心臟衰竭病人出院後6個月死亡預測

工作分配:

張哲安:

1.文獻查證

2.撰寫目的、動機及資料集來源

3.串取MIMIC-III資料及資料前處理

4.撰寫資料及處理-納入、排除條件及特殊條件、特徵選擇

B組員:

1.模型訓練

2.撰寫實驗流程、實驗結果與討論

3.製作圖表

C組員:製作PPT

D組員:上台報告呈現

一、目的、動機及資料集來源:

Savarese et al.(2023)中提到直至2017年統計，心衰竭影響全球超過6430萬人。Chen et al.(2024)表示隨著醫療的進步，心衰竭的治療也取得重大進展，但仍然預後不佳。Abdin et al.(2021)提到心衰竭病人於出院後30至90天被稱為脆弱期。這段期間，因各種原因導致的死亡率，從30天內比原本高出10%，並且持續攀升至90天的31%。Y et al.(2022)表示這種脆弱期導致的死亡率攀升，延續至180天才有逐漸下降的趨勢。基於上述，心衰竭病人出院後180天內有較高的死亡風險，本組預建立機器學習模型，以預測心衰竭病人出院後180天內的死亡機率。期望能及早識別高風險族群，進而採取介入措施，降低病人死亡風險。

心衰竭病人於出院後出現的脆弱期，主要是因器官受到損傷、身體虛弱，尚未完全從疾病中恢復，導致死亡率增加。由於加護病房的病人具有較嚴重的病況，因此本研究選擇使用MIMIC-III資料庫作為研究資料來源。

二、資料及處理-納入、排除條件及特殊條件:

1.納入入院主診斷ICD-9所有包含心衰竭的1639位病人

2.排除院內死亡者199位

最終共1440筆資料，333位死亡，1107位存活。

3.身高3筆、體重4筆為異常值改為空值

4.MIMIC-III因隱私原因，當病人年齡超過90歲時，不給予真實出生日期。於資料庫手冊中提己超過90歲之病人年齡中位數為91.4歲，故研究中169位年齡皆以此代替。

5.空值依照是否為常態使用平均數或中位數填補

三、特徵選擇:

前面提到心臟衰竭病患因器官損傷、身體虛弱，尚未完全從疾病中恢復，而出現的脆弱期，導致死亡率增加，主要因素為:

1.病況嚴重不穩定:電解質不平衡、腎功能惡化、感染，進而影響心臟功能。Abdin et al.(2021), Y et al.(2022), Chen et al.(2024)

2.多重共病:許多心臟衰竭病患同時患有其他疾病，這些疾病會加重心臟衰竭的嚴重程度，並增加脆弱期的風險。Y et al.(2022)

3.行為因素:未遵循醫囑，如:控制水份，或酗酒等不良習慣。Y et al.(2022)

4.低社經地位:獲得醫療保健機會低、健康素養較低、生活方式。Phan et al.(2023)

5.心理因素:焦慮、憂鬱症會導致生活品質下降，治療遵從性下降。Y et al.(2022)

因此本研究以此建立特徵，包含基本資訊、生命徵象、抽血數值、共病及過去病史，共72個特徵。

標籤:

|  |  |  |
| --- | --- | --- |
| 標籤名稱 | 說明 | 類型 |
| 180DAYS\_DEAD | 出院後180天內死亡 | 類別0:活、1:死 |

A chart with numbers and a blue circle

Description automatically generated

特徵:

