，再以 Precision 指標判斷模型預測為正例樣本的正確率，以及

Recall 指標觀察模型對於所有正例樣本的正確率，最後透過 F1 Score 指標觀察 Precision 與 Recall 兩者，是否能夠取得良好的平衡。

## 四、實驗結果

本研究依前述的設計方法，以公開的資料集驗證本研究所提出的方法與其他方法的差異。實驗主要分為兩個面向驗證，分別為準確度、執行時間，以此來評估判斷是否能夠改善乳癌預測以及觀察其中的運算成本。實驗成果如下所列：

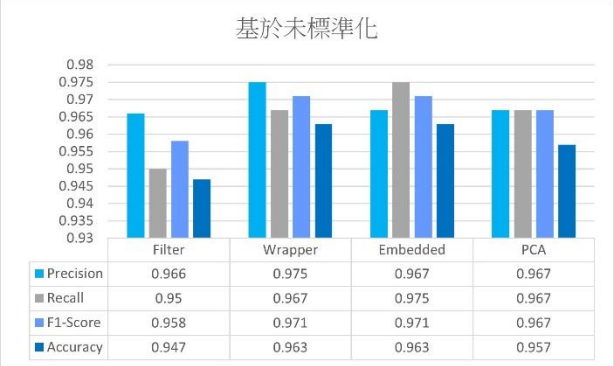
### （一）實驗資料來源

實驗資料集採用來自 UCI（University of California, Irvin）數據資料庫中的威斯康辛乳癌資料集（Wisconsin Breast Cancer Diagnostic Database，WDBC）[3]，針對乳腺癌診斷資料集總共有 569 筆診斷資料，其中診斷（Diagnosis）為良性有 359 筆（63.1%），惡性診斷則有 210 筆（36.9%）。資料集中 32 個屬性包含了一個 ID 屬性、一個目標屬性、以及 30 個特徵屬性。 特徵屬性包括了乳腺組織細胞的 10 個不同特徵的計算結果，這些特徵被分為三組，每組包含 10 個特徵的平均值、標準差以及最大值。

從資料集可以觀察到部分特徵值的範圍差異過大，例如 area_mean 特徵的最大值高達 2500，但 smoothness_mean 特徵的最大值卻只有 0.1634，這可能導致模型在訓練過程中過度關注數值較大的特徵，而忽略數值較小的特徵。因此本實驗中希望透過標準化將不同特徵調整到相同的數值範圍內，以確保模型不會偏重於某些特徵，進而提升模型的穩定性和預測性能。

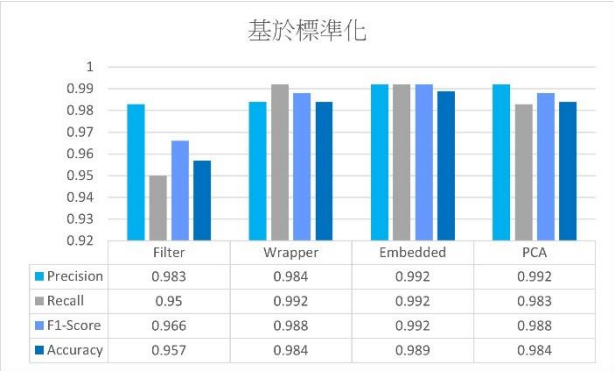
### （二）特徵工程的實驗結果

本節比較了常用的特徵擷取方法 PCA，與三種特徵選擇的方法：過濾法、包裝法、以及嵌入法，並且分別以特徵值標準化與未標準化比較在各特徵工程方法的準確性。在這四種方法的中，同時比較資料是否標準化的結果如圖 5 與圖 6 所示。在未經過標準化的特徵工程比較中，由圖 5 可以觀察到嵌入法有較佳的準確率（Accuracy）達 0.963 且召回率（Recall）達 0.975 也優於其他方法。此外也可以注意到包裝法（Wrapper）的準確率與 F1 Score 也達到了與嵌入法相同的數值，兩者之間的差別在於包裝法的結果著重於精確度，也就是模型帶來的預測更為可信，但對於陽性樣本的預測能力降低了；嵌入法的結果則是著重於召回率，代表他能夠找到更多陽性的樣本，但也因為如此導致誤判率也增加了。




| | Filter | Wrapper | Embedded | PCA |
|---|---|---|---|---|
| Precision | 0.966 | 0.975 | 0.967 | 0.967 |
| Recall | 0.95 | 0.967 | 0.975 | 0.967 |
| F1-Score | 0.958 | 0.971 | 0.971 | 0.967 |
| Accuracy | 0.947 | 0.963 | 0.963 | 0.957 |

