

研究主題 - 馴服生理脈動：針對連續血糖數據的非定態時間序列分析與深度學習實踐

隨著感測技術與計算能力的飛速發展，醫療領域正產生前所未有的複雜生物醫學數據集。這些數據橫跨實驗室研究、臨床環境以及去中心化的居家照護場景，其核心特徵在於高度的非線性、非定態性（Nonstationarity）以及不確定性[1]。針對 2026 年大學部學生暑期研究計畫（USRP），本主題旨在引導學生深入探索時間序列預測的核心引擎，結合傳統統計模型 ARIMA 與現代深度學習架構 LSTM，並將其應用於具有重大臨床意義的連續血糖監測（CGM）數據分析中。本計畫旨在引導學生超越單純的演算法應用，深入探索時間序列分析的數學本質。過去研究結果指出，儘管深度學習在許多大數據領域佔據主導地位，但在生理指標的特定預測場景中，ARIMA 展現了出色的線性特徵識別能力與短期預測準確性。這反映了 LSTM 對非線性規律捕捉能力強，但需要龐大的訓練數據量來避免過度擬合，而 ARIMA 則在小規模且趨勢明顯的數據集上更具魯棒性[2]。

本計畫不僅強調數據科學的應用，更著重於「數據科學背後的科學」，即如何構建能夠編碼生理機制、現象行為和系統動力學的模式，進而引導演算法設計與統計推論，提供具備可解釋性與臨床可靠性的見解。

時間序列預測的跨領域重要性與科學背景

生物醫學時間序列往往具備複雜的動態特性。傳統的預測模型假設數據具有定態性，即均值與變異數不隨時間改變。然而，現實世界中的生理信號通常受多種因素干擾（如進食、運動、壓力與藥物代謝），呈現出同質或異質的非定態特徵。血糖值在進食、代謝與藥物干擾下，呈現出明顯的非定態特徵，即其統計性質（如均值與變異數）隨時間漂移。本計畫將重點放在如何「馴服」這些趨勢，並區分定態序列與非定態序列的建模策略。

線性與同質非定態建模：ARIMA 模型理論架構

ARIMA（整合移動平均自迴歸模型）是處理時間序列中趨勢與季節性的經典工具，被視為分析同質非定態數據的基石。ARIMA 模型由「三位一體」的組成部分構成，分別代表自迴歸（AR）、整合（I）與移動平均（MA），其參數化表示為 $ARIMA(p, d, q)$ 。

非線性動力學與深度學習：深度學習與 LSTM 架構

雖然 ARIMA 模型在處理線性趨勢與短期預測方面表現優異，但生物醫學數據往往包含複雜的非線性規律與長期依賴關係，這超出了傳統統計模型的捕捉範圍。因此，計畫的第二部分轉向基於神經網路的深度學習方法，特別是長短期記憶網路（LSTM）。這種結構允許 LSTM 在長時間跨度內保持穩定的資訊流，從而有效地捕捉非定態數據中的非線性動態模式。

臨床背景：住院連續血糖監測（CGM）指標與目標

本計畫的應用核心在於分析連續血糖監測數據。CGM 技術透過每 1 至 5 分鐘記錄一次皮下間質液中的葡萄糖濃度，提供了比傳統指尖採血（POC）更為完整的血糖軌跡。對於住院患者，特別是那些血糖波動劇烈或有低血糖風險的高危人群，血糖預測具有極高的臨床

價值。根據糖尿病技術學會的專家共識，住院臨床試驗應採用標準化的 CGM 指標來量化血糖控制品質：

- **目標範圍內時間 (Time-in-Range, TIR)**：定義為血糖處於 70–180 mg/dL 之間的時間百分比。TIR 是評估血糖穩定性的核心指標，但在住院場景下需根據患者的臨床急迫性進行個體化調整。
- **平均絕對相對誤差 (Mean Absolute Relative Difference, MARD)**：這是衡量 CGM 準確性的主要指標，透過與實驗室參考標準或 POC 血糖值進行對比得出。住院環境下的 MARD 通常在 9.4% 至 15.6% 之間[3]。

活動規劃與目標

將依循「從想法到分析再到報告」的完整研究循環，配合 NCTS 共同活動（週三下午博雅講談與交流報告），進行為期六週的密集訓練。計畫內容將圍繞以下兩個核心模組展開：

