

MIMIC-IV 資料庫介紹

郭芷彤 副統計分析師

上期 eNEWS 中，介紹了 Python 的安裝方式，之後幾期會示範如何使用 Python 執行資料清理、統計分析、機器學習及繪製視覺化圖表等，本期將針對後續會使用的資料庫-MIMIC 來進行介紹。

一、簡介

MIMIC 全名為 Medical Information Mart for Intensive Care，目前最新版本為第四版(MIMIC-IV)，採取線上申請免費使用，資料來源為美國麻薩諸塞州波士頓的貝斯以色列女執事醫療中心(Beth Israel Deaconess Medical Center)。該資料庫提供了這間醫院在 2008-2019 年超過 25 萬名加護病房(ICU)住院患者相關的去識別化訊息，屬於單一機構的縱向資料，囊括了如患者人口學特徵、實驗室檢查、用藥情況、生命徵象、手術、疾病診斷等，特別的是除了結構化資料外，也有非結構化資料。

二、資料內容

1. 整個資料庫分為六大部分：Core、Hosp、ICU、ED、CXR、Note，而 Core、Hosp、ICU 則為資料庫最主要的部份，而目前 Note 的部分尚未公開，以下依序介紹各部分所包含的資料：
 - i. Core：包含了三個部分，分別為 Admissions table、Patients table 以及 Transfers table，主要為患者人口學相關資料(如：性別、年齡)以及住院相關資料(如住院和出院日期、轉院)。其中較特別是，為求保護個人隱私，所有的日期皆固定增加一個隨機的天數，不影響日期彼此間天數計算，但無法辨識原始的日期，年齡部分則是已算出年齡，未再提供出生年月日，而超過 89 歲以上患者則是統一標註為 91 歲。
 - ii. Hosp：總共分為 17 個部分，主要是醫院病歷或檢查相關的電子

數據，如 ICD 代碼、手術紀錄、藥品紀錄(藥物劑量、處方劑量數、給藥頻率、用藥途徑和處方持續時間)、其他處置紀錄(如放射科、營養科)、實驗室數據、細菌培養數據及抗生素判別、收費相關等。

- iii. ICU：主要針對的是 ICU 中如靜脈注射的藥品、劑量、藥物、藥物注射速度、呼吸器參數，其中特別的是詳細記錄了殘餘藥量、病人體重、若是治療被取消，也會在此檔案中附註。
- iv. ED：主要是來自急診的詳細資料，包含了入院時的狀況(包含疾病診斷、進出時間)、分診(包含了病患到急診後，由護理師進行的評估，像是目前身體健康狀況、測量生命徵象如血氧濃度、呼吸平瀘、自我評估疼痛程度等)、生命徵象(包含了呼吸、心跳、血壓等數值外，也包含了緊急嚴重性指數(ESI)及疼痛狀況)、以及藥物重整(包含了入院前所使用的藥品、入院後的藥物使用、給藥時間)。
- v. CXR：目前的版本為 MIMIC-CXR v2.0.0，為此間醫療機構的 2011-2016 年的 64,588 名患者之影像檢查，共有 377,110 張圖像。除了提供影像外，另外也提供由放射科醫師以及護理師所撰寫的非結構化的報告，裡面的病患資料已經以底線代替，內容大致為(1)簡單的歸因總結潛在的醫療狀況、(2)此次檢查的原因，以及(3)先前進行的影像學研究。
- vi. Note：雖然目前尚未公開資料，其包含了心電圖報告及其說明、超音波報告及其說明、放射報告及其說明、出院報告，而出院報告中則需描述入院原因、住院過程以及出院的總體相關說明。

三、已發表文獻

以下列舉幾篇各國團隊利用 MIMIC 資料庫所發表的文獻：

1. 中國南方醫科大學團隊於 2021 年在 *Ann Intensive Care (Impact factor :6.925)* 上發表的 *Early combination of albumin with crystalloids administration might be beneficial for the survival of septic patients: a retrospective analysis from MIMIC-IV database*，主要是針對敗血症患者進行白蛋白輸液的生存率研究，利用了 MIMIC-IV 中的約 25 萬位患者的年齡、性別、種族、保險條件、第一次 24 小時連續器官衰竭評估 (SOFA) 評分、以往病史、藥物使用種類及劑量、血液、尿液檢驗結果等來進行生存風險的估算。
2. 由日本順天堂大學團隊所研究發表的 *Machine Learning Approach to Predict Positive Screening of Methicillin-Resistant Staphylococcus aureus During Mechanical Ventilation Using Synthetic Dataset From MIMIC-IV Database* 刊登於 *Frontiers in Medicine (Impact factor:4.468)*，主旨是利用機器學習的方式建立模型，使得耐甲氧西林金黃色葡萄球菌(MRSA)的抗生素治療可以更有效，研究將個案隨機分配為 7：3 作為建模，從 MIMIC-IV 中選取了 28 個變項(如：過去病史、Charlson 合併症指數、呼吸機使用天數、先前使用皮質類固醇或抗生素的種類、介入手術等)來進行模型的預測，其結果 AUC0.89，sensitivity 達 0.98。
3. *Feature Explanations in Recurrent Neural Networks for Predicting Risk of Mortality in Intensive Care Patients* 是由泰國宋卡王子大學團隊於 2021 年在 *Journal of Personalized Medicine (Impact factor:4.453)* 上發表的研究，同時利用了 MIMIC-III、MIMIC-IV 以及 eICU 三個資料庫來進行，主要是利用資料庫中的來進行循環神經網路 (Recurrent Neural Network, RNN) 藉由深度學習來建立預測加護病房患者死亡風險的模型，利用了如患者基本資料、生命徵象、實驗室數據等。
4. *Plasma Anion Gap and Risk of In-Hospital Mortality in Patients with Acute Ischemic Stroke: Analysis from the MIMIC-IV Database* 是 2021 年發表在

