

Data Science Project on Student Depression

Team: 6

Members: 312554010 周鈺祥, 312554036 陳胤宏

Part I: Survey on the Core Topic

List of all surveyed papers

- #1: D. Salas-Rueda, "Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis," JMIR Formative Research, vol. 6, no. 8, p. e37618, Aug. 2022
- #2: A. Yoo et al., "Prediction of Adolescent Depression from Prenatal and Childhood Data from ALSPAC Using Machine Learning," Scientific Reports, vol. 14, no. 23282, Oct. 2024
- #3: A. R. Lyles, J. K. Smith, M. T. Nguyen, and L. M. Johnson, "Identifying adolescent depression and anxiety through real-world data and social determinants of health: Machine learning model development and validation," JMIR Ment. Health, vol. 12, no. 1, p. e66665, Jan. 2025.
- #4: X. Liu, R. Cang, Z. Zhang, P. Li, H. Wu, W. Liu, and S. Li, "Identification of depressive symptoms in adolescents using machine learning combining childhood and adolescence features," BMC Public Health, vol. 25, no. 264, 2025.
- #5: A. F. Bastos, O. Fernandes-Jr, S. P. Liberal, A. J. L. Pires, L. A. Lage, O. Grichtchouk, A. R. Cardoso, L. de Oliveira, M. G. Pereira, G. M. Lovisi, R. B. De Boni, E. Volchan, and F. S. Erthal, "Academic-related stressors predict depressive symptoms in graduate students: A machine learning study," Behavioural Brain Research, vol. 478, Art. no. 115328, 2025.
- #6: M. Gil, S.-S. Kim, and E. J. Min, "Machine learning models for predicting risk of depression in Korean college students: Identifying family and individual factors," Frontiers in Public Health, vol. 10, Art. no. 1023010, Nov. 2022.
- #7: D. Khadayat and P. Poudel, "Predicting Student Depression Using Machine Learning," Preprint, ResearchGate, Mar. 2025.
- #8: C. El Morr, M. Jammal, I. Bou-Hamad, S. Hijazi, D. Ayna, M. Romani, and R. Hoteit, "Predictive machine learning models for assessing Lebanese university students' depression, anxiety, and stress during COVID-19," Journal of Primary Care & Community Health, vol. 15, pp. 1–12, 2024.
- #9: L. López Steinmetz, J. Godoy, and M. F. Fong, "Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine," Psychiatry Research, vol. 295, p. 113629, 2021.
- #10: M. S. Hossain, M. A. K. Azad, M. I. R. Sarker, M. B. Ahsan, and K. Andersson, "Comparison of the Performance of Machine Learning-based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh: A Result of the First Wave of the COVID-19 Pandemic," BMC Psychiatry, vol. 22, no. 1, pp. 1–14, Mar. 2022

Brief introduction to each paper

1. Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis

1.1 Target problem

心理健康狀況為全球健康優先議題，其中憂鬱症是青少年中主要疾病之一，大多數的心理健康疾病會在 14 歲以前出現，但常因未即時治療而對兒童的心理、身體健康造成長期嚴重後果。作者希望透過 ML 技術提前預測與學童憂鬱與焦慮相關的風險因子。

1.2 Dataset

此篇論文使用的資料集來自 2013 至 2014 年間，由巴勒斯坦教育部與 Al-Quds University 合作，收集來自 100 所學校共 3984 位年齡介於 10 至 15 歲的學生，共有 23 項特徵變數，包含性別、年齡、家長教育程度、身體活動、飲食、學業成績、社交支持、PTSD、生活滿意度、健康感知、暴力經驗與欺凌行為等。

1.3 Overview of the framework

此篇將資料集切分為訓練集 70%，測試集 20%，驗證集 10%，使用 5 種基礎的機器學習模型預測學童是否有憂鬱或焦慮症狀，並找出重要的風險因子。

1.4 Results

憂鬱預測 SVM 的準確率(92.6%)最高、RF(92.4%)，焦慮預測 RF(78.4%)、SVM(76.5%)，在其他表現也皆為最佳，證明 SVM 及 RF 適合用於預測學童心理健康風險。

| Model, mental Health condition | AUC ^a , % | CA ^b , % | Error rate, % | F1-score ^c , % | Precision, % | Recall, % |
|----------------------------------|----------------------|---------------------|---------------|---------------------------|--------------|-----------|
| Decision tree | | | | | | |
| Depression | 86.7 | 88.5 | 88.5 | 88.5 | 88.5 | 86.7 |
| Anxiety | 73.7 | 74.4 | 74.1 | 74.2 | 74.4 | 73.7 |
| Support vector machine | | | | | | |
| Depression | 96.8 | 92.5 | 92.6 | 93.7 | 92.5 | 96.8 |
| Anxiety | 82.1 | 76.4 | 76.5 | 76.8 | 76.4 | 82.1 |
| Random forest | | | | | | |
| Depression | 97.2 | 92.4 | 92.4 | 93.3 | 92.4 | 97.2 |
| Anxiety | 86.8 | 78.6 | 78.4 | 78.5 | 78.6 | 86.8 |
| Artificial neural network | | | | | | |
| Depression | 96.8 | 91.9 | 91.9 | 92.3 | 91.9 | 96.8 |
| Anxiety | 84 | 75.9 | 75.7 | 75.7 | 75.9 | 84 |
| Naive Bayes | | | | | | |
| Depression | 95.5 | 86.9 | 87.1 | 89.9 | 86.9 | 95.5 |
| Anxiety | 82.3 | 73 | 72.7 | 72.8 | 73 | 82.3 |

▲ Performance measures analysis for the different machine learning models

2. Prediction of Adolescent Depression from Prenatal and Childhood Data from ALSPAC Using Machine Learning

2.1 Target problem

作者認為造成憂鬱的發展可能早在孕期就已經開始，根據壓力敏感化的假說，父母患

有精神疾病或家庭經濟困難也會提高日後罹患憂鬱症的風險，因此若能在 10 歲前辨別出未來可能罹患憂鬱症的青少年，將有助於降低對社會成本的影響。

2.2 Dataset

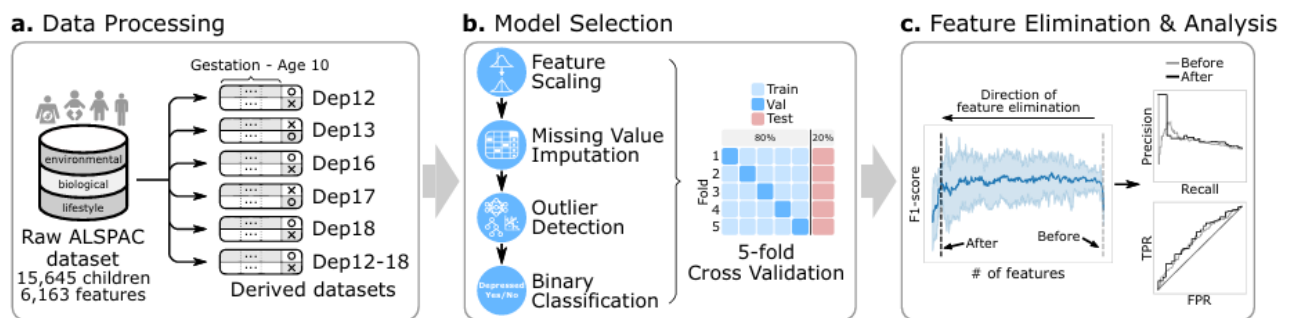
此研究所使用的資料集來自 ALSPAC(Avon Longitudinal Study of Parents and Children)，樣本數為 15645 位孩童，追蹤時長從孕期至成年，紀錄方式包含問卷、觀察資料、行政紀錄、生物樣本等等，其中有 8467 筆資料具有 12 至 18 歲間某個時間點的憂鬱症資料，憂鬱症狀以 SMFQ 量表作為依據，分數範圍為 0 到 26 分，得分大於等於 12 為憂鬱，反之則為未憂鬱。

2.3 Overview of the framework

Data Processing：此論文選定 12、13、16、17、18 歲作為預測點(其餘缺乏資料)，以及一個合併預測 12 至 18 歲任一時間點，經過資料清洗、資料合併與缺失值處理後，剩餘 885 個特徵做為模型輸入。

Model Selection：作者將資料集隨機分為訓練集(80%)與測試集(20%)，分別建構了橫斷式模型與時序模型(RNN、LSTM 架構，考慮時間資訊)，接著使用 5-fold cross validation 與 Grid Search 最佳化。

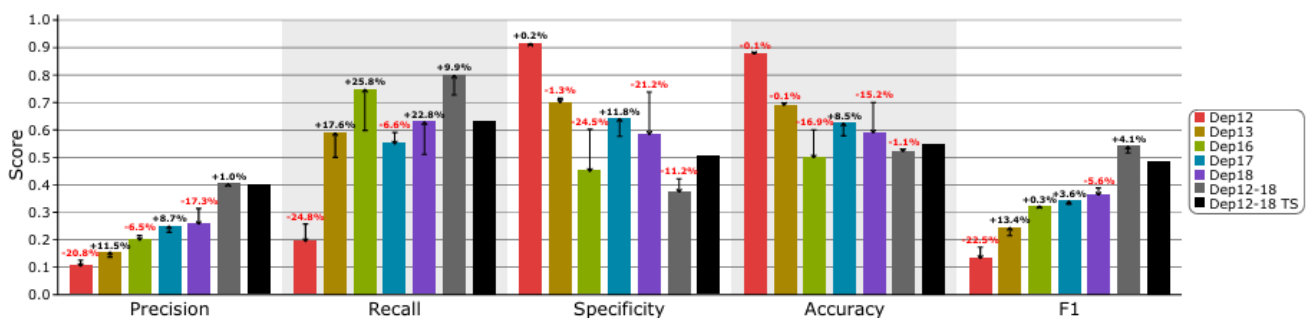
Feature Elimination & Analysis：作者採用 SBFS 方法用做特徵消除，逐步移除不重要特徵，僅保留對預測有幫助的特徵，透過交叉驗證確保效能不下降的前提精簡特徵。



▲ Overview of the adolescent depression prediction framework

2.4 Results

在 Dep12-18 資料集中，準確率達 64%、召回率 0.59、特異度 0.61、F1-score 0.54，可以觀察到時序模型(TS)表現略低於橫斷模型，原因為資料遺漏率高導致資料不連續影響模型效能，作者也發現到每個年齡階段主要影響特徵也會隨著改變。



▲ Performance and evaluation of the machine learning models

3. Identifying adolescent depression and anxiety through real-world data and social determinants of

health: Machine learning model development and validation

3.1 Target Problem

作者想利用機器學習技術結合真實世界資料(RWD)與社會健康決定因子(SDoH)，辨別青少年的憂鬱與焦慮狀況。

3.2 Dataset

資料源自於佛羅里達大學的 UF-IDR，總共 52054 名青少年資料，包含 12572 名焦慮症患者、7812 名憂鬱症患者、14019 名患有任一症狀者(移除重複計算)，將資料結合至 National Historical Geographic Information System，使病歷資料能連結到所屬區域的相應資訊，作者也將地區劣勢指數(ADI)連結到每個病患的地址，ADI 為衡量社區層級的重要指標，ADI 數值越高表示該地區的社會劣勢程度越高。

3.3 Overview of the framework

Data Processing：對數值進行標準化與補值缺失值採用方法為 KNN。

Feature Selection：此論文使用到卡方檢定移除無顯著差異的類別變數，連續變數則使用 T 檢定方法，接著使用 LASSO 進一步精簡特徵。

最終使用 XGBoost 作為主模型，總共訓練了 9 種模型，依據預測目標與特徵類型分為三組，預測目標分別為焦慮、憂鬱、兩者皆有，特徵類型分別為使用完整特徵、使用 ADI、完全不使用 SDoH，與使用 5*5 Nested Cross Validation。

3.4 Results

最終預測結果發現含有 SDoH 特徵焦慮與憂鬱模型 AUC 為 0.78、憂鬱模型 AUC 為 0.81、焦慮模型 AUC 為 0.80，不使用 SDoH 特徵表現為綜合模型 AUC 0.78、憂鬱模型 AUC 為 0.81、焦慮模型為 0.78，兩者差距不大，SDoH 對模型沒有幫助，SHAP 分析則顯示重要因子包含性別、種族、教育程度、生理指標。

| | Accuracy, mean (SD) | Balanced accuracy, mean (SD) | Sensitivity, mean (SD) | Specificity, mean (SD) | AUC, ^a mean (SD) |
|------------------------|---------------------|------------------------------|------------------------|------------------------|-----------------------------|
| Depression and anxiety | 0.73 (0) | 0.70 (0.01) | 0.65 (0.01) | 0.75 (0.01) | 0.78 (0) |
| Depression | 0.76 (0.01) | 0.73 (0) | 0.68 (0.02) | 0.78 (0.01) | 0.81 (0.01) |
| Anxiety | 0.75 (0.01) | 0.72 (0) | 0.66 (0.02) | 0.78 (0.01) | 0.80 (0) |

▲ Predictive performance of ML models by incorporating SDoH features

| | Accuracy, mean (SD) | Balanced accuracy, mean (SD) | Sensitivity, mean (SD) | Specificity, mean (SD) | AUC, ^a mean (SD) |
|------------------------|---------------------|------------------------------|------------------------|------------------------|-----------------------------|
| Depression and anxiety | 0.72 (0.01) | 0.70 (0.01) | 0.66 (0.01) | 0.74 (0.01) | 0.78 (0) |
| Depression | 0.76 (0) | 0.73 (0) | 0.68 (0) | 0.77 (0) | 0.81 (0) |
| Anxiety | 0.75 (0.01) | 0.71 (0.01) | 0.65 (0.01) | 0.78 (0.01) | 0.79 (0) |