抽血及體重皆有同一次住院中第一筆及最後一筆，了解病人最一開始疾病嚴重程度，及經過治療最後恢復狀況。

|  |  |  |
| --- | --- | --- |
| 特徵名稱 | 說明 | 類型 |
| AGE | 年齡 | 數值 |
| GENDER | 性別 | 類別0:女、1:男 |
| INSURANCE | 保險 | 類別 |
| HOS\_DAYS | 住院天數 | 數值 |
| HEIGHT | 身高 | 數值 |
| WEIGHT | 體重 | 數值 |
| BMI | 身體質量指數 | 數值 |
| WEIGHT\_CHANGE | 體重變化(最後一筆減第一筆) | 數值 |
| DIABETES\_WITHOUT\_CH | 無慢性併發糖尿病 | 類別0:無、1:有 |
| DIABETES\_WITH\_CH | 有慢性併發糖尿病 | 類別0:無、1:有 |
| PER\_VASCULAR\_DIS | 周邊血管疾病 | 類別0:無、1:有 |
| CERE\_VASCULAR\_DIS | 腦血管疾病 | 類別0:無、1:有 |
| DEMENTIA | 失智症 | 類別0:無、1:有 |
| COPD | 慢性肺病 | 類別0:無、1:有 |
| PEPTIC | 消化性潰瘍 | 類別0:無、1:有 |
| RHEUMATIC | 風濕病 | 類別0:無、1:有 |
| MILD\_LIVER | 輕度肝病 | 類別0:無、1:有 |
| MOD\_LIVER | 中度肝病 | 類別0:無、1:有 |
| PARAPLEGIA | 半側麻痺或下肢癱瘓 | 類別0:無、1:有 |
| RENAL | 腎病 | 類別0:無、1:有 |
| TUMORT | 腫瘤 | 類別0:無、1:有 |
| LYMPH | 淋巴瘤 | 類別0:無、1:有 |
| LEUKE | 白血病 | 類別0:無、1:有 |
| META\_TUMOR | 轉移性腫瘤 | 類別0:無、1:有 |
| MI | 心肌梗塞 | 類別0:無、1:有 |
| CCI\_SCORE | 查爾森合拼症嚴重度指標 | 數值 |
| AFIB | 心房顫動 | 類別0:無、1:有 |
| DEPRESSION | 憂鬱症 | 類別0:無、1:有 |
| HYPERTENSION | 高血壓 | 類別0:無、1:有 |
| HYPERLIPIDEMIA | 高血酯 | 類別0:無、1:有 |
| HB | 血紅素 | 數值 |
| WBC | 白血球 | 數值 |
| PLT | 血小板 | 數值 |
| NEU | 嗜中性 | 數值 |
| ESR | 紅血球沉降速率 | 數值 |
| MONO | 單核球 | 數值 |
| CRP | C-反應蛋白 | 數值 |
| NA | 鈉 | 數值 |
| MAG | 鎂 | 數值 |
| K | 鉀 | 數值 |
| BNP | NT-proBNP | 數值 |
| BUN | 尿素氮 | 數值 |
| CREA | 肌酸酐 | 數值 |
| ALB | 白蛋白 | 數值 |
| LAC | 乳酸 | 數值 |
| O | 血氧 | 數值 |
| HR | 心跳 | 數值 |
| RR | 呼吸 | 數值 |
| SBP | 收縮壓 | 數值 |
| SMOKE | 抽菸史 | 類別0:無、1:有 |

在共病特徵中排除AIDS，因抓出資料中，無任何人有此疾病。

四、實驗結果:

為了有效預測心衰竭病患的死亡風險，設計了一套完整的機器學習實驗流程。首先清理臨床數據，對於數值型數據，使用平均數或中位數填補缺失值並標準化後將之與類別型資料合併為一個完整的特徵矩陣，作為模型完整特徵輸入。

在模型選擇上，我們對三種不同的機器學習模型進行比較：Naive Bayes、Logistic Regression和XGBoost。為了確保模型學習能力，實驗中採用Stratified 10-Fold Cross-Validation進行評估，確保每個fold中的正負樣本比例與整體數據一致。過程中，各模型會將數據分為10等份，輪流選取其中一個作為測試集，其餘9個作為訓練集，進行模型訓練和測試。這一過程重複10次後，最終計算各fold的平均性能指標。

考量到數據中存在的正負樣本比例失衡問題，實驗中採用了三種不同的策略進行處理：

1. 不進行任何調整，直接以原始數據建模
2. 使用Under Sampling方法減少多數類樣本
3. 採用SMOTE技術生成合成少數類樣本，以平衡數據分佈。

另外，實驗中也討論了利用XGBoost模型進行特徵篩選，基於重要性得分選取排名前80%的特徵，並構建兩種特徵組合：使用所有特徵的完整特徵組合，以及基於篩選結果的精簡特徵組合。