**圖 5. 基於未標準化之分類性能評估**




| | Filter | Wrapper | Embedded | PCA |
|---|---|---|---|---|
| Precision | 0.983 | 0.984 | 0.992 | 0.992 |
| Recall | 0.95 | 0.992 | 0.992 | 0.983 |
| F1-Score | 0.966 | 0.988 | 0.992 | 0.988 |
| Accuracy | 0.957 | 0.984 | 0.989 | 0.984 |

**圖 6. 基於標準化之分類性能評估**

資料經過標準化的特徵工程實驗結果如圖 6 所示，由實驗結果可以發現，基於標準化的表現相較於未經標準化的表現好，尤其嵌入法的表現最為亮眼，準確率高達 0.989，並且在精確率、召回率與 F1 Score 也取得了最佳表現。由實驗結果也可以觀察到，包裝法與特徵擷取 PCA 在準確率上也達到了出色的 0.984，包裝法在召回率上取得了較好的成績，這表示他能夠找到更多陽性樣本，即使它可能將一些陰性樣本判斷為陽性；PCA 則是在精確率上取得了更好的表現，代表其訓練出來的模型預測的答案更為準確。

以上述四種特徵工程方法的實驗結果的比較，以乳癌資料集作為資料來源，在資料未經標準化的實驗中可以發現，嵌入法與包裝法都能夠幫助預測模型達到最高的準確率，但是嵌入法在召回率的表現更勝一籌，表示以此方法可以找到更多陽性樣本。

以同樣的資料來源但是經過標準化後，各項評估指標都得到全面的提升。經實驗結果顯示嵌入法一樣擁有最高的準確率，略勝 PCA 與包裝法 0.005。PCA 雖然準確率也相當高，但 PCA 是一種特徵擷取的工程，會破壞特徵可解釋性。此特性在醫療檢測領域中會有負面的影響，因為理想中特徵

工程能夠盡量保留特徵意義，讓醫療人員也能透過機器學習找到特徵與疾病間的關係。

**（三）堆疊集成式架構效能實驗結果比較**

學者 Ariana Tulus Purnomo 等人在其研究中[11]針對醫療資料集，使用了多種分類演算法，包括了 MLR、DT、RF、SVM、XGB、LGBM、CB、MLP，以及三種堆疊集成學習架構，即 SEC、BTSC 和 NSEM。該文獻實驗結果指出採用 NSEM（Neural Stacked Ensemble Model）架構，可以得到最佳的準確率，因此在本實驗中選擇 NSEM 架構作為對比。

Neural Stack Ensemble Model 是一種結合多個神經網路模型的架構，其主要目的是透過「集成學習」（ensemble learning）來提高模型的預測準確性和穩健性。其運作方式在第一層有多個基礎神經網絡（如 CNN、RNN 或 Transformer），每個網絡都獨立學習並進行預測。這些基礎模型的預測結果會作為輸入，傳遞給更上層的模型。最後上層模型會學習如何最佳地結合下層模型的預測，產生最終結果。

為保持實驗公正性，首先需要依照 NSEM 文獻提供之分類器設定如表 2 所示，完整還原 NSEM 的 Base-Models 參數設置。且因本篇實驗專注於架構改善能夠提升準確率與執行時間，考慮到 NSEM 架構並無經特徵工程，因此將經本篇特徵工程（標準化和嵌入法）情況與無經特徵工程情況的實驗結果分別列出，以公平進行實驗對照。

在資料未經特徵工程處理的實驗結果如圖 7 所示，本研究所提出的基於多樣性堆疊（Diversity Stacking）模型的表現，在各項評估指標中都勝過於 NSEM 的結果。實驗結果顯示多樣性堆疊模型在預測方面，比 NSEM 找到更多的陽性樣本，且預測得更為精確，如此也說明了為什麼最終準確率能夠優於 NSEM 架構。

在資料經過特徵工程的實驗結果如圖 8 所示，由實驗結果可以觀察到再經過本研究提出之特徵工程後，兩者的評估指標都有明顯的提。雖然 NSEM 的精確率（Precision）從 0.974 大幅度提升至 0.991，但由於召回率（Recall）只有微幅度的提升，導致各項指標不如多樣性堆疊（Diversity Stacking）模型。與之不同的是多樣性堆疊的各項指標獲得全面性的提升，這讓多樣性堆疊的最終準確率勝過了 NSEM 的最終準確率。