▲ Predictive performance of ML models by incorporating ADI only

| | Accuracy, mean (SD) | Balanced accuracy, mean (SD) | Sensitivity, mean (SD) | Specificity, mean (SD) | AUC, ^a mean (SD) |
|------------------------|---------------------|------------------------------|------------------------|------------------------|-----------------------------|
| Depression and anxiety | 0.73 (0) | 0.70 (0) | 0.64 (0.01) | 0.76 (0) | 0.78 (0) |
| Depression | 0.76 (0.01) | 0.73 (0) | 0.68 (0) | 0.77 (0.01) | 0.81 (0) |
| Anxiety | 0.72 (0) | 0.71 (0) | 0.69 (0.01) | 0.73 (0.01) | 0.78 (0) |

▲ Predictive performance of ML models without incorporating SDoH features

4. Identification of depressive symptoms in adolescents using machine learning combining

childhood and adolescence features

4.1 Target Problem

主要是想了解在孩童時期與青少年時期的特徵，是否能讓 Machine Learning Model 更能準確預測在青少年時期的憂鬱症狀。以往常使用人口統計模型來進行比較，但畢竟在不同時期的心理、家庭關係、行為舉止的特徵都不盡相同，這篇論文提到想了解哪些特徵對憂鬱症的預測準確度最重要，並也比較不同的 Machine Learning Model 方法。

4.2 Dataset

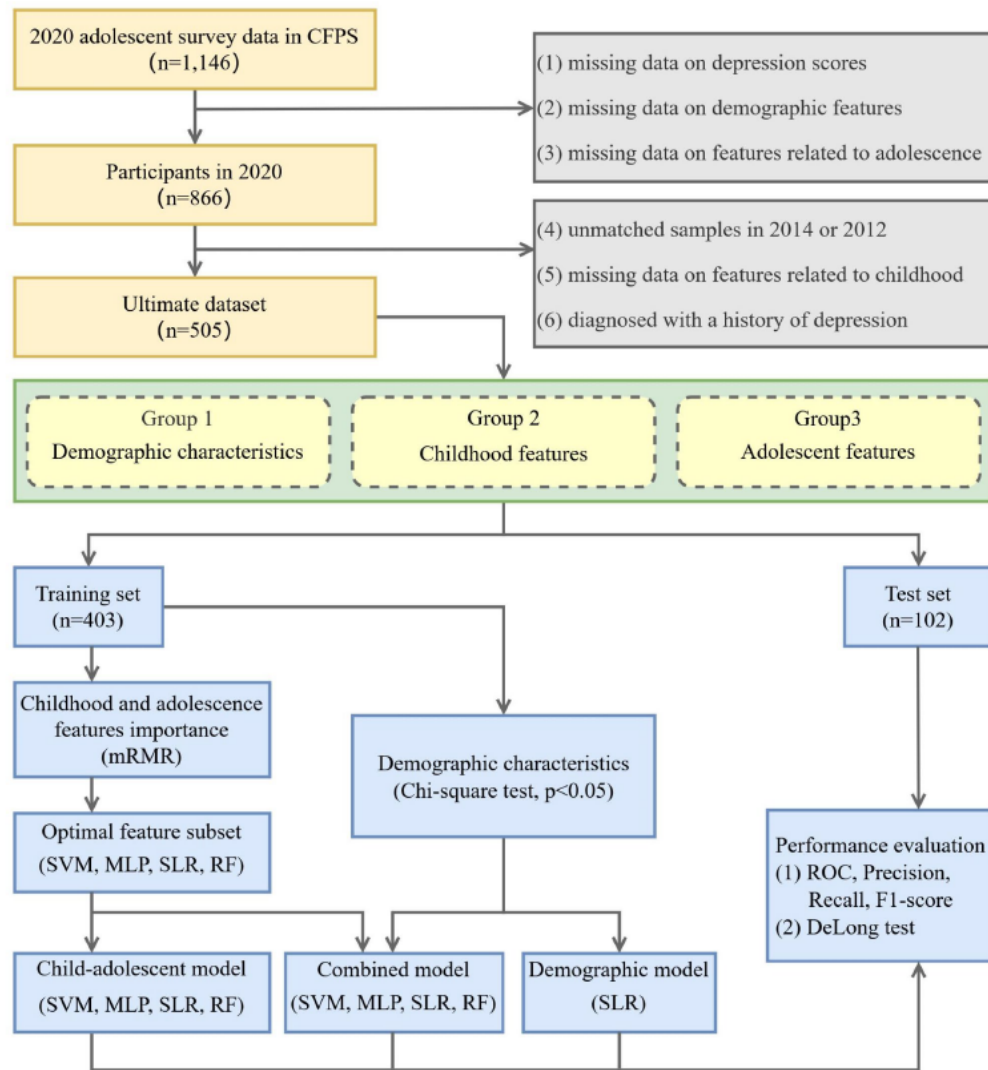
這篇使用到的資料來自於中國官方家庭追蹤調查(CFPS)，事由北京大學中國社會科學調查中心收集的，本身屬於全國性且長期追蹤的多學科大型調查。而研究使用的是 2020 年的 13 到 15 歲的青少年樣本資料，往回推估至 2012 到 14 年，當時這些青少年在他們 6 到 9 歲孩童時期的特徵資料，且經過篩選後，最終有 505 個青少年樣本進行分析。使用的特徵總共有 39 個，有包括 9 個童年特徵(父母缺席、積極人格、與同儕關係品質、自主性、關心孩子學業、因成績不好而懲罰、父母陪伴與照顧、限制看電視、教育期待)、30 個青少年時期特徵(網路使用頻率、網路娛樂重要性、網路社交重要性、網路學習重要性、運動頻率、午休習慣、睡眠時數、是否與父母爭吵、是否能與父母談心、是否能與他人分享煩惱、零用錢數量、正向行為表現)。

(<https://www.issp.pku.edu.cn/cfps/>)。

4.3 Overview of the framework

Data Preprocess: 首先將 CFPS 的資料進行移除青少年時期與孩童時期的缺失值、人口統計學特徵(性別、年齡、居住地、家庭成員數、健康狀況自評)、拆成三個資料集來訓練，分別是人口統計學資料、童年+青少年特徵資料、人口統計+精選特徵資料。將剛提到的資料使用 Z-score 標準化以集 Borderline-SMOTE 解決類別不平衡的問題。

Feature Engineering: 為了篩選哪些特徵是比較重要的，作者這邊使用最大關聯最小冗餘 (mRMR) 方法，找出哪些是比較相關的特徵，然後再根據最大關聯最小冗餘的排序結果，進一步用 Model 來 5-fold cross-validation.



▲ Flowchart of this study

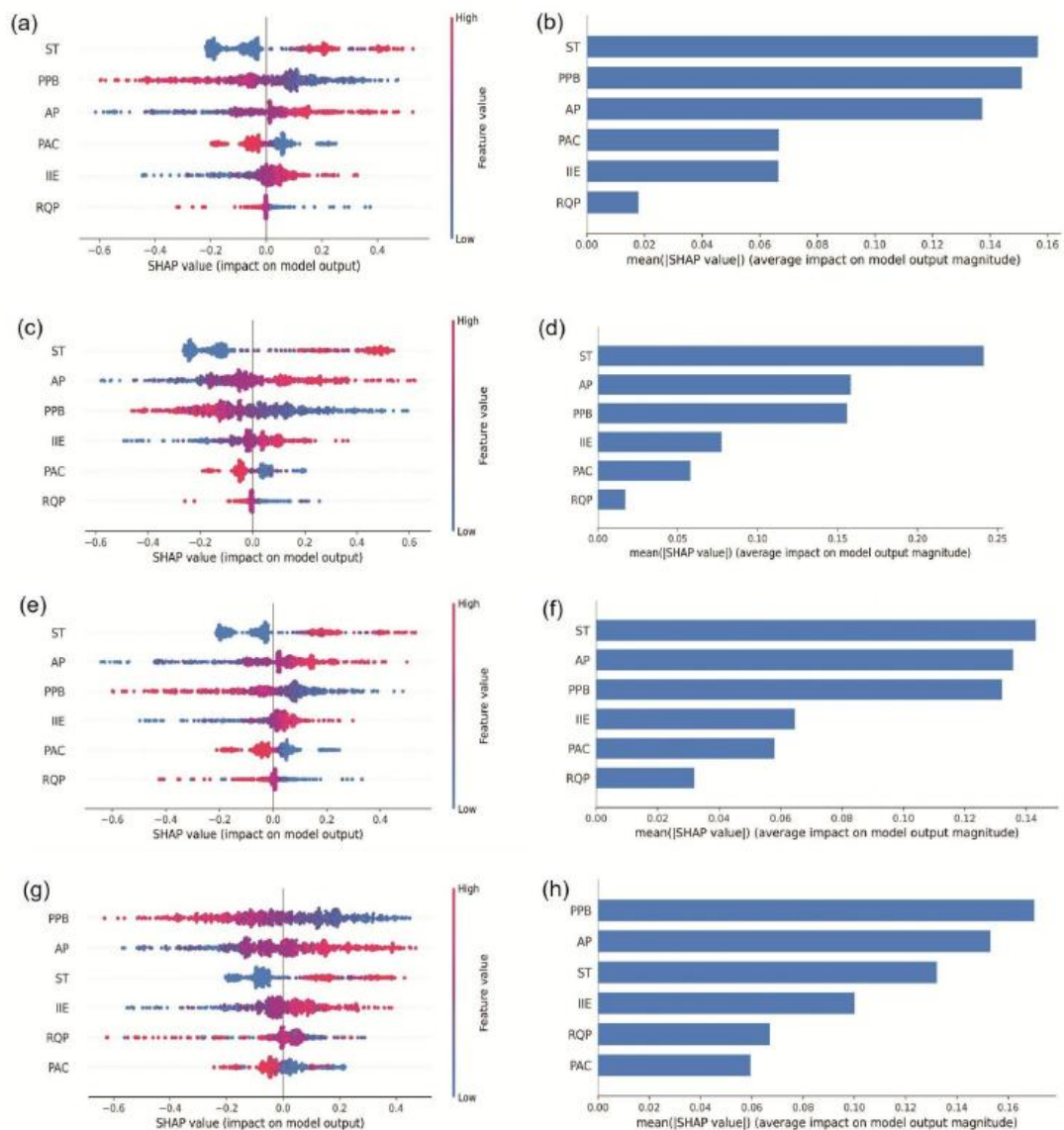
最後使用 Random Forest、SVM、MLP、Stepwise Logistic Regression 四種常見的 Machine Learning Model 來預測結果，之後再使用 Precision、Recall、F1 分數、AUC 作為評估指標，也有使用 SHAP (SHapley Additive exPlanations) 分析特徵貢獻與模型可解釋性。

| Data sets | Models | Precision | Recall | F1-score | AUC |
|--------------|----------------------|--------------|--------------|--------------|--------------|
| Training set | Demographic | 0.614 ± 0.03 | 0.613 ± 0.03 | 0.612 ± 0.03 | 0.628 ± 0.03 |
| | Child-adolescent_RF | 0.806 ± 0.06 | 0.762 ± 0.09 | 0.750 ± 0.09 | 0.884 ± 0.02 |
| | Child-adolescent_SLR | 0.838 ± 0.01 | 0.834 ± 0.00 | 0.833 ± 0.00 | 0.879 ± 0.01 |
| | Child-adolescent_SVM | 0.857 ± 0.01 | 0.852 ± 0.01 | 0.851 ± 0.01 | 0.885 ± 0.01 |
| | Child-adolescent_MLP | 0.855 ± 0.02 | 0.851 ± 0.01 | 0.850 ± 0.01 | 0.907 ± 0.02 |
| | Combined_RF | 0.795 ± 0.04 | 0.792 ± 0.04 | 0.792 ± 0.04 | 0.861 ± 0.01 |
| | Combined_SLR | 0.834 ± 0.01 | 0.830 ± 0.01 | 0.829 ± 0.01 | 0.882 ± 0.01 |
| | Combined_SVM | 0.841 ± 0.02 | 0.837 ± 0.02 | 0.836 ± 0.02 | 0.886 ± 0.01 |
| | Combined_MLP | 0.850 ± 0.03 | 0.847 ± 0.03 | 0.847 ± 0.03 | 0.901 ± 0.02 |
| Test set | Demographic | 0.798 | 0.495 | 0.591 | 0.530 |
| | Child-adolescent_RF | 0.872 | 0.763 | 0.803 | 0.835 |
| | Child-adolescent_SLR | 0.913 | 0.845 | 0.868 | 0.879 |
| | Child-adolescent_SVM | 0.920 | 0.814 | 0.846 | 0.876 |
| | Child-adolescent_MLP | 0.889 | 0.784 | 0.820 | 0.866 |
| | Combined_RF | 0.879 | 0.804 | 0.832 | 0.840 |
| | Combined_SLR | 0.900 | 0.845 | 0.865 | 0.864 |
| | Combined_SVM | 0.907 | 0.814 | 0.845 | 0.876 |
| | Combined_MLP | 0.875 | 0.784 | 0.817 | 0.857 |

▲ Performance of all models predicting depressive symptoms on the training and test sets