通過上述模型選擇、採樣方式以及是否採納feature selection等三個層面，進一步設計了多種模型組合實驗。例如，比較了未採樣與SMOTE或Under Sampling處理數據的效果，並評估全特徵與篩選後特徵對模型性能的影響。這些組合的實驗結果通過準確率、精確率、召回率、F1分數和ROC-AUC等指標進行量化。最後，通過結果分析並將AUROC與AUPRC 曲線視覺化比較了各模型在不同實驗條件下的性能差異，並討論了特徵篩選與數據平衡技術的影響。

五、實驗結果與討論

本次實驗比較了三種機器學習模型（XGBoost、Logistic Regression、Naive Bayes）在處理心衰竭患者180天內死亡預測任務中的表現，並探討了數據平衡技術（SMOTE和Under Sampling）以及特徵篩選對模型性能的影響。實驗結果如表一和表二所示，並輔以ROC曲線和Precision-Recall曲線進行詳細分析。

首先，從表一和表二的結果可以看出，XGBoost模型在多數組合中的表現最優。未採用數據平衡的XGBoost（AllFeature）模型，在存活患者（Label\_0）上的F1分數達到0.87，死亡患者（Label\_1）的F1分數為0.34，顯示其在主類別（存活患者）上的預測能力極為突出。然而，對少數類別（死亡患者）的召回率僅為0.24，表明模型在處理不平衡數據時仍存在挑戰。當結合SMOTE技術進行數據平衡後，XGBoost的少數類別召回率提高至0.43，且F1分數提升至0.41，說明SMOTE能有效改善模型對死亡患者的識別能力，但也伴隨主類別性能的輕微下降。

從綜合性能指標（表二）來看，XGBoost在未進行數據平衡（AllFeature）時的準確率（Accuracy）為0.81，AUROC達到0.74，是所有組合中的最佳表現。此外，XGBoost結合特徵篩選和SMOTE後，AUROC從0.737微弱下降至0.729，但AUPRC略有提升，達到0.483，顯示特徵篩選在一定程度上提升了模型的精確性和召回率平衡能力。相比之下，Logistic Regression在大多數情況下的性能居中，當結合SMOTE時，其AUROC約為0.72，但對少數類別的AUPRC表現稍弱。Naive Bayes模型整體表現最差，AUROC僅達0.59左右，且對少數類別的召回率和F1分數均低於其他模型，這可能與其假設特徵條件獨立的模型局限性相關。

在ROC曲線分析中，XGBoost未平衡數據的曲線（AUC=0.737）在整體上優於其他組合，顯示其對不同閾值下的預測穩定性更高。當應用SMOTE進行數據平衡後，曲線略有下降，但仍保持較高的穩定性。另一方面，Precision-Recall曲線進一步證實了XGBoost在不平衡數據中的穩健表現，其AUPRC最高達到0.485，顯著高於Logistic Regression（0.43）和Naive Bayes（0.36）。這表明在處理少數類別樣本時，XGBoost更能保持較高的精確率和召回率平衡。

**實驗限制**

儘管本次實驗中透過不同組合的模型詳細討論了特徵以及採樣方式等對於模型學習及預測的影響，但仍存在一些限制。首先，從圖表中可以發現數據不平衡問題仍未完全解決。雖然SMOTE技術在一定程度上改善了少數類別的召回率，但生成的合成數據可能會影響模型對實際數據的泛化能力。而Under Sampling技術減少了多數類別樣本量，導致整體性能下降，尤其對於Naive Bayes模型影響尤為明顯。其次，特徵篩選僅基於XGBoost的重要性排序，未考慮其他特徵選擇方法（如Lasso回歸或PCA）可能帶來的潛在提升。這種單一特徵篩選方式可能限制了模型在不同特徵組合下的表現。此外，本實驗僅考察了三種機器學習模型，未涉及深度學習方法（如LSTM），可能未能充分挖掘數據中的複雜模式，且由於原始資料量龐大，本次實驗中未能納入如檢傷主訴等文字型資料，大幅限制了資料類型的多元性，因此本次實驗中也尚未探討基於Transformer架構的BERT、RoBERTa、BioClinicalBERT等預訓練語言模型。