1. 研究方法論與科學基礎

學生將學習如何執行科學研究，這不僅包括演算法的堆疊，更包括對數據生成的生理背景進行深入理解。重點在於如何界定問題、設計實驗並對結果進行嚴謹的統計檢驗。

2. 實務操作：真實生物醫學案例

所有的理論都將應用於真實世界的案例中，透過 CGM 血糖信號資料達到做中學的目的。學生將在老師與助教的協助下，從數據清理、模型構建到性能評估進行端到端的實作。

週次	教學重點	實踐目標
1~2	研究方法與定態序列建模	掌握線性代數、殘差分析與 ARIMA 實作
3~4	深度學習與非定態信號分析	LSTM 模型、學習時頻分析；參加第三週期中交流報告
5~6	綜合案例研究與成果發表	針對 CGM 數據進行預測性能對比，撰寫研究報告並簡報。

學生先備知識與能力要求

本計畫具備高度跨學科性質，強烈建議具備數學、統計、電資或生醫背景之學生跨系組隊，以促進多元學術交流。

數學與統計背景

- **迴歸(regression)分析**：必須深入理解最小平方法 (OLS)、參數估計以及最關鍵的「殘差分析 (Residual Analysis)」。殘差分析是判斷模型是否完整捕捉數據規律的重要依據。
- **基礎數學**：包含線性代數、多變量微積分以及機率論。
- **深度學習基礎**：需了解類神經網路的基本原理，特別是前向傳播、反向傳播以及循環神經網路 (RNN) 的循環機制。

編程技能要求 - 學生必須熟練掌握以下工具之一：

- **R 語言**：建議使用 RStudio 搭配 R Markdown 進行分析。重點在於熟練使用 forecast 軟體包中的 auto.arima 函數來自動優化統計模型。
- **Python**：熟練使用 NumPy、Pandas 進行數據預處理，並能操作 PyTorch 或 TensorFlow 框架來搭建 LSTM 模型。

推薦/活動前必備讀物

為了在暑期計畫中取得卓越成果，學生應在計畫開始前研讀以下關鍵文獻與工具文檔：

1. **統計建模基礎**：《Forecasting: Principles and Practice》（第 2 版或第 3 版），由 Hyndman 教授撰寫。此書對 ARIMA 模型的原理與實作有極佳的闡述，特別是第 8 章與第 9 章。
2. **臨床共識文件**：《A Consensus Statement for Continuous Glucose Monitoring Metrics for Inpatient Clinical Trials》。這將幫助學生理解數據的醫療意義，以及為何預防 Level 2 低血糖是臨床的首要任務。
3. **深度學習導論**：《Dive into Deep Learning》（d2l.ai），專注於序列模型、RNN 與 LSTM 的章節，了解梯度截斷與參數初始化等數學推導細節。

結論：跨學科合作的必然性

2026 年大學部暑期研究計畫（USRP）是一個獨特的平台，它打破了學科間的藩籬，將數學的嚴謹、統計的穩健與深度學習的靈活結合在一起。對於學生而言，這是一次從學術理論邁向臨床應用、從「怎麼做」深入到「為什麼這樣做」的科學洗禮。透過對連續血糖監測數據的分析，學生不僅能掌握 ARIMA 與 LSTM 等先進工具，更能深刻體會數據科學在改善人類健康方面的巨大潛力。

透過嚴格執行此計畫書中規劃的教學模組與實踐循環，我們有信心培養出既懂數學本質，又能解決醫療實際問題的跨領域研究者。這不僅符合 NCTS 對科學卓越的追求，更是應對未來智慧醫療挑戰的最佳策略方案。我們期待在這個暑假，見證新一代生物醫學人工智慧人才的誕生。

相關參考資料

- [1] M. Cheng *et al.*, “A Comprehensive Survey of Time Series Forecasting: Concepts, Challenges, and Future Directions,” Apr. 10, 2025. doi: 10.36227/techrxiv.174430535.53879341/v1.
- [2] U. Krishnamoorthy, V. Karthika, M. K. Mathumitha, H. Panchal, V. K. S. Jatti, and A. Kumar, “Learned prediction of cholesterol and glucose using ARIMA and LSTM models – A comparison,” *Results in Control and Optimization*, vol. 14, p. 100362, Mar. 2024, doi: 10.1016/j.rico.2023.100362.
- [3] E. K. Spanakis *et al.*, “A Consensus Statement for Continuous Glucose Monitoring Metrics for Inpatient Clinical Trials,” *J. Diabetes Sci. Technol.*, vol. 17, no. 6, pp. 1527–1552, Nov. 2023, doi: 10.1177/19322968231191104.