4.4 Results

作者在三種資料集的模型架構下分別使用各種 Machine Learning Model 的結果，主要可以發現在 Child-Adolescent Model 表現最佳，四種 Machine Learning Model 中以 SLR 達到最高 AUC 0.879；而相比之下可以發現，Demographic Model 的 AUC 只有 0.530；然而在 Combined Model 中看不出來在哪個指標下有比較好的結果。而最後作者有使用 SHAP 分析顯示：社會信任、學業壓力、正向教養行為 為預測憂鬱最重要的三個特徵。



▲ SHAP summary plots and feature importance visualizations for the child-adolescent models

5. Academic-related stressors predict depressive symptoms in graduate students: A machine learning study

5.1 Target Problem

此研究室針對在疫情期間研究生心理健康惡化的情況，傳統都是以統計學方法分析壓力與心理疾病的相關性，只是比較少能夠具體指出到底是哪些因子對憂鬱症影響程度多寡，而此篇目的是想了解能否應用 Machine Learning Model 方法，根據研究生們在疫情期間的學業相關的壓力，來預測是否有憂鬱的情況產生，並探討哪些相關壓力對憂鬱的影響甚大。

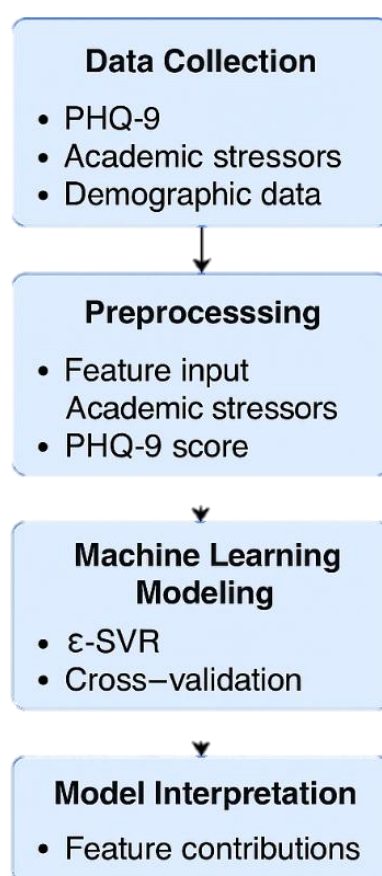
5.2 Dataset

此研究 Dataset 未公開，但是資料集來自於在 2021 年 10 月到 2022 年 3 月間，巴西里約熱內盧大學由四個研究所的學生建立的線上問卷調查資料，此時還是處於 COVID-

19 疫情期間的遠距教學狀態。有效填答樣本數為 172 名研究生，有 67.4% 為女性，年齡落在 20-50 歲之間，學位層級分為碩士和博士，種族分為白人 64%、黑人、混血其他族群占比 36%。

其中 dataset 裡包含三個主要的變項：(一) 心理健康量表，PHQ-9 憂鬱量表，一共有 9 題，每題以 0-3 評分，故總分為 0-27 分，一旦 $PHQ-9 \geq 9$ 被視為「可能有憂鬱症狀」。(二) 學業壓力因子有 10 個(對學業進度與未來計劃的擔憂、對學業表現的擔憂、對學業截止日期的擔憂、財務困難、社交與人際關係困擾、家庭關係困擾、適應遠距學習的困難、與指導教授的關係問題、學習資源取得困難、家務負擔壓力)，每項壓力都已 0-4 分 Likert 五點量表評分 (0 = 完全沒有，4 = 極度困擾)。(三) 人口統計特徵(性別、年齡種族與族群、修習學位層級)。

5.3 Overview of the framework



▲ Flowchart

Data Preprocess: 首先將從 Google forms 收集到的資料進行移除缺失值、不合格及無效性填答。初步利用一些統計軟體來計算中位數、IQR、比率作為描述性統計。

Feature Engineering: 並沒有做什麼特別處理，直接輸入 10 個與學業相關的壓力來源 Likert 分數，把 PHQ-9 總分（憂鬱症狀程度）當作目標標籤，至於性別、年齡、種族等

人口特徵使用交叉驗證平衡處理。

此篇採用 ϵ -SVR (epsilon-insensitive Support Vector Regression)，這是 SVM 的回歸版本，主要特點是使用 ϵ -insensitive 損失函數，只有當預測誤差超過 ϵ 時，才會被計入損失，這有助於減少對小誤差的過度反應。而作者後續也採用 5-fold 交叉驗證作為主策略，後續也加入 2-fold (半數交叉驗證) 做為輔助確認模型穩定性，避免 overfitting 現況產生。

至於作者採用的評估指標有皮爾森相關係數 r ，預測值與實際 PHQ-9 分數的線性相關程度、 R^2 預測的結果和解釋原始分數變異的比例、NMSE 為標準化均方誤差。

5.4 Results

| | | |
|---|-----|--------|
| Gender (n, %) | | |
| Female | 116 | 67.4 % |
| Male | 56 | 32.6 % |
| Race and ethnicity (n, %) | | |
| White | 110 | 64.0 % |
| Black, mixed & others | 62 | 36.0 % |
| Level of academic course (n, %) | | |
| Master's | 71 | 41.3 % |
| Doctorate | 101 | 58.7 % |
| PHQ-9 (med, IQR) | 14 | 11 |
| Probable depression diagnosis (n, %) | | |
| No | 46 | 26.7 % |
| Yes | 126 | 73.3 % |

▲ Sample characteristics (n=172). Rio de Janeiro, 2021–22

由上圖可以看到在疫情期間研究生的憂鬱風險達到 126 人 (73.3%)。

| Regression Models | r | p -value | R^2 | p -value | NMSE | p -value |
|-----------------------------|------|------------|-------|------------|------|------------|
| ϵ -SVR ($k = 5$) | 0.44 | 0.001 | 0.21 | 0.001 | 0.84 | 0.001 |
| ϵ -SVR ($k = 2$) | 0.47 | 0.001 | 0.22 | 0.001 | 0.78 | 0.001 |

▲ Measurements of agreement between the actual and predicted depressive symptoms based on academic-related stressors (n=172)

這邊數據可以明顯看出來在 $r=0.44$ 、 $R^2=0.21$ 、NMSE0.84，顯示 ϵ -SVR 模型能有效預測研究生的憂鬱症狀，並驗證學業壓力與心理健康之間存在顯著可預測性關係。

| Stressors/difficulties (n, %) | | |
|--|-----|--------|
| Fear and worry about academic progress and plans | 155 | 90.1 % |
| Fear and worry about academic performance | 136 | 79.1 % |
| Fear and worry about academic deadlines | 135 | 78.5 % |
| Financial difficulties | 95 | 55.2 % |
| Social/interpersonal relationships | 83 | 48.3 % |
| Family relationships | 76 | 44.2 % |
| The need to perform housework | 72 | 41.9 % |
| Difficulty adapting to distance learning | 56 | 32.6 % |
| Relationship with supervisors | 42 | 24.4 % |
| Difficulty in accessing study materials | 27 | 15.7 % |

▲ The most prevalent stressors/difficulties, described as interfering "quite a bit" or "extremely" with their lives and academic activities (n=172). Rio de Janeiro, 2021–22

上圖可分辨在 Fear and worry about academic progress and plans 高達 90.1%、Fear and worry about academic performance 也有 79.1%、Fear and worry about academic deadlines 為 78.5%。

6. Machine learning models for predicting risk of depression in Korean college students: Identifying family and individual factors

6.1 Target Problem

在韓國大學生普遍憂鬱情況較高，然而社會上對於這種疾病往往不是很善良，因此得到此症狀的學生經常會忽略病情不去就醫，因此增加許多未及時治療的風險。故此篇論文提出眾多 Machine Learning 技術，結合個人背景及家庭層面的資料（包含父母與大學生本人），建立可以有效預測大學生憂鬱症風險的 Model，並同時出與憂鬱症風險高度相關的關鍵家庭與個人因子。

6.2 Dataset

作者提到他們的資料集是從學校裡面或一些宗教場所，透過發放問卷或傳單的方式自行收集，並不是官方的數據來源，因此也未提供公開的資料集。而這份資料是包含韓國的大學生家庭，包括父親、母親和大學生本人（每個家庭三人一組，共 171 個家庭）。

資料內容特徵包含:人口統計學(年齡、性別、宗教信仰、教育背景、父母婚姻年數、家庭成員數、子女人數)、家庭關係互動(與家庭成員相處時間（每週小時數）、參與社交團體與滿意度、家庭收入滿意度)、健康情形(父母健康狀況、睡眠品質、飲酒與吸菸行為、有無從事有氧與無氧運動、自我感知的心理健康與身體健康)、心理特徵量表(外向性、親和性、盡責性、神經質、開放性、精神性量表、自我超越量表)、依附風格與家庭功能(安全型、疏離型、焦慮型、恐懼-迴避型、父母教養行為量表、家庭適

應與凝聚力量表)、憂鬱症量表。

6.3 Overview of the framework

Data Preprocess:將收集到的 171 組家庭(共 513 人) 以大學生憂鬱症狀表 CES-D Score ≥ 13 定義為「憂鬱風險組」, <13 為「正常組」, 並且也移除 26 組缺失值的家庭數據, 隨後再進行標準化。

後續分別使用三種機器學習 Model 來建模, 分別為 SLR, LASSO 可以懲罰進行特徵篩選, 同時避免 overfitting、SVM 尋找最佳超平面將「憂鬱風險組」與「正常組」分開、RF 可以擁有多棵決策樹, 利用多數投票機制進行分類。最終使用 Accuracy、Precision、Sensitivity、Specificity、F1 Score、AUC 做評估。

6.4 Results

根據三種模型結果可以看出在 Random forest 這個 Model, 不論是 Accuracy、Precision、Sensitivity、Specificity、F1 都是最高的; 只有 SLR 在 AUC 有達到最高 90.32%表示排序能力最好。

作者主要從 SLR 找出哪些特徵是重要影響因素, 只要 p-value < 0.05 以下就是為特別顯著影響, 可以發現 Father's severity of cancer、Mather's severity of respiratory diseases、College student's self-perceived mental health、College student's conscientiousness、College student's neuroticism, 這些特徵都是從 SLR 選出來且 p 值有顯著的影響。

| | Accuracy | Precision | Sensitivity | Specificity | F1 score | AUC |
|-----------------|----------|-----------|-------------|-------------|----------|--------|
| Sparse logistic | 0.7843 | 0.7143 | 0.7500 | 0.8065 | 0.7317 | 0.9032 |
| SVM | 0.8039 | 0.7500 | 0.7500 | 0.8387 | 0.7500 | 0.7944 |
| Random forest | 0.8627 | 0.8059 | 0.8500 | 0.8710 | 0.8274 | 0.8605 |

▲ Performance of machine learning algorithms

| | Estimate | Std error | Exp coef | Low CL | Upper CL | p-Value |
|---|----------|-----------|----------|--------|----------|---------|
| Father's severity of cancer | -1.646 | 0.704 | 0.193 | 0.041 | 0.675 | 0.019 |
| Father's aerobic exercise | -0.283 | 0.228 | 0.753 | 0.462 | 1.147 | 0.213 |
| Mather's severity of cardiovascular disease | 0.770 | 0.591 | 2.159 | 0.751 | 8.154 | 0.193 |
| Mather's severity of respiratory diseases | 1.216 | 0.520 | 3.373 | 1.345 | 11.313 | 0.019 |
| College student's self-perceived mental health | -2.377 | 0.638 | 0.093 | 0.021 | 0.272 | 0.000 |
| Father's fearful-avoidant attachment | 0.496 | 0.288 | 1.643 | 0.958 | 3.024 | 0.085 |
| Mather's depression | -0.016 | 0.037 | 0.984 | 0.914 | 1.060 | 0.660 |
| College student's conscientiousness | -1.337 | 0.610 | 0.263 | 0.071 | 0.805 | 0.028 |
| College student's neuroticism | 1.391 | 0.519 | 4.021 | 1.562 | 12.420 | 0.007 |
| College student's fearful-avoidant attachment | 0.330 | 0.219 | 1.391 | 0.914 | 2.183 | 0.131 |
| Mother's care | -0.079 | 0.072 | 0.924 | 0.797 | 1.060 | 0.271 |
| College student's income satisfaction-neutral | | | | | | |
| College student's income satisfaction-dissatisfaction | 1.666 | 0.859 | 5.290 | 1.069 | 33.254 | 0.053 |
| College student's income satisfaction-satisfaction | 0.831 | 0.965 | 2.295 | 0.351 | 16.548 | 0.389 |

▲ Logistic regression analysis results

7. Predicting Student Depression Using Machine Learning

7.1 Target Problem

學生憂鬱已成為現代社會日益嚴重的問題，長期憂鬱會影響學業表現、心理健康、同儕交流等等，而傳統臨床評估的方法主觀性值較高，需要專業介入等限制，且許多學生因社會壓力而不願求助醫生或諮商師，所以作者想探索如何運用 Machine Learning Model，從學生的人口統計學特徵、學業表現與生活型態資料，來自動預測學生是否有憂鬱症的傾向。

7.2 Dataset

作者在文中只有提到是從 Kaggle，但看裡面的欄位跟我們所使用的資料集一樣，我們這資料集是 2025/3 月才上傳自 Kaggle，而作者發在 ResearchGate 也是 2025/3，我們認為應該就是跟我們一樣的資料集 (<https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset/code>)。

主要資料夾包含是否曾有自殺想法、財務壓力、工讀/讀書時數、學業滿意度、飲食習慣、工作滿意度、性別、城市、年齡、學業壓力、睡眠時間、工作壓力、CGPA，職業、是否憂鬱、家人是否有心理集病史。