**未來方向**

基於上述結果與限制，未來可以從以下幾個方面改進。第一是探索更先進的數據平衡技術，例如基於Cost-Sensitive Learning的方法，直接優化模型對少數類別的學習能力，而無需依賴數據合成或抽樣。其次，可以結合多模型集成方法（如Stacking或Blending），進一步提升模型的泛化性能。此外，特徵篩選方法的多樣化也是未來的重點，例如引入統計檢驗（卡方檢驗、ANOVA）或非監督學習的特徵提取技術（如Autoencoder），可能有助於發掘更具區分力的特徵。最重要的是，引入深度學習模型（如基於LSTM的序列模型或Transformer架構），以捕捉數據中的時序關係與高維非線性模式，進一步提升模型對死亡患者的預測能力。

最後，未來還可考慮擴大數據樣本量，特別是增加死亡患者的數據樣本，或者引入外部數據源以進行數據擴充，改善模型的訓練效果和泛化能力。這些措施將有助於進一步完善模型，提升其在心衰竭疾病預測任務中的應用價值。

表一：各模型預測心衰竭患者180天內死亡標籤指標

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Label\_0 (Survive) |  |  |  | Label\_1 (Death) |  |
| Models | Precision | Recall | F1 Score |  | Precision | Recall | F1 Score |
| XGBoost\_AllFeature | 0.81 | **0.95** | **0.87** |  | **0.58** | 0.24 | 0.34 |
| XGBoost\_\_AllFeature\_UnderSampling | **0.87** | 0.66 | 0.75 |  | 0.37 | **0.68** | **0.48** |
| XGBoost\_\_ FeatureSelection \_UnderSampling | 0.86 | 0.63 | 0.73 |  | 0.35 | 0.66 | 0.46 |
| XGBoost\_\_AllFeature\_SMOTE | 0.82 | 0.87 | 0.84 |  | 0.46 | 0.36 | 0.40 |
| XGBoost\_\_ FeatureSelection\_SMOTE | 0.82 | 0.91 | 0.86 |  | 0.54 | 0.32 | 0.40 |
| LogisticRegression\_\_UnderSampling | 0.86 | 0.67 | 0.75 |  | 0.37 | 0.64 | 0.47 |
| LogisticRegression\_UnderSampling\_FeatureSelection | 0.86 | 0.67 | 0.75 |  | 0.37 | 0.64 | 0.47 |
| LogisticRegression\_\_SMOTE | 0.82 | 0.80 | 0.81 |  | 0.39 | 0.43 | 0.41 |
| LogisticRegression\_SMOTE\_FeatureSelection | 0.82 | 0.80 | 0.81 |  | 0.39 | 0.43 | 0.41 |
| NaiveBayes\_\_UnderSampling | 0.81 | 0.81 | 0.81 |  | 0.37 | 0.36 | 0.35 |
| NaiveBayes\_UnderSampling\_FeatureSelection | 0.82 | 0.80 | 0.80 |  | 0.38 | 0.38 | 0.36 |
| NaiveBayes\_\_SMOTE | 0.81 | 0.51 | 0.63 |  | 0.27 | 0.60 | 0.38 |
| NaiveBayes\_SMOTE\_FeatureSelection | 0.81 | 0.55 | 0.66 |  | 0.28 | 0.58 | 0.38 |