Journal of Personalized Medicine 的文章(Impact factor:4.453)，由台灣 Hong-Jie Jhou 等人所發表，針對急性缺血性中風患者 plasma anion gap(AG)進行在院死亡相關研究，有別於健保資料庫中只能控制共病及用藥外，還包含了如 ICU 入院 24 小時內生命徵象及實驗數據、全身性發炎候群分數(Systemic inflammatory response syndrome, SIRS)及續器官衰竭評估分數(The sequential organ failure assessment, SOFA)等數據來進行多變項 Cox Hazard Model 分析。

5. 2022 年刊登在 Frontiers in Neurology (Impact factor:3.552)上，由台灣 Chen-Shu Wu 等人所發表的 *Atrial Fibrillation Is Not an Independent Determinant of Mortality Among Critically Ill Acute Ischemic Stroke Patients: A Propensity Score-Matched Analysis From the MIMIC-IV Database* 同樣是針對急性缺血性中風患者，提出心房顫動並非這類患者死亡的獨立決定因子，除了共病、CCI 分數以及抗血小板藥物外，這篇使用了大量的相關指標，如血紅蛋白、血小板、鈉、鉀、白血球記數平均血壓、心跳、體溫等。

四、申請及 Demo

因 MIMIC 資料庫申請較為繁雜且瑣碎，因此本篇不再進行 MIMIC 資料庫申請的解說，網路上也有需多教學網站可供參閱，這裡僅提供 MIMIC 申請網址：<https://mimic.mit.edu/>，目前 MIMIC 官方有釋出 100 筆病患資料做為 DEMO 讓使用者測試，後續幾期的操作也將會以此資料來進行，若有需要請參考 <https://physionet.org/content/mimic-iv-demo-omop/0.9/>

五、總結

本期針對 MIMIC-IV 進行簡單的介紹，後續也將使用 DEMO 資料來進行操作示範，此資料庫有別於健保資料庫，不僅提供了疾病及治療處方，更包含了病患的相關訊息及醫學非結構化報告的部分，使得研究上資料更

加完整。

六、參考資料

1. <https://www.frontiersin.org/articles/10.3389/frai.2021.691626/full>
2. <https://ppfocus.com/0/te99974dc.html>
3. MIMIC 官方網站 <https://mimic.mit.edu/docs/>
4. Zhou S, Zeng Z, Wei H, Sha T, An S. Early combination of albumin with crystalloids administration might be beneficial for the survival of septic patients: a retrospective analysis from MIMIC-IV database. *Ann Intensive Care*. 2021;11(1):42. Published 2021 Mar 10. doi:10.1186/s13613-021-00830-8
5. Hirano Y, Shinmoto K, Okada Y, Suga K, Bombard J, Murahata S, Shrestha M, Ocheja P, Tanaka A. Machine Learning Approach to Predict Positive Screening of Methicillin-Resistant Staphylococcus aureus During Mechanical Ventilation Using Synthetic Dataset From MIMIC-IV Database. *Front Med (Lausanne)*. 2021 Nov 16;8:694520. doi: 10.3389/fmed.2021.694520. PMID: 34869405; PMCID: PMC8635043.
6. Na Pattalung T, Ingviya T, Chaichulee S. Feature Explanations in Recurrent Neural Networks for Predicting Risk of Mortality in Intensive Care Patients. *J Pers Med*. 2021 Sep 19;11(9):934. doi: 10.3390/jpm11090934. PMID: 34575711; PMCID: PMC8465577.
7. Jhou, H. J., Chen, P. H., Yang, L. Y., Chang, S. H., & Lee, C. H. (2021). Plasma Anion Gap and Risk of In-Hospital Mortality in Patients with Acute Ischemic Stroke: Analysis from the MIMIC-IV Database. *Journal of personalized medicine*, 11(10), 1004. <https://doi.org/10.3390/jpm11101004>
8. Wu CS, Chen PH, Chang SH, Lee CH, Yang LY, Chen YC, Jhou HJ. Atrial Fibrillation Is Not an Independent Determinant of Mortality Among Critically Ill

Acute Ischemic Stroke Patients: A Propensity Score-Matched Analysis From the MIMIC-IV Database. *Front Neurol.* 2022 Jan 17;12:730244. doi: 10.3389/fneur.2021.730244. PMID: 35111120; PMCID: PMC8801535.