7.3 Overview of the framework

Data Preprocess: 作者在前處理的階段並不是直接移除缺失值，若為數值的資料型態，作者用計算中位數來填補遺失值;若為類別型資料，作者用眾數來填補。接下來使用 Z-score 絕對值只要>3 就當作異常值將其剔除，接者將類別型態的資料用 label encoding 將其編碼成數字，最後使用 StandardScaler 標準化數值特徵。

Feature Engineering: 使用 Pearson 相關係數分析篩選出與憂鬱症最相關的特徵(學業壓力、財務壓力、是否曾有自殺想法)。

後續分別使用三種機器學習 Model 來建模，分別為 Logistic Regression，Random Forest Gradient Boosting。目標變數是預測憂鬱症狀態 (Depression 0 = 無憂鬱、1 = 有憂鬱)，最終使用 Accuracy、Precision、Recall、F1 Score、ROC-AUC 做評估。

7.4 Results

| Model Performance Comparison: | | | | | |
|--------------------------------------|---------------------|----------|-----------|----------|----------|
| | Modal | Accuracy | Precision | Recall | F1 Score |
| 1 | Logistic Regression | 0.844126 | 0.848765 | 0.887609 | 0.867752 |
| 2 | Random Forest | 0.840897 | 0.848158 | 0.881694 | 0.864601 |
| 3 | Gradient Boosting | 0.845202 | 0.848206 | 0.890722 | 0.868945 |

▲ Model Performance Comparison

從上圖實驗結果可看出 Gradient Boosting 在 Accuracy、Recall、F1-score 及 ROC AUC 等評估指標上都優於其他 Model，看的出來在 Gradient Boosting 區分憂鬱與非憂鬱學生上有高度穩定性。相比之下在 Logistic Regression 與 Random Forest 的準確率略低，分別為 84.41%與 84.08%。

從特徵重要性分析來看，心理壓力相關的特徵在憂鬱症的預測有著至關重大的影響。尤其有三個最具影響力的是「是否曾有自殺想法」、「學業壓力」以及「財務壓力」，而睡眠時數則呈現與憂鬱症的負相關。

8. Predictive machine learning models for assessing Lebanese university students' depression, anxiety, and stress during COVID-19

8.1 Target Problem

之前在 COVID-期間曾經嚴重的造成學生心理壓力徒增的情況，不論是在憂鬱、焦慮及壓力狀況的學生逐漸地增加，作者把焦點關注在黎巴嫩大學的學生上，希望可以不依賴生理裝置測量到的數據，而是透過機器學習 Model 來預測，藉由這樣的方式來提早找出可能含有心理健康風險的人，並再提供給學校輔導單位介入調查。

8.2 Dataset

作者未提供資料集，不過他們是在 2021 年 11 月到 2022 年 2 月，當時正值 Omicron 變異株出現時期，分別在 American University of Beirut、Lebanese University 兩所學校利用線上問卷的方式，最後共有 329 位 18 歲以上的學生填答，欄位名稱包含：年齡、性別、是否有交往對象、就讀學位、COVID-19 疫情後 GPA 變化、家庭收入是否高於最低工資、整體自評健康狀況、宗教是否在日常生活中扮演重要角色、是否相信 COVID-19 病毒/疫苗背後有陰謀、是否遵守 COVID-19 防疫措施、是否曾感染 COVID-19、是否曾接受私人心理諮詢、飲食健康程度、香菸抽煙行為變化、水煙抽煙行為變化、酒精飲用行為變化、身體活動量變化、每日睡眠時間等。

8.3 Overview of the framework

Data Preprocess: 若資料中含有遺失值，會用眾數來填補。另外也使用 Stratified Shuffle Split，確保在訓練集與測試集中，各類別（如憂鬱/非憂鬱）樣本的比例一致，解決資料不平衡的問題。

Feature Engineering: 文中沒特別提到，但我們認為應該也是用 label encoding 跟 標準化的方式去處理。

後續分別使用八種機器學習 Model 來建模，分別為 Logistic Regression，Random Forest SVM、KNN、AdaBoost、XGBoost、Naïve Bayes、MLP。目標變數是預測 PHQ-9 憂鬱量表只要 ≥ 10 判定為有重大憂鬱傾向、BAI 焦慮量表只要 ≥ 16 判定為有臨床焦慮、PSS-10 壓力量表只要 ≥ 27 判定為高壓力，最終使用 Accuracy、Precision、F1 Score、Sensitivity、AUC 做評估。

8.4 Results

| Model | AUC (%) | Sensitivity (%) | Precision (%) | F1-score (%) |
|---------------------|---------|-----------------|---------------|--------------|
| MLP | | | | |
| Depression | 73.90 | 74.28 | 63.41 | 68.42 |
| Anxiety | 72.60 | 73.68 | 63.63 | 68.29 |
| Stress | 70.30 | 66.66 | 47.05 | 55.17 |
| Logistic regression | | | | |
| Depression | 74.12 | 55.00 | 78.00 | 64.00 |
| Anxiety | 74.89 | 74.00 | 75.00 | 74.00 |
| Stress | 66.51 | 7.00 | 100.00 | 13.00 |
| AdaBoost | | | | |
| Depression | 76.25 | 65.00 | 72.00 | 68.00 |
| Anxiety | 74.89 | 100.00 | 53.00 | 69.00 |
| Stress | 72.96 | 33.00 | 69.00 | 45.00 |
| Random forest | | | | |
| Depression | 78.27 | 60.00 | 75.00 | 67.00 |
| Anxiety | 69.93 | 62.00 | 70.00 | 66.00 |
| Stress | 72.42 | 12.00 | 75.00 | 19.00 |
| XGBoost | | | | |
| Depression | 75.55 | 51.00 | 79.00 | 68.00 |
| Anxiety | 67.67 | 71.00 | 65.00 | 68.00 |
| Stress | 66.87 | 30.00 | 72.00 | 42.00 |
| SVM | | | | |
| Depression | 74.36 | 67.00 | 65.00 | 66.00 |
| Anxiety | 74.94 | 72.00 | 72.00 | 72.00 |
| Stress | 72.37 | 3.00 | 100.00 | 7.00 |
| Naïve Bayes | | | | |
| Depression | 74.12 | 62.00 | 71.00 | 65.00 |
| Anxiety | 76.37 | 77.00 | 75.00 | 76.00 |
| Stress | 63.36 | 37.00 | 47.00 | 41.00 |
| KNN | | | | |
| Depression | 66.63 | 35.00 | 66.00 | 46.00 |
| Anxiety | 61.05 | 53.00 | 57.00 | 62.00 |
| Stress | 63.84 | 12.00 | 60.00 | 18.00 |

Abbreviations: AUC, area under curve; F1-score, harmonic mean between precision and recall.

▲ Models' Performance Measurements

機器學習模型評估結果顯示，在憂鬱症狀預測中，Random Forest 表現最好，AUC 為 78.27%，在焦慮症狀預測中 Naïve Bayes 最好，AUC 為 76.37%，在壓力的話是 AdaBoost 最高，AUC 為 72.96%。SVM、MLP、XGBoost、Logistic Regression 也都有到 70%以上 AUC，也還算可接受的預測能力。

而在心理健康評估方面，學生的 PHQ-9 有 10.18 分，BAI 有 18.81 分，PSS-10 有 21.97 分。52.3%的學生有輕度至中度憂鬱症狀，24.6%有重度憂鬱；42.9%有輕度至中度焦慮，29.3%有重度焦慮；61.7%有輕度至中度壓力，27.6%達到高壓力水準。整體而言來說，中度至重度憂鬱、焦慮與壓力的占比分別為 75.9%、72.2%、89.3%。

9. Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine

9.1 Target Problem

在 COVID-19 期間隔離的阿根廷大學生們容易有抑鬱的情況產生，目的是能預測 COVID-19 隔離期間阿根廷大學生出現憂鬱症狀的風險，作者也希望找出那些心理疾

病、社會或個人特徵能夠準確並有效預防憂鬱症發生，就能提前預防。

9.2 Dataset

此篇論文並未提供資料集，是作者們在隔離期間利用線上問卷調查收集的，參加的學生都是自願填答，包含來自阿根廷不同地區的大學生，問卷內容包含年齡、性別、是否與家人或朋友同住、隔離天數、隔離期間的生活改變、是否有陪伴者、焦慮感受、壓力感受、孤獨感、憂鬱症狀、睡眠模式變化、飲食習慣變化、運動頻率、憂鬱程度、焦慮程度、壓力量表分數。

9.3 Overview of the framework

Data Preprocess: 若問卷中含有遺失值，會直接刪除該筆資料。

Feature Engineering: 使用 Label Encoding 將類別型資料編碼成數字。文章中有提到用統計方法來篩選特徵，更詳情節文章中就沒提到了。

後續分別使用五種機器學習 Model 來建模，分別為 Logistic Regression，Random Forest SVM、KNN、Decision Tree。目標變數是預測是否有憂鬱情況，最終使用 Accuracy、Specificity、F1 Score、Sensitivity 做評估。

9.4 Results

這項研究建立了很多 Machine Learning Model 來預測 COVID-19 隔離時期阿根廷大學生出現憂鬱症狀的，而後經過交叉驗證評估各模型表現後，發現 RF 模型的表現最好，尤其在 Accuracy, Sensitivity 和 Specificity 以及 F1 Score 上。

RF Accuracy 大約是 85%，Sensitivity 和 F1Score 也在 80%-85%之間，而 Specificity 大約是 85%-90%。這個結果表示，這一模型在識別有憂鬱症狀的人及排除沒有憂鬱症狀的人方面都有好的表現。

然而作者在進一步的特徵重要性分析發現，孤獨感、焦慮水平、壓力感受、睡眠品質變化及隔離時是否有陪伴是影響憂鬱症狀預測重要的一些特徵。此篇論文證實，機器學習方法能有效協助在疫情情境下早期辨識高風險群體，為心理健康防治提供依據。

10. Comparison of the Performance of Machine Learning-based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh: A Result of the First Wave of the COVID-19 Pandemic

10.1 Target Problem

孟加拉在 2020 年 5 月到六月期間 COVID-19 疫情初次大規模爆發、封鎖期間學校關閉、社交也隔離，學生壓力劇增的階段，因此作者想了解透過機器學習 Model 去預

測在這段時間出現憂鬱症及焦慮症狀的風險比例。

10.2 Dataset

作者們在 2020 年 5 月至 6 月收集孟加拉大學的學生們的數據，主要是透過線上問卷的方式，總共收集到 292 筆有效問卷資料，這些資料也並未公開，而問卷上的欄位資料有：人口統計特徵(年齡、性別、婚姻狀態、居住地區、教育階段)、生活方式特徵(運動習慣、睡眠時間、飲食規律性、吸菸狀態、飲酒狀態)、健康狀態特徵(是否有慢性病、是否有心理健康問題史)、COVID-19 疫情影響特徵(是否本人或親友感染 COVID-19、COVID-19 導致的學業影響、COVID-19 導致的經濟壓力、對未來職涯的擔憂)、學業與經濟壓力特徵(學業壓力感受、家庭經濟壓力感受)、心理健康量表分數(PHQ-9 憂鬱量表分數、GAD-7 焦慮量表分數)。

10.3 Overview of the framework

Data Preprocess: 文章中提到會處理遺失值，但具體怎麼做卻沒說。目標函數是 PHQ-9 和 GAD-7 的得分來分類，只要 $\text{PHQ-9} \geq 10$ 就定義為有憂鬱症狀，而 $\text{GAD-7} \geq 10$ 只要定義為有焦慮症狀，這樣就分為兩類：有無憂鬱和有無焦慮。

Feature Engineering: 作者有使用到 Chi-square test 和 ANOVA test 來選重要特徵。Chi-square test 是用來類別型資料與目標函數是否憂鬱和是否焦慮之間的關聯性，ANOVA 是分辨連續型函數與目標函數之間的差異，都是以 p-value 小於 0.05，代表該特徵與目標函數間有統計上顯著的關聯。

後續分別使用六種機器學習 Model 來建模，分別為 Logistic Regression，Random Forest SVM、KNN、Decision Tree、Naive Bayes。目標變數是預測是否有憂鬱情況，最終使用 Accuracy、Specificity、PPV、NPV、Sensitivity、 κ 做評估。

10.4 Results

以憂鬱準確度來說，從表格上可看出在 Random Forest 在所有指標上表現最好。像 Accuracy 88.66%、Specificity 95.88%、PPV 85.84%、NPV 89.42%、Sensitivity 65.4%、 κ 0.6903。

| Table 4: Performance indicators of all five machine learning algorithms to predict depression status | | | | | | |
|--|-------------|-------------|-------------|-------------|-------------|-------------|
| | Algorithms | | | | | |
| | LR | RF | SVM | LDA | KNN | NB |
| Training set | | | | | | |
| Accuracy (%) | 75.63 | 89.76 | 74.62 | 75.31 | 89.38 | 70.79 |
| 95% CI | 73.44-77.72 | 88.17-91.21 | 72.41-76.75 | 73.12-77.42 | 87.77-90.86 | 68.49-73.02 |
| κ | 0.2537 | 0.7220 | 0.1835 | 0.2474 | 0.7099 | 0.1504 |
| Sensitivity (%) | 28.17 | 71.36 | 19.95 | 28.17 | 69.72 | 26.06 |
| Specificity (%) | 92.97 | 96.48 | 94.60 | 92.54 | 96.57 | 87.14 |
| PPV | 59.40 | 88.12 | 57.43 | 57.97 | 88.13 | 42.52 |
| NPV | 77.98 | 90.22 | 76.38 | 77.90 | 89.72 | 76.33 |
| Test set | | | | | | |
| Accuracy (%) | 74.48 | 88.66 | 74.1 | 74.29 | 88.28 | 68.24 |
| 95% CI | 70.54-78.14 | 85.64-91.23 | 70.15-77.79 | 70.34-77.97 | 85.23-90.9 | 64.09-72.19 |
| κ | 0.1968 | 0.6903 | 0.1664 | 0.1931 | 0.6769 | 0.1027 |
| Sensitivity (%) | 22.69 | 68.79 | 19.14 | 22.69 | 66.67 | 25.53 |
| Specificity (%) | 93.29 | 95.88 | 94.07 | 93.04 | 96.13 | 83.76 |
| PPV | 55.17 | 85.84 | 54.00 | 54.23 | 86.24 | 36.36 |
| NPV | 76.85 | 89.42 | 76.20 | 76.80 | 88.81 | 75.58 |