表二：各模型預測心衰竭患者180天內死亡之綜合指標

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy |  |  |  |  |  |
|  |  | Macro |  | AUROC | AUPRC |
| Models | Precision | Recall | F1 Score |  |  |
| XGBoost\_AllFeature | **0.81** | **0.95** | **0.87** | 0.58 | 0.24 | 0.34 |
| XGBoost\_\_AllFeature\_UnderSampling | 0.66 | 0.62 | 0.67 | 0.61 | 0.72 | 0.44 |
| XGBoost\_\_ FeatureSelection \_UnderSampling | 0.64 | 0.61 | 0.65 | 0.59 | 0.72 | 0.44 |
| XGBoost\_\_AllFeature\_SMOTE | 0.75 | 0.64 | 0.62 | 0.62 | 0.71 | 0.45 |
| XGBoost\_\_ FeatureSelection\_SMOTE | 0.78 | 0.68 | 0.62 | **0.63** | **0.73** | **0.48** |
| LogisticRegression\_\_UnderSampling | 0.66 | 0.62 | 0.66 | 0.61 | 0.72 | 0.43 |
| LogisticRegression\_UnderSampling\_FeatureSelection | 0.66 | 0.62 | 0.66 | 0.61 | 0.72 | 0.43 |
| LogisticRegression\_\_SMOTE | 0.71 | 0.61 | 0.61 | 0.61 | 0.69 | 0.41 |
| LogisticRegression\_SMOTE\_FeatureSelection | 0.71 | 0.61 | 0.61 | 0.61 | 0.69 | 0.41 |
| NaiveBayes\_\_UnderSampling | 0.71 | 0.59 | 0.59 | 0.58 | 0.68 | 0.37 |
| NaiveBayes\_UnderSampling\_FeatureSelection | 0.70 | 0.60 | 0.59 | 0.58 | 0.67 | 0.36 |
| NaiveBayes\_\_SMOTE | 0.54 | 0.54 | 0.56 | 0.50 | 0.59 | 0.33 |
| NaiveBayes\_SMOTE\_FeatureSelection | 0.56 | 0.55 | 0.57 | 0.52 | 0.59 | 0.33 |
| XGBoost\_\_UnderSampling | 0.66 | 0.62 | 0.67 | 0.61 | 0.72 | 0.44 |

圖一：各模型AUROC、AUPRC曲線

A graph of different colored lines

Description automatically generated

六、參考文獻:

Abdin, A., Anker, S. D., Butler, J., Coats, A. J. S., Kindermann, I., Lainscak, M., Lund, L. H., Metra, M., Mullens, W., Rosano, G., Slawik, J., Wintrich, J., & Böhm, M. (2021). 'Time is prognosis' in heart failure: time-to-treatment initiation as a modifiable risk factor. *ESC heart failure*, *8*(6), 4444–4453.

Çavuşoğlu, Y., Altay, H., Nalbantgil, S., Temizhan, A., & Yılmaz, M. B. (2022). Pre-Discharge and Post-Discharge Management and Treatment Optimization in Acute Heart Failure. *Turk Kardiyoloji Dernegi arsivi : Turk Kardiyoloji Derneginin yayin organidir*, *50*(5), 378–394.

Chen, J., Yang, L., Han, J., Wang, L., Wu, T., & Zhao, D. (2024). Interpretable Machine Learning Models Using Peripheral Immune Cells to Predict 90-Day Readmission or Mortality in Acute Heart Failure Patients. *Clinical and applied thrombosis/hemostasis : official journal of the International Academy of Clinical and Applied Thrombosis/Hemostasis*, *30*, 10760296241259784.

Gracia, E., Singh, P., Collins, S., Chioncel, O., Pang, P., & Butler, J. (2018). The Vulnerable Phase of Heart Failure. *American journal of therapeutics*, *25*(4), e456–e464.

Phan, J., Barroca, C., & Fernandez, J. (2023). A Suggested Model for the Vulnerable Phase of Heart Failure: Assessment of Risk Factors, Multidisciplinary Monitoring, Cardiac Rehabilitation, and Addressing the Social Determinants of Health. *Cureus*, *15*(2), e35602.

Radovanovic, D., Seifert, B., Urban, P., Eberli, F. R., Rickli, H., Bertel, O., Puhan, M. A., Erne, P., & AMIS Plus Investigators (2014). Validity of Charlson Comorbidity Index in patients hospitalised with acute coronary syndrome. Insights from the nationwide AMIS Plus registry 2002-2012. *Heart (British Cardiac Society)*, *100*(4), 288–294.

Savarese, G., Becher, P. M., Lund, L. H., Seferovic, P., Rosano, G. M. C., & Coats, A. J. S. (2023). Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovascular research*, *118*(17), 3272–3287.

Sabouri, M., Rajabi, A. B., Hajianfar, G., Gharibi, O., Mohebi, M., Avval, A. H., Naderi, N., & Shiri, I. (2023). Machine learning based readmission and mortality prediction in heart failure patients. *Scientific reports*, *13*(1), 18671.