綜合本研究中實驗結果如表 3 所示，經由上述實驗可以得出，集成學習堆疊的 Base-Models 具備多樣性，能夠提高幫助模型準確率，且於未經特徵工程之實驗與經特徵工程之實驗兩者情況，都達成了一致的實驗結果。另外在結果中也顯示，本研究所提出之特徵工程架構，也提升了 NSEM 模型的預測準確率。這表示在醫療資料集中，透過標準化再進行嵌入法，也適用於其他模型，並且是能夠有效幫助提升預測準確率。

**表 2. NSEM 實驗參數 [14]**

| Models | Hyperparameter | Values | Final Value |
|---|---|---|---|
| XGB | learning_rate | 0.01, 0.2, 0.3, 0.5 | 0.2 |
| | subsample | 0.5, 0.6, …, 1 | 1 |
| | colsample_tytree | 0.5, 0.6, …, 1 | 1 |
| | max_depth | 3, 4, …, 8 | 7 |
| | n_estimators | 100, 200, …, 1000 | 700 |
| | alpha | 0.7, 1, 1.3 | 0.7 |
| | gamma | 0, 0.5, 1 | 0 |
| LGBM | eta | 1e-1, 1e-2, 1e-3 | 1e-1 |
| | min_child_weight | 1e-5, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3, 1e4 | 1e-5 |
| | subsample | 0.1, 0.2, …, 1 | 0.4 |
| | colsample_bytree | 0.1, 0.2, …, 1 | 0.7 |
| | max_depth | 3, 4, 5, 6, 7, 8 | 5 |
| | n_estimators | 100, 200, …, 1000 | 900 |
| | num_leaves | 10, 20, …, 100 | 10 |
| CB | learning_rate | 0.03, 0.001, 0.01, 0.1, 0.2, 0.3 | 0.1 |
| | depth | 3, 4, …, 10 | 8 |
| | iterations | 100, 200, …, 1000 | 400 |
| | l2_leaf_reg | 1, 3, 5, 10, 100 | 1 |

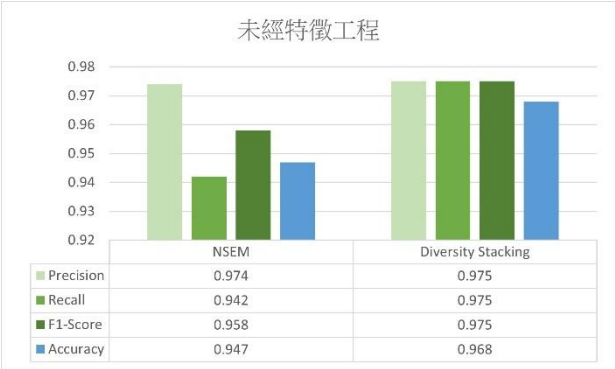



**圖 7. 未經標準化與嵌入法實驗結果**


**圖 8. 經標準化與嵌入法實驗結果**

**表 3. NSEM 與多樣性堆疊結果比較**

| **未經特徵工程** | | |
|---|---|---|
| | NSEM | Diversity Stacking |
| Precision | 0.974 | 0.975 |
| Recall | 0.942 | 0.975 |
| F1-Score | 0.958 | 0.975 |
| Accuracy | 0.947 | 0.968 |
| **經特徵工程** | | |
| | NSEM | Diversity Stacking |
| Precision | 0.991 | 0.992 |
| Recall | 0.950 | 0.992 |
| F1-Score | 0.970 | 0.992 |
| Accuracy | 0.963 | 0.989 |

## 五、結論與未來展望

本研究提出一種多樣性集成學習架構，以改善集成學習的預測性能，並透過比較不同特徵工程的方法，挑選出預測表現最佳的組合，幫助集成學習提高預測上限。在特徵工程比較的實驗中，使用特徵降維中的過濾法、包裝法、嵌入法與特徵擷取四者做比較，其中以嵌入法的基於正則化的羅吉斯迴歸特徵選擇表現最佳。在集成學習實驗中，透過將不同原理與優勢的基礎模型組合，使集成學習架構更具多樣性。實驗透過與過往文獻提出的集成學習架構 NSEM 比較，發現本研究提出之多樣性集成學習架構，在各項評估指標中都勝過於 NSEM，這表示多樣性能夠有效提升集成模型的預測性能。