LR: Logistic regression, RF: Random forest, SVM: Support vector machine, LDA: Linear discriminant analysis, KNN: K-nearest neighborhood, NB: Naïve Bayes, PPV: Positive predictive value, NPV: Negative predictive value, CI: Confidence interval

▲ Performance indicators of all five machine learning algorithms to predict depression status

至於焦慮預測表現上 Random Forest 和 KNN 表現都非常好，其中，Random Forest 在 Test Accuracy 91.3%、Specificity 97.55%、PPV 89.47%、NPV 91.71%、 κ 0.7334。至於 KNN 在 Test Accuracy 91.3%是一樣的，但 Sensitivity 75%比 Random Forest 70.45%還稍微高一點點，不過整體指標很接近 Random Forest，所以文章最後還是以 Random Forest 為焦點，因為整體平衡性與穩定性最佳。

| Table 5: Performance indicators of the five machine learning algorithms to predict anxiety status | | | | | | |
|---|-------------|-------------|-------------|-------------|-------------|-------------|
| | Algorithms | | | | | |
| | LR | RF | SVM | LDA | KNN | NB |
| Training set | | | | | | |
| Accuracy (%) | 80.78 | 93.09 | 92.27 | 81.78 | 92.9 | 76.26 |
| 95% CI | 78.76-82.69 | 91.73-94.29 | 90.85-93.54 | 79.8-83.65 | 91.53-94.11 | 74.09-78.33 |
| κ | 0.2565 | 0.7852 | 0.7525 | 0.2775 | 0.7834 | 0.1448 |
| Sensitivity (%) | 22.54 | 76.59 | 70.81 | 22.25 | 78.61 | 28.09 |
| Specificity (%) | 96.95 | 97.67 | 98.23 | 98.31 | 96.87 | 95.18 |
| PPV | 67.24 | 90.14 | 91.76 | 78.57 | 87.46 | 31.81 |
| NPV | 81.84 | 93.76 | 92.38 | 81.99 | 94.22 | 78.85 |
| Test set | | | | | | |
| Accuracy | 78.07 | 91.3 | 91.49 | 79.58 | 91.3 | 73.53 |
| 95% CI | 74.3-81.53 | 88.57-93.56 | 88.78-93.73 | 75.89-82.94 | 88.57-93.56 | 69.56-77.25 |
| κ | 0.1583 | 0.7334 | 0.7333 | 0.1909 | 0.7399 | 0.1827 |
| Sensitivity (%) | 14.87 | 70.25 | 67.77 | 14.87 | 73.55 | 24.13 |
| Specificity (%) | 96.81 | 97.55 | 98.53 | 98.77 | 96.57 | 94.11 |
| PPV | 58.06 | 89.47 | 93.18 | 78.26 | 86.41 | 17.24 |
| NPV | 79.31 | 91.71 | 91.16 | 79.64 | 92.49 | 76.80 |

CI: Confidence interval, LR: Logistic regression, RF: Random forest, SVM: Support vector machine, LDA: Linear discriminate analysis, KNN: K-nearest neighbors, NB: Naïve Bayes, PPV: Positive predictive value, NPV: Negative predictive value

▲ Performance indicators of the five machine learning algorithms to predict anxiety status

Summarized Table

| Paper Number | Publication Year | Target Problem | Approach/Technique | Datasets | Results & Contribution |
|--------------|------------------|----------------------|--|--|---|
| #1 | 2022 | Depression & Anxiety | 5 ML Models (SVM, RF, ANN, DT, NB) | West Bank school dataset | SVM 表現最佳 (Depression: 92.5%, Anxiety: 92.4%)，首次使用學校調查資料預測心理狀況。 |
| #2 | 2024 | Depression | Cross-sectional & Time-series ML (SVC, RF, MLP, RNN, LSTM) + RFE | ALSPAC longitudinal cohort | 預測青少年 12–18 歲憂鬱 (平均 Accuracy: 0.64, Recall: 0.59) |
| #3 | 2025 | Depression & Anxiety | XGBoost + SHAP | UF EHR + Social Determinants of Health | 預測 AUC: Depression (0.81), Anxiety (0.80), Combined (0.78) |
| #4 | 2025 | Depression | RF、SVM、MLP、SLR、SMOTE、mRMR、5-fold cross-validation | Peking University Institute of Social Science Survey | Child-Adolescent Model 表現最佳，SLRAUC 0.879 |
| #5 | 2025 | Depression | ε -SVR、5-fold cross-validation | Questionnaire survey (Non-public) | $r=0.44$ 、 $R^2=0.21$ 、NMSE0.84 驗證學業壓力與心理存在顯著可預測 |

| | | | | | |
|-----|------|-----------------------------|---|------------------------------------|---|
| | | | | | 性關係 |
| #6 | 2022 | PHQ-9 depression score | LASSO、SVM、RF | Questionnaire survey (Non-public) | Random forest 這個 Model，不論是 Accuracy、Precision、Sensitivity、Specificity、F1 都是最高 |
| #7 | 2025 | Depression | Logistic Regression，Random Forest Gradient Boosting | Kaggle(Student Depression Dataset) | Gradient Boosting 在 Accuracy、Recall、F1-score 及 ROC AUC 都是最高 |
| #8 | 2021 | Depression, Anxiety, Stress | LR、RF、SVM、KNN、AdaBoost、XGBoost、Naïve Bayes、MLP | Questionnaire survey (Non-public) | Random Forest 表現最好，Depression AUC 為 78.27% |
| #9 | 2021 | Depression | Logistic Regression，Random Forest SVM、KNN、Decision Tree | Questionnaire survey (Non-public) | RF 在 Accuracy, Sensitivity Specificity F1 Score 上最好 |
| #10 | 2022 | Depression | Logistic Regression，Random Forest SVM、KNN、Decision Tree、Naive Bayes | Questionnaire survey (Non-public) | 焦慮預測表現上 Random Forest 和 KNN 表現都非常好 |

Summary

➤ #1

● Technical Side Summary

Insights: 作者使用 SVM 在預測憂鬱(0.925)及焦慮(0.924)表現最佳，使用的資料涵蓋心理、生理、社會三個層面。x

Possible improvements and extensions: 可以進一步使用時間序列的資料，讓模型學習學生成長變化，也可以加入特徵重要性排序提升解釋性。

- Application Side Summary

Insights: 結果顯示 ML 技術具有整合學生多元資料以辨識心理健康的能力，校園若能引入此技術可以系統化評估學生心理健康並早期介入。

Possible improvements and extensions: 真正在使用時須針對不同地區不同文化背景重新設計問卷篩選特徵。

- #2

- Technical Side Summary

Insights: 使用 ALSPAC 長期追蹤資料集，並同時比較 cross-sectional 與 time-series 模型，選擇最佳模行將特徵數精簡至 39 個。

Possible improvements and extensions: time-series 模型可以引入 attention 機制，融合 Transformer 架構。

- Application Side Summary

Insights: 能作為兒童健康檢查工具，幫助評估孩童心理健康。

Possible improvements and extensions: 能與學校輔導系統串接，建立一套完整的心理健康評估系統。

- #3

- Technical Side Summary

Insights: 使用美國健保資料庫(EHR)，進行敏感性分析後發現即使不使用 SDoH，模型仍表現相似，表示醫療數具足夠支持模型預測。

Possible improvements and extensions: 龐大的資料沒有針對族群進行子群體分析，可能出現種族或性別偏誤。

- Application Side Summary

Insights: 醫療資料結合社會資料可支援實際應用，基層醫療機構就可判別心理健康風險。

Possible improvements and extensions: 可與政府政策連動，擴大應用場景。

- #4

- Technical Side Summary

Insights: 找出童年時期特徵: 與同儕的關係品質、父母缺席; 青少年時期特徵: 社會信任、學業壓力、網路娛樂重要性、正向教養行為。這六個特徵在 SLR 中帶來最高 AUC 高達 0.879，而 SHAP 分析確認過他的影響力與貢獻。

Possible improvements and extensions: 因資料有多時間點性質，所以也可用 LSTM、

Temporal Fusion Transformer 等 sequence-aware 模型看看。

- Application Side Summary

Insights: 作者使用的 CFPS dataset 長期追蹤型問卷資料，結構化且具時間性，涵蓋心理、行為、家庭與社經背景等因素。

Possible improvements and extensions: 未來可建置互動式決策支援系統，讓諮詢師、社工等，能夠即時查詢預測結果與風險因子解釋。

➤ #5

- Technical Side Summary

Insights: ϵ -SVR Model 提供每個特徵權重因子，可以清楚知道哪些學業壓力對於憂鬱症狀來說影響最大，也幫助模型可解釋性

Possible improvements and extensions: 評估指標有點少，未來也可以擴展以 MAE、 R^2 等。

- Application Side Summary

Insights: 壓力因子來自 Meo 及 Wang 的研究，項目明確結構也穩定，可以幫助未來其他學校使用。

Possible improvements and extensions: 未來可以加上行為、感測器或社交媒體的資料來豐富資料集，來增強預測能力

➤ #6

- Technical Side Summary

Insights: RF 實現了最高 86.27% 預測準確率和 F1 score 82.74%，而 SLR 有最好 AUC 0.9032，至於 LASSO 的 SLR 在高維資料集中進行特徵選擇，提供可解釋的係數。

Possible improvements and extensions: 可以使用 Cross-validation 來提升 Model 的泛化能力，也可以嘗試用 MLP 等深度模型，可以更有效地捕捉複雜的非線性模式。

- Application Side Summary

Insights: 作者很特別的做法是把家庭背景也考慮進來，能捕捉世代和文化影響，這在同類研究中算是很少見。

Possible improvements and extensions: 可以嘗試進行縱向資料收集以分析時間模式並建立因果關係。

➤ #7

- Technical Side Summary

Insights: 作者有使用 Pearson 相關係數與特徵重要度分析，其中發現課業壓力、睡眠時數、財務壓力、自殺念頭等特徵對憂鬱傾向有關鍵影響，模型的可解釋性相當高。

Possible improvements and extensions: 未來也可以 CNN、RNN 這種處理多維度或時間序列心理行為資料，提升特徵表達能力與預測準確度。

- Application Side Summary

Insights: 資料來源為 Kaggle 公開問卷型資料，包含人口統計、學業、生活與心理壓力等特徵，比較有公開可以度。

Possible improvements and extensions: 未來也可以使用穿戴式裝置、網路行為的資料納入特徵，亦可以將其他國家的資料考慮進來。

➤ #8

● Technical Side Summary

Insights: 有使用 RF 來做特徵重要性程度分析，最後同時預測三項指標:憂鬱、焦慮與壓力，在特徵重要性分析上可看出，健康狀況和年齡始是所有結果中最重要特徵。

Possible improvements and extensions: 也可以結合性能最好的模型優勢的整合或混合模型。

● Application Side Summary

Insights: 資料來自兩所環境文化不同的大學，有助於模型通用性。

Possible improvements and extensions: 雖模型有很好的預測能力，但未實際部署，之後建議可將部署在學生諮商系統平台上，即早預防憂鬱的發生。

➤ #9

● Technical Side Summary

Insights: 隨機森林表現最好，Accuracy 有 85%、F1score 有 80-85%，展現良好的分類能力。

Possible improvements and extensions: 可以引入 Recursive Feature Elimination、Lasso 等篩選技術，能提升特徵組合的最佳性。

● Application Side Summary

Insights: 特別針對 COVID-19 隔離期間的大學生進行分析，有跟上時代背景下的心理挑戰。

Possible improvements and extensions: 資料來源單一固定、樣本侷限於阿根廷的學生，未來若能加上其他國家的學生資料更能穩定模型。

➤ #10

● Technical Side Summary

Insights: 特徵選擇方法主要使用 Chi-square test 和 ANOVA，可以篩選出與心理健康狀態顯著相關特徵。

Possible improvements and extensions: 可以使用 Cross-validation 來提升 Model 的泛化能力，也可以加入 SHAP 等技術來提升模型解釋性，讓臨床應用更有說服力。

● Application Side Summary

Insights: 作者將童年與青少年時期的特徵都考慮進來，實驗證明早期的特徵對憂鬱預測是有價值。

Possible improvements and extensions: 可以加入比較多學校、多個地區甚至國外的資料，

提升資料多樣性與泛化能力。

➤ Other Findings/Issues

List all findings that could be useful for the term project

1. SMOTE 平衡資料分布。

2. RF Gini index 計算出哪些特徵在模型中權重較高，藉此篩選出更高的特徵或結合產生複合性特徵。

3. Cross-validation 來避免模型 Overfitting。

List all issues that could be encountered for the term project conduction

因我們憂鬱資料分布比為六比四，介於平衡與不平衡交界，因此我們還是有座

SMOTE 來平滑數據;就算使用 RF Gini index 還是發現特徵數量過少或影響力較少等問題;數據若過於極端，模型可能無法有效學習。

Part II: Survey on Extended Topics

- #11: H. Ding, N. Li, L. Li, Z. Xu, and W. Xia, “Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study,” *Journal of Affective Disorders*, vol. 368, pp. 537–546, 2025.
- #12: B. Kadirvelu, T. B. Bel, A. Freccero, M. Di Simplicio, D. Nicholls, and A. A. Faisal, “Digital phenotyping for adolescent mental health: A feasibility study employing machine learning to predict mental health risk from active and passive smartphone data,” *medRxiv*, preprint, Apr. 2024.
- #13: S. K. Sharma, A. I. Alutaibi, A. R. Khan, G. G. Tejani, F. Ahmad, and S. J. Mousavirad, “Early detection of mental health disorders using machine learning models using behavioral and voice data analysis,” *Scientific Reports*, vol. 15, no. 16518, pp. 1–18, 2025.
- #14: G. Sharma, A. M. Joshi, and E. S. Pilli, “DepML: An Efficient Machine Learning-Based MDD Detection System in IoMT Framework,” *SN Computer Science*, vol. 3, no. 394, 2022.
- #15: S. Hassantabar, J. Zhang, H. Yin, and N. K. Jha, “MHDeep: Mental Health Disorder Detection System Based on Body-Area and Deep Neural Networks,” *IEEE Transactions on Computers*, vol. 72, no. 1, pp. 1–12, 2023.

11. Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study

11.1 Target Problem

作者認為青少年由於心智尚未成熟，在經歷壓力性生活事件後(例如 COVID-19)面臨更為明顯的心理健康挑戰，例如憂鬱、焦慮及認知障礙，影響心理健康的因素複雜多樣，使得精準預測較為困難，此篇論文以 Brief Symptom Inventory-53 (BSI-53) 作

為應對結果判定標準，若 BSI-53 的總分小於 63 則視為心理狀況良好，反之則為心理狀況不佳，並透過機器學習技術選取重要特徵，建構預測模型，同時進行內外部驗證，以評估模型的穩健性與泛化能力。

11.2 Dataset

此篇論文的資料集來自兩波針對中國地區青少年的橫斷面線上問卷調查。篩選條件為

年齡介於 14 至 25 歲、自 COVID-19 以來持續居住在中國、能夠閱讀並以中文溝通，其中排除條件為曾經由專業人士臨床診斷患有心理或精神疾病，補充資料中有詳細的問卷內容。第一波資料用做模型訓練及內部驗證，資料收集時間為 2022 年 7 月 1 日至 8 月 30 日，共招募到 3038 位參與者。樣本平均年齡為 19.93 歲（標準差 2.85），其中 678 人（22.3%）年齡小於 18 歲，1560 人（51.3%）為男性。第二波資料用做模型外部驗證，資料收集時間為 2023 年 4 月 30 日至 5 月 7 日，共 539 位參與者。樣本平均年齡為 19.86 歲（標準差 2.69），其中 154 人（28.6%）年齡小於 18 歲，266 人（49.4%）為男性。此研究使用了共 25 個變數，主要包含以下幾個大類別：人口學特徵、壓力性生活事件、應對策略、社會支持、情緒智力，此研究所使用的資料已儲存於資料庫中。作者之後將會公開這些資料，但在使用前須先透過電子郵件聯繫通訊作者說明使用目的。

11.3 Overview of the framework

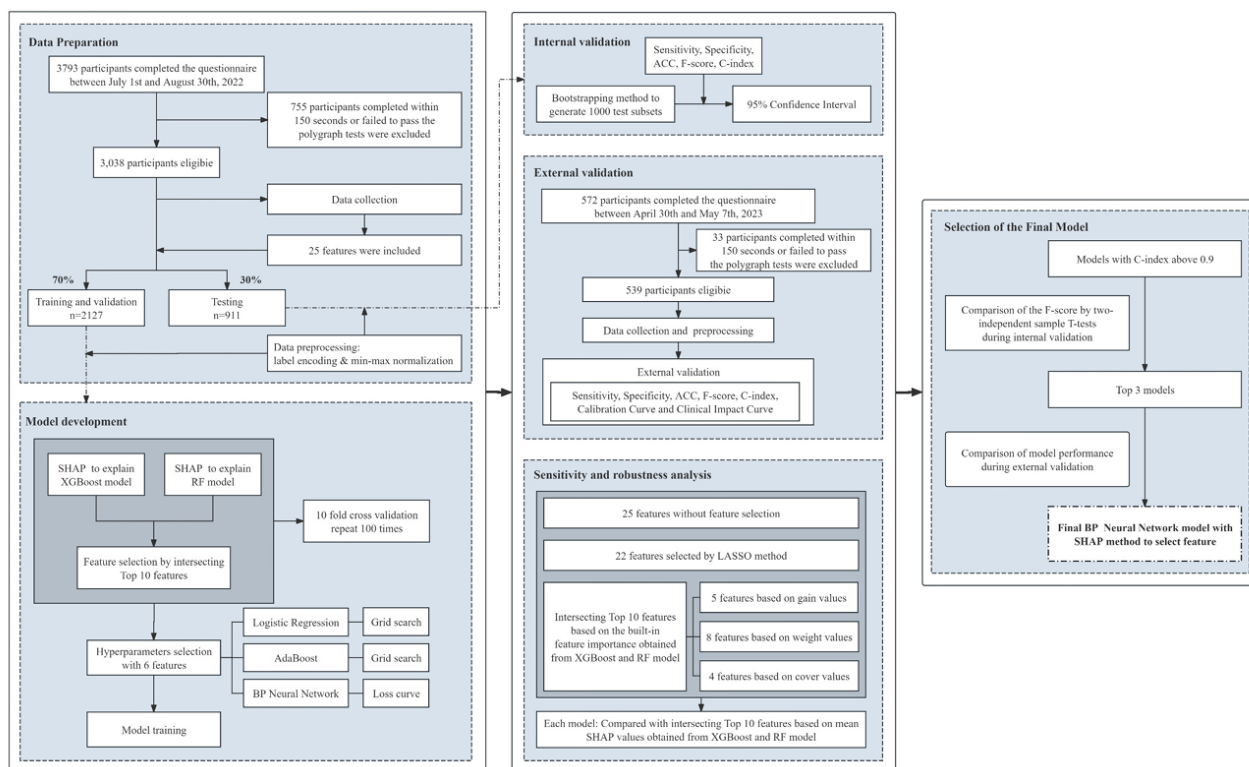
此研究提出用於預測青少年心理健康狀態的完整機器學習流程：

Data Preprocessing：此研究使用了異常資料剔除、缺失值處理、資料切分(Wave 1 70%用於訓練與驗證，再以 9:1 做 10 fold cross validation，30%為內部測試集。Wave 2 僅用於外部驗證)、資料轉換。

Feature Selection：使用 Shapley Additive Explanation 方法來衡量每項特徵對預測結果的重要性，採用 Random Forest 與 XGBoost 分別計算 SHAP 值，重複計算 1000 次取平均後，取兩模型前十名特徵的交集，作為模型最終輸入特徵(共 6 個)。

Model Training：共訓練三種模型 Logistic Regression、AdaBoost、Back Propagation Neural Network，每個模型都基於前述 6 個選定特徵進行訓練。

Validation & Evaluation：評估指標包含：Accuracy、Sensitivity/Specificity、F-score、C-index、Brier Score、Bias-corrected Calibration Curve、Clinical Impact Curve。



▲ The whole modelling processes

11.4 Results

在 Sensitivity 指標中 LR(86.66%)表現最好，BPNN 較低(83.29%)，結果顯示 BPNN 判斷可能較為保守容易漏判部分 BSI-positive，Specificity 指標中 BPNN 表現最為突出 (98.10%)，Accuracy 三者準確率皆高但 BPNN(93.81%)略高於其於兩者，F-score 指標中 BPNN(88.61%)略高於 LR(86.98%)及 AdaBoost(88.38%)，LR 和 AdaBoost 有較高的 C-index (>95%)，但 BPNN 稍低(91.20%)，但仍具有良好的預測能力。根據外部驗證結果選出最終模型 SHAP + BPNN 為最終模型。

| Model | Sensitivity (%) | Specificity (%) | Accuracy (%) | F-score (%) | C-index (%) |
|------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Internal validation (mean [95%CI]) | | | | | |
| LR ^a | 86.66 [86.54, 86.79] | 94.91 [94.85, 94.96] | 92.53 [92.47, 92.58] | 86.98 [86.88, 87.07] | 95.20 [95.14, 95.26] |
| AdaBoost ^b | 85.53 [85.40, 85.67] | 96.76 [96.72, 96.80] | 93.52 [93.47, 93.57] | 88.38 [88.28, 88.47] | 95.29 [95.23, 95.34] |
| BPNN ^c | 83.29 [83.14, 83.45] | 98.10 [98.04, 98.15] | 93.81 [93.77, 93.86] | 88.61 [88.53, 88.69] | 91.20 [91.11, 91.29] |
| External validation | | | | | |
| LR ^a | 91.40 | 94.17 | 93.69 | 83.33 | 98.02 |
| AdaBoost ^b | 88.17 | 95.52 | 94.25 | 84.10 | 96.11 |
| BPNN ^c | 90.32 | 95.07 | 94.25 | 84.42 | 97.49 |

▲ the performance of the models developed with different algorithms in primary analysis

11.5 Main Contributions

- 建構一套使用 SHAP + BPNN 的心理預測模型，模型經內外部驗證均具高準確性，泛化能力強。
- 提出以非單一模型特徵交集方式進行重要特徵選擇，將特徵數從 25 縮減至 6，提升模型計算效率與解釋性。

11.6 Discussions

- I. Discuss what data analytics experiences you obtained from the term project can be used or extended for this extended topic.
 - 可以根據我們提出的資料擴增的方法，對選出的 6 個 feature 進行處理提升模型表現
- II. Discuss what data analytics methods/experiences given in this paper can be helpful for your term project.
 - 情緒智力、社會支持等等結構化輸入資料來源，適合擴展我們的 feature
 - Bias-corrected Calibration Curve、Clinical Impact Curve 啟發了新的指標評估工具

12. Digital Phenotyping for Adolescent Mental Health: A Feasibility Study Employing Machine Learning to Predict Mental Health Risk From Active and Passive Smartphone Data

12.1 Target Problem

此論文在評估使用創新的機器學習模型架構，將主動與被動的智慧型手機數據（Digital Phenotyping）整合，來預測非臨床青少年的心理健康的可行性。

12.2 Dataset

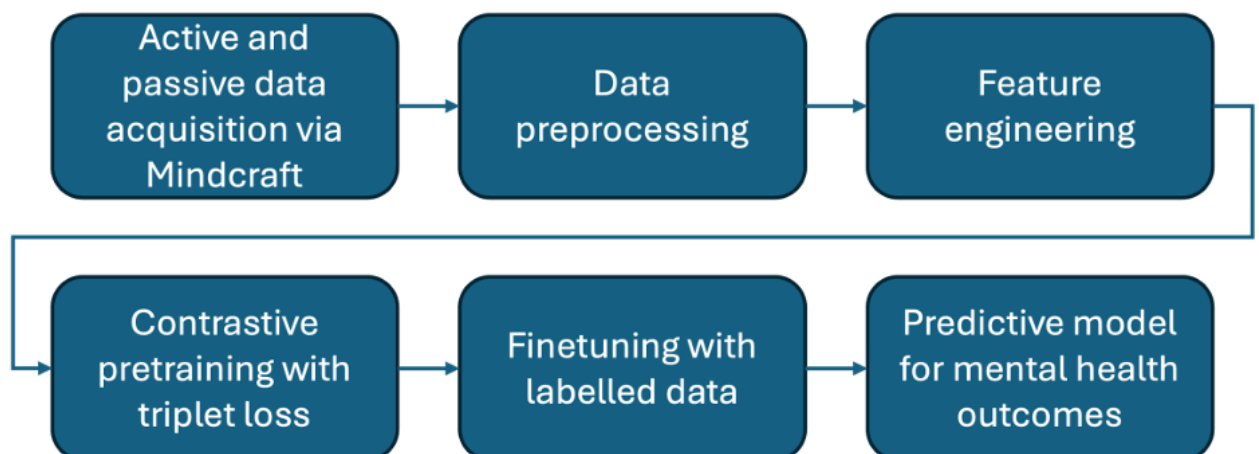
樣本來自於倫敦三所中學，共 103 位青少年，平均年齡為 16.1 歲，持續使用 Minecraft APP 作為回報工具 14 天，收集主動數據(情緒、睡眠、孤獨感、自我照顧等等)、被動數據(GPS 位置、步數、背景噪音、螢幕亮度、APP 使用時間等等，共計 92 種特徵)。

12.3 Overview of the framework

Feature Engineering：作者將設計的 92 個被動數據使用 Cumulative Median 平滑化以穩定模型輸入。

Contrastive Pretraining with triplet loss：此論文提到使用 Triplet Margin Loss 想讓模型學會辨別同一使用者不同天的資料及不同使用者的資料，接者使用兩層 MLP 產生 Embedding

Supervised Fine-Tuning：每個任務皆為二元分類問題，包含 SDQ 高風險、失眠、飲食障礙、自殺意念



▲ Workflow of the machine learning pipeline

12.4 Results

此論文使用的評估指標包含 Balanced Accuracy、AUC、F1、Precision、Recall 等，整體表現(Balanced Accuracy)SDQ 高風險(0.71)、失眠(0.67)、自殺意念(0.77)、飲食障礙(0.70)，藉由 SHAP 分析得到關鍵特徵分別是主動數據：負面思考、飄忽思緒、自我照顧、希望感、孤獨感，被動數據：Location Entropy、步數、夜間噪音、螢幕亮度，預訓練模型平均準確度高於未預訓練，提升約 0.03。

| | SDQ-High Risk | Insomnia | Suicidal ideation | Eating disorder |
|--------------------------|----------------------|-----------------|--------------------------|------------------------|
| Balanced accuracy | 0.71 ± 0.03 | 0.67 ± 0.04 | 0.77 ± 0.03 | 0.70 ± 0.03 |
| AUC | 0.77 ± 0.03 | 0.74 ± 0.02 | 0.82 ± 0.03 | 0.73 ± 0.02 |
| AUC-PR | 0.53 ± 0.04 | 0.52 ± 0.05 | 0.64 ± 0.05 | 0.52 ± 0.03 |
| F1 | 0.61 ± 0.04 | 0.59 ± 0.04 | 0.70 ± 0.04 | 0.61 ± 0.03 |
| F1 macro | 0.69 ± 0.03 | 0.66 ± 0.04 | 0.76 ± 0.03 | 0.68 ± 0.03 |
| Sensitivity | 0.71 ± 0.06 | 0.68 ± 0.03 | 0.78 ± 0.05 | 0.74 ± 0.03 |
| Specificity | 0.71 ± 0.05 | 0.66 ± 0.07 | 0.77 ± 0.04 | 0.67 ± 0.04 |
| Precision | 0.53 ± 0.04 | 0.52 ± 0.05 | 0.64 ± 0.05 | 0.52 ± 0.03 |
| Recall | 0.71 ± 0.06 | 0.68 ± 0.03 | 0.78 ± 0.05 | 0.74 ± 0.03 |

▲ Detailed Performance Metrics for Mental Health Outcome Predictions

12.5 Main Contributions

- 首創於非臨床青少年族群整合主動與被動資料進行心理風險預測
- 首創引入 Contrastive Learning，提升特徵一致性

12.6 Discussions

- I. Discuss what data analytics experiences you obtained from the term project can be used or extended for this extended topic.
 - 此研究使用 pretraining (contrastive learning) 搭配 fine-tuning MLP 進行風險分類，我們在 project 中也使用過不同模型融合策略能理解將表徵學習與任務導向模型解耦設計的優勢，可延伸至此文章的方法構建具有分支 attention 的跨模態結構（例如：一條處理主動資料、一條處理被動資料），並以 transformer 實現 cross-attention 整合，有機會進一步提升模型性能。
- II. Discuss what data analytics methods/experiences given in this paper can be helpful for your term project
 - 這篇文章證實多模態(主動+被動)資料融合大幅提升模型效能，
 - 對比學習可作為特徵處理時的 pretrain 策略

13. Early detection of mental health disorders using machine learning models using behavioral and voice data analysis

13.1 Target Problem

作者指出心理疾病具有全球性、潛伏性與多因性等挑戰，傳統臨床診斷方式耗時、主觀且易延誤治療。尤其在青春期中與成年初期，憂鬱、焦慮等心理困擾更容易因壓力、社交問題、行為模式或語音表徵而被忽略。這篇文章提出一套融合語音與行為數據的多模態預測架構 NeuroVibeNet，目標是實現對心理異常的早期偵測與分類。作者指出，透過整合語音與行為的非侵入性特徵，可以有效提升模型準確率，並強調可拓展性、臨床適用性與即時性。

13.2 Dataset

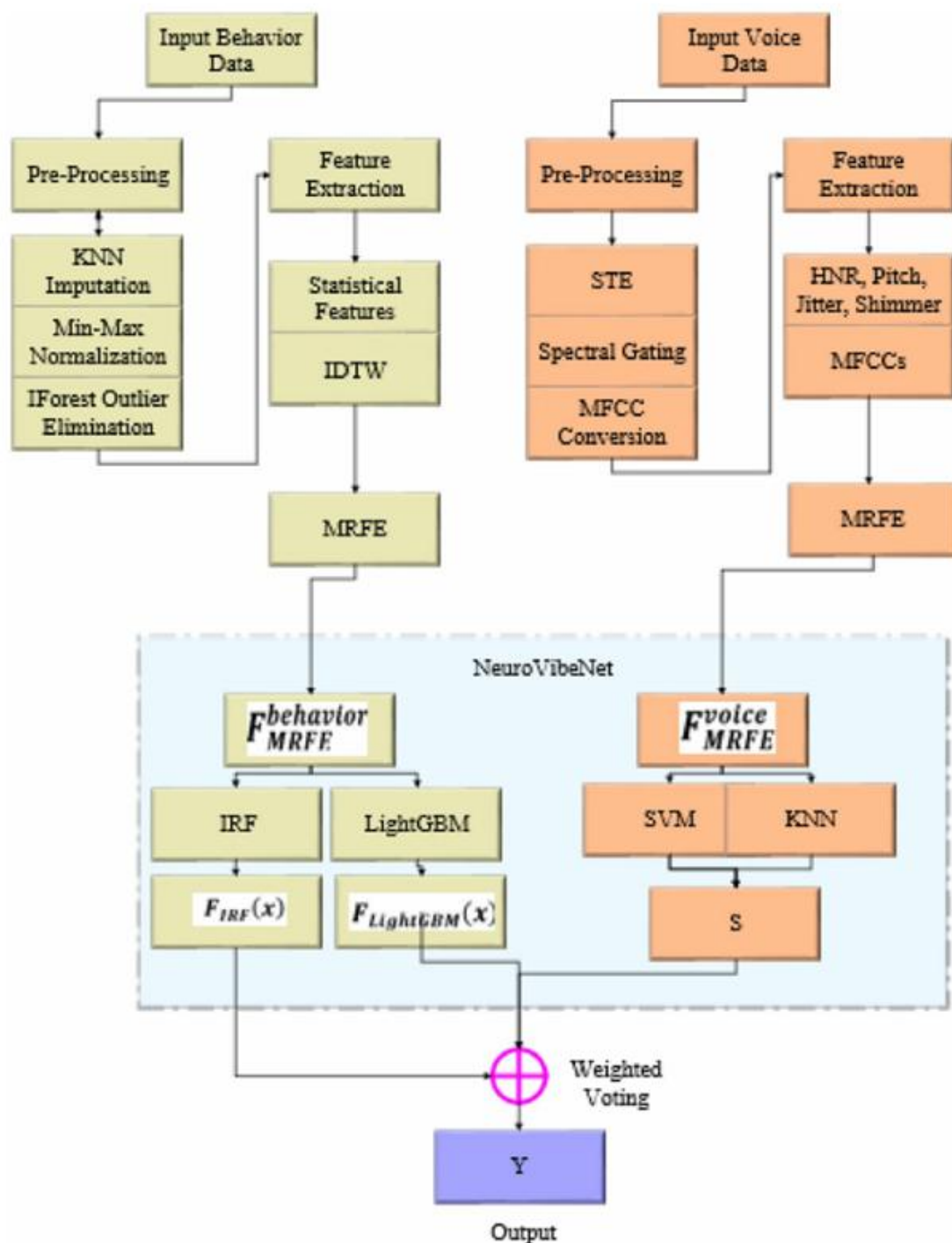
此研究整合兩筆公開資料集 **Mental Disorder Classification Dataset**（行為數據）資料內容包括參與者的生活行為、壓力指標、運動習慣、社交互動等，分類包含健康者與不同心理疾病患者（憂鬱、焦慮、雙相情緒障礙等）。**MODMA Dataset**（語音數據）提供受試者錄製的語音樣本，涵蓋病患與健康對照組。包含的語音特徵有 pitch、MFCC、jitter、shimmer、HNR 等。特徵資料經過配對排序與對齊，建立完整的多模態訓練樣本。

13.3 Overview of the framework

Data Preprocessing：對行為資料使用 KNN 補齊缺失值，採用 Min-Max normalization 將特徵值進行標準化，利用 Isolation Forest (iForest) 偵測並剔除離群樣本。對語音資料以 Short-Time Energy (STE) 演算法將語音切割為短段，使用 spectral gating 去除背景噪音。再來依 label 對兩模態資料排序並逐 row 對齊後合併。

Feature Extraction：對行為數據用 IDTW 分析後擷取的三類統計特徵（mean, variance, skewness）再從語音樣本中擷取以下五種常見語音表徵（pitch、MFCCs、jitter、shimmer、HNR）。

NeuroVibeNet 採用雙分支結構，各模態獨立訓練後再進行投票整合，分為 Behavioral 分支模型使用 Improved Random Forest (IRF)、LightGBM 及 Voice 分支模型使用 Hybrid SVM + KNN，接者使用 Weighted Voting 形成最終分類結果。使用 MRFE（Modified Recursive Feature Elimination）針對每個模態單獨進行特徵重要性排序與遞迴刪減，選出最關鍵的子特徵集，提升泛化能力與解釋性。使用 8:2 分割（交叉驗證）與 7:3 外部驗證，以 Accuracy、F1-Score、Sensitivity、Specificity、MCC（Matthews Correlation Coefficient）、PR-AUC 做為評估指標。



▲ Overview of model

13.4 Results

在 Accuracy 指標上，NeuroVibeNet 表現最好，達到了 99.06%，遠高於基礎模型如 Random Forest (95.32%) 與 LightGBM (96.85%)。在 F1-score 方面亦明顯領先，NeuroVibeNet 達到 99.00%，顯示其在正類與負類的辨識能力上皆十分穩定。Sensitivity 部分（模型偵測患病個體的能力）則由 NeuroVibeNet 以 98.90% 領先，其判斷偏向積極，有利於早期識別心理異常個體，但也相對提高了誤報的可能，在 Specificity（正確

排除健康個體的能力)上, NeuroVibeNet 同樣表現出色 (99.20%), 展現出高度的平衡性, 避免模型過度傾向某一類別。在整體二元分類相關性評估的 MCC 指標上, NeuroVibeNet 則達到 0.97, 顯示其輸出結果與實際標記高度一致。整體來看, 經過 MRFE 特徵優化與多模態融合後的 NeuroVibeNet 在所有指標上全面領先。

| Model | Accuracy | Precision | Specificity | Sensitivity | F1-Score | MCC | NPV | FPR | FNR |
|-----------------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|-------------|-------------|
| RF | 94.85 | 94.50 | 95.70 | 94.30 | 94.40 | 0.87 | 94.10 | 4.30 | 5.70 |
| SVM | 92.90 | 92.40 | 93.50 | 92.70 | 92.55 | 0.83 | 92.30 | 6.50 | 7.30 |
| LightGBM | 96.20 | 95.80 | 96.80 | 96.10 | 96.00 | 0.90 | 95.90 | 3.20 | 3.90 |
| KNN | 89.80 | 89.40 | 90.60 | 89.50 | 89.45 | 0.76 | 89.10 | 9.40 | 10.50 |
| LR | 88.00 | 87.80 | 88.90 | 87.60 | 87.70 | 0.72 | 87.50 | 11.10 | 12.40 |
| XGBoost | 96.80 | 96.50 | 97.10 | 96.40 | 96.45 | 0.91 | 96.20 | 2.90 | 3.60 |
| Proposed NeuroVibeNet | 98.50 | 98.40 | 98.80 | 98.20 | 98.30 | 0.96 | 98.00 | 1.20 | 1.80 |

▲ Performance of proposed NeuroVibeNet

13.5 Main Contributions

- 首次整合行為 + 語音多模態資料進行心理健康預測
- 提出 **NeuroVibeNet** 架構, 整合 IRF、LightGBM、Hybrid SVM+KNN 並用 weighted voting 增強模型穩定性
- 使用 **改良型 MRFE 特徵選擇法**, 同時維持可解釋性與泛化能力

13.6 Discussions

- I. Discuss what data analytics experiences you obtained from the term project can be used or extended for this extended topic.
 - 若能結合深度融合策略、引入序列學習架構如 Transformer, 或增加模型可解釋性 (例如 attention 機制), 或許有助於進一步提升模型實用性與臨床應用可信度。
- II. Discuss what data analytics methods/experiences given in this paper can be helpful for your term project
 - 結合語音模態 (如 MFCC, jitter, shimmer) 啟發導入語音問卷或視訊記錄作為新資料來源
 - MRFE 概念提升了特徵選擇動態性與上下文感知性, 適合用於目前的 feature pruning 流程

14. DepML: An Efficient Machine Learning-Based MDD Detection System in IoMT Framework

14.1 Target Problem

作者針對重度憂鬱症 (MDD) 自動化辨識進行探討, 指出傳統如 BDI (Beck Depression Inventory) 與 DRS (Depression Rating Scale) 主觀性高, 且實施成本高、易延誤治療時機, 因此提出一套低複雜度、高準確率、可部署於 IoMT 架構的 MDD 偵測系統 DepML。DepML 利用 EEG 生理訊號結合機器學習模型, 實作一套能自動萃取、選擇關鍵特徵並進行憂鬱判斷的架構, 適用於遠距或穿戴式智慧裝置應用場景。