在本研究所提出的方法於實驗中達到超果 99%的準確度，對乳癌判斷提供相當可靠的參考性。但考量實驗所使用數據集資料量是否具代表性並未有明確證明，因此對數據代表性以及資料量適當性，在未來需要更多的研究。此學習模型所達到的效果，應用在其他資料集是否能有相近的成果，在未來的研究中需更多的實驗數據以支持此模型的可靠度。

在集成學習實驗中，主要關注於利用多樣性提升集成學習的預測性能，但由於目前為止多樣性並沒有一個公認定義，也沒有統一測量多樣性的方法，因此本研究提出之多樣性集成學習架構，無法完全證明多樣性與預測準確率的關係，只能透過準確率判斷是否能夠有效提升預測性能。而近期許多學者，針對多樣性提出測量方法，若是將來多樣性能夠有效被測量且定義，未來研究中可以透過量測多樣性數值與預測性能之間的關係，從而更明確證明本研究集成學習預測準確率提升，是與集成學習的多樣性改善有直接關係。

此外，與集成學習類似的聯邦學習（Federated Learning）近來也受到相當的關注，兩者都是結合多個模型的學習方式，以提升學習效能與穩定性。兩者學習方式的差異主要在集成學習在一次訓練完成後即進行整合，而聯邦學習則是反覆進行本地訓練與參數的彙整更新。兩者效能比較與應用，在未來可以更深入的研究。

## 參考文獻

1. Aggarwal, C. (Ed.) (2014) *Data Classification: Algorithms and Applications*, 1st ed., Chapman and Hall/CRC, Boca Raton, FL.
2. Chandrashekar, G. and F. Sahin (2014) A survey on feature selection methods, *Computers & electrical engineering*, 40(1), 16-28.
3. Diagnostic Wisconsin Breast Cancer Database (1995) Retrieved October 10, 2024, from https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic.
4. ElectricityLoadDiagrams20112014 (2015) Retrieved October 10, 2024, from https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014.
5. Fitriyani, N. L., M. Syafrudin, G. Alfian and R. Jongtae

(2019) Development of disease prediction model based on ensemble learning approach for diabetes and hypertension, *IEEE Access*, 7, 144777-144789.

6. Lee, Y. and H. Kim (2018) A data partitioning method for increasing ensemble diversity of an eSVM-based P300 speller, *Biomedical Signal Processing and Control*, 39, 53-63.
7. Masisi, L., V. Nelwamondo and T. Marwala (2008) The use of entropy to measure structural diversity, Proceedings of the 2008 IEEE International Conference on Computational Cybernetics, Stará Lesná, Slovakia.
8. Mienye, I. D. and Y. Sun (2022) A survey of ensemble learning: concepts, algorithms, applications, and prospects, *IEEE Access*, 10, 99129-99149.
9. Nascimento, Diego S. C. N., A. M. P. Canuto, L. M. M. Silva and A. L. V. Coelho (2011) Combining different ways to generate diversity in bagging models: An evolutionary approach, The 2011 International Joint Conference on Neural Networks, San Jose, CA.
10. Polikar, R. (2012) Ensemble learning. In: *Ensemble Machine Learning: Methods and Applications*, 1-34, C. Zhang and Y. Ma Eds. Springer Science + Business Media, New York, NY.
11. Purnomo, A. T., K. S. Komariah, D. Lin, W. F. Hendria, B. Sin and N. Ahmadi (2022) Non-contact supervision of COVID-19 breathing behavior with FMCW radar and stacked ensemble learning model in real-time, *IEEE Transactions on Biomedical Circuits and Systems*, 16(4), 664-678.
12. Sagi, O. and L. Rokach (2018) Ensemble learning: a survey, *WIREs Data Mining and Knowledge Discovery*, 8(4), e1249。
13. Sakkis, G., I. Androutsopoulos, G. Paliouras, V. Karkaletsis, C. Spyropoulos and P. Stamatopoulos (2001) Stacking classifiers for anti-spam filtering of e-mail, *Proceedings of Empirical Methods in Natural Language Processing*, 44-50.
14. Turner, C. Reid, F. Alfonso, L. Lavazza and A. F. Wolf (1999) A conceptual basis for feature engineering, *Journal of Systems and Software*, 49(1), 3-15.
15. Wolpert, D. (1992) Stacked generalization, *Neural networks*, 5(2), 241-259.

**收件：113.10.15　修正：113.12.12　接受：114.02.26**