14.2 Dataset

資料來源為 Mumtaz 等人公開之 EEG 資料集，共 64 位受試者（34 名 MDD、30 名健康對照），年齡介於 12–77 歲之間，平均 20.54 歲，男女比例約 3:2。每位受試者記錄 10 分鐘腦波（眼閉與眼開各 5 分鐘），取樣率為 256 Hz，採用國際 10–20 系統共 19 個 EEG 頻道。

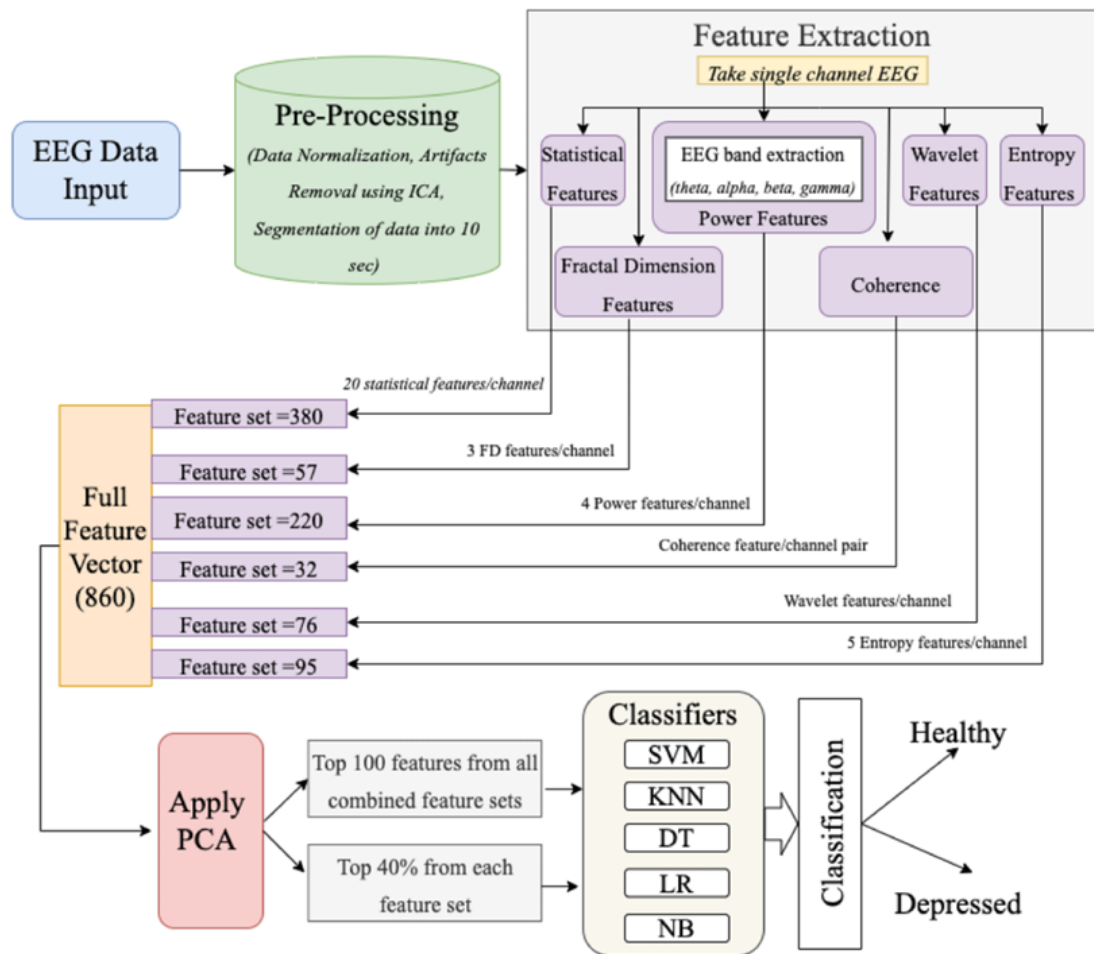
14.3 Overview of the framework

Preprocessing: 使用 ICA 去除眼動與肌肉干擾與 Z-score 標準化再以 10 秒為單位進行分段，每位受試者產生 60 筆樣本，最終建構出一個 860 維度的特徵矩陣。

Feature Extraction: 總共萃取 860 維特徵，包含：統計特徵（mean, std, skewness）頻譜功率特徵（Power spectral density）非線性特徵（Entropy, Fractal Dimension）時域特徵（Wavelet）神經連結性（Coherence）。

Feature Selection: 使用 PCA、RBA、NCA 三種方法比較最終選擇 **PCA** 擷取 top 40% 關鍵特徵，有效降低模型計算成本而不損失準確性。

共使用了五種分類器 RBF-SVM、KNN、Logistic Regression、Decision Tree、Naïve Bayes，最後選擇 RBF-SVM 為最終模型，使用 10-fold cross-validation，以 Accuracy、F1-score、Sensitivity、Specificity、MCC、分類時間作為評估指標。



▲ Architectural design of machine learning based depression detection system

14.4 Results

在非線性特徵與 RBF-SVM 模型的組合下，DepML 達成最高的分類準確率 98.90%。Sensitivity 達 99.2%，顯示模型具有極高的檢出能力，在不犧牲特異性下仍能有效抓出 MDD 患者。與其他模型（如 LR、NB、KNN）比較下，DepML 準確率高出約 2~4%，且在特徵維度減少 60% 的情況下分類效能並無明顯下滑。分類時間也從 491 秒降至 259 秒，證明系統能兼顧效能與效率。

| Classifiers | SVM | KNN | DT | LR | NB |
|--------------------|-------|-------|-------|-------|-------|
| Linear | | | | | |
| Without FS | 0.868 | 0.823 | 0.801 | 0.854 | 0.862 |
| With FS | 0.885 | 0.814 | 0.765 | 0.812 | 0.851 |
| Non-linear | | | | | |
| Without FS | 0.989 | 0.984 | 0.978 | 0.98 | 0.969 |
| With FS | 0.987 | 0.981 | 0.971 | 0.978 | 0.96 |
| FD | | | | | |
| Without FS | 0.782 | 0.728 | 0.713 | 0.766 | 0.752 |
| With FS | 0.78 | 0.73 | 0.706 | 0.756 | 0.742 |
| Statistical | | | | | |
| Without FS | 0.964 | 0.958 | 0.96 | 0.955 | 0.948 |
| With FS | 0.965 | 0.956 | 0.95 | 0.951 | 0.942 |
| Time | | | | | |
| Without FS | 0.801 | 0.793 | 0.741 | 0.768 | 0.772 |
| With FS | 0.793 | 0.79 | 0.73 | 0.752 | 0.75 |
| Coherence | | | | | |
| Without FS | 0.965 | 0.955 | 0.902 | 0.913 | 0.939 |
| With FS | 0.964 | 0.956 | 0.891 | 0.884 | 0.920 |

▲ Classification accuracy comparison without and with feature selection (FS) method

| Classifiers | SVM | KNN | DT | LR | NB |
|--------------------|--------|--------|--------|--------|--------|
| Linear | | | | | |
| Without FS | 515.34 | 620.91 | 231.12 | 430.22 | 489.33 |
| With FS | 243.88 | 377.57 | 102.01 | 256.34 | 279.90 |
| Non-linear | | | | | |
| Without FS | 491.23 | 575.14 | 210.40 | 398.09 | 451.13 |
| With FS | 259.90 | 323.38 | 122.91 | 176.97 | 210.00 |
| FD | | | | | |
| Without FS | 459.20 | 543.85 | 183.33 | 372.34 | 429.90 |
| With FS | 214.19 | 271.67 | 99.21 | 241.30 | 203.88 |
| Statistical | | | | | |
| Without FS | 505.10 | 589.07 | 229.04 | 418.43 | 470.76 |
| With FS | 290.54 | 259.87 | 115.54 | 202.23 | 253.45 |
| Time | | | | | |
| Without FS | 477.19 | 571.67 | 197.90 | 389.28 | 435.69 |
| With FS | 246.56 | 223.34 | 99.10 | 171.29 | 231.29 |
| Coherence | | | | | |
| Without FS | 433.79 | 481.69 | 159.36 | 393.88 | 380.80 |
| With FS | 209.46 | 239.63 | 80.54 | 179.09 | 149.73 |

▲ Classification time comparison without and with feature selection (FS) method (in seconds)

14.5 Main Contributions

- 提出一個針對重度憂鬱症的機器學習診斷架構 (DepML)，可實現於 IoMT 環境
- 使用 PCA 有效減少 60% 特徵數，運算速度提升約 2 倍

14.6 Discussions

- I. Discuss what data analytics experiences you obtained from the term project can be used or extended for this extended topic.
 - Transformer 模型可替代現有的獨立分類器架構，使 EEG 各通道間具備跨維 attention 能力，有助於模型捕捉跨頻道互動特徵
- II. Discuss what data analytics methods/experiences given in this paper can be helpful for your term project
 - 強調非線性腦波特徵的重要性 (entropy, fractal dimension) 可啟發未來嘗試更多非傳統特徵

15. MHDeep: Mental Health Disorder Detection System Based on Body-Area and Deep Neural Networks

15.1 Target Problem

傳統精神疾病診斷往往依賴自我回報或問卷量表，主觀性與延遲風險高，因此作者提出一個名為 MHDeep 的系統，利用可穿戴設備（如智慧手錶和智慧型手機）收集的多模態感測數據，結合 DNN 模型，以檢測三種主要的心理健康障礙：思覺失調情感障礙 (Schizoaffective Disorder)、重度憂鬱症 (Major Depressive Disorder) 和雙相情感障礙 (Bipolar Disorder)，實現即時監測、邊緣運算與早期介入。

15.2 Dataset

實驗資料由 74 位受試者收集，包含 25 位健康者、23 位雙相情緒障礙患者、10 位憂鬱症患者、16 位分裂情感障礙患者，每位參與者皆配戴 Empatica E4 手錶與 Samsung Galaxy S4 手機，收集共 8 種生理與環境感測數據 (GSR、ST、IBI、Acc、Temp、Grav、Vel 等)，每位收集約 1.5 小時，製成超過 120,000 筆 15 秒長的樣本。

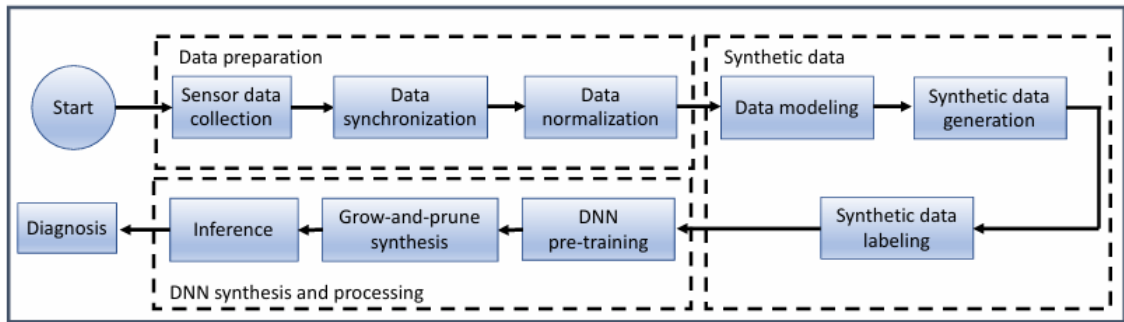
15.3 Overview of the framework

Preprocessing: 對所有數據做同步對齊，然後做標準化 (normalization) 和 sliding window 切割，每 15 秒切出一筆樣本，每筆資料就變成一段時間內的生理行為狀況，總共會形成一個 2325 維的特徵向量。

Synthetic Data Generation: 因為原始資料數量偏少，而且分布也不均，作者特別設計了一個 clever 的資料擴增步驟。他們先使用高斯混合模型 (GMM) 根據原始資料的分布隨機生成 10 萬筆新資料（但這些資料一開始是沒標籤的），再用先訓練好的傳統模型（像是 Random Forest）來幫這些資料標上最有可能的分類標籤。

Model Training: 訓練由 4~5 層全連接層組成的 DNN 模型，輸入是 2325 維的感測特徵向量，再利用 **Grow-and-Prune** 的方法去修剪不重要的權重。

使用 Parameters、FLOPs、Accuracy、False Positive Rate (FPR)、False Negative Rate (FNR)、F1 Score 做為評估指標。



▲ Overview of the MHDeep framework

15.4 Results

在三類疾病分類任務中（Healthy vs. Schizoaffective/MDD/Bipolar），MHDeep 表現為 Schizoaffective disorder：平均 Accuracy 為 90.4%、F1-score 達 85.7%，Major depressive disorder：平均 Accuracy 為 87.3%、F1-score 約 74.3%，Bipolar disorder：平均 Accuracy 為 82.4%、F1-score 約 81.2%，透過 Grow-and-Prune 訓練流程，大幅壓縮模型參數與 FLOPs（最多達 2.8×）。

15.5 Main Contributions

- 使用 GMM 合成資料以進行 pretraining，有效提升小樣本下的模型穩定性
- 建構首套使用商用穿戴裝置進行憂鬱、雙相與分裂情感障礙分類的深度學習系統

15.6 Discussions

- I. Discuss what data analytics experiences you obtained from the term project can be used or extended for this extended topic.
 - Transformer 能加強跨模態訊號（如 GSR × Motion）間的特徵互動學習，有望進一步提升 F1-score 與 Recall。
- II. Discuss what data analytics methods/experiences given in this paper can be helpful for your term project
 - MHDeep 對合成資料的使用（GMM-based sampling + 傳統模型標記）提供了極具啟發性的 low-data training 策略。