

## 比較多種統計模型與深度學習模型於 PM<sub>2.5</sub> 預測

翁政雄

國立彰化師範大學資訊管理學系

[peteweng@gmail.com](mailto:peteweng@gmail.com)

陳毓欣\*

國立彰化師範大學資訊管理學系

[chenyuxin821@gmail.com](mailto:chenyuxin821@gmail.com)

### 摘要

本研究旨在應用統計模型(ARIMA, Prophet)、機器學習(SVR, 回歸樹)與深度學習模型(CNN, RNN, LSTM, GRU, Transformer 等)對 PM<sub>2.5</sub> 進行預測分析。基於多種氣象變量(如風向、氣溫、氣壓)構建模型，並比較各模型的性能指標(MAE、RMSE、MAPE)。實驗結果表明，RNN 的基礎模型在預測中具有最佳表現，相較於其他基礎模型如 Prophet 及 SVR，顯著降低了誤差。本研究同時指出數據來源單一、模型泛化能力不足以及計算資源限制等挑戰，並提出未來改進方向。本研究證實深度學習模型在空氣品質預測中的潛力，為環境管理與政策制定提供了重要參考依據，有助於提升空氣污染預警能力。

關鍵詞：機器學習、統計模型、時間序列預測、混合神經網路、空汙預測

# Comparison of Various Statistical and Deep Learning Models for PM<sub>2.5</sub> Prediction

Weng, Cheng-Hsiung

National Changhua University of Education, Department of Information Management

[peteweng@gmail.com](mailto:peteweng@gmail.com)

Chen, Yu-Xin

National Changhua University of Education, Department of Information Management

[chenyuxin821@gmail.com](mailto:chenyuxin821@gmail.com)

## Abstract

This study aims to apply statistical models (ARIMA, Prophet), machine learning techniques (SVR, regression trees), and deep learning models (CNN, RNN, LSTM, GRU, Transformer, etc.) to forecast PM<sub>2.5</sub> concentrations. The models are constructed based on various meteorological variables (e.g., wind direction, temperature, atmospheric pressure) and their performance metrics (MAE, RMSE, MAPE) are compared. Experimental results indicate that foundational RNN models exhibit superior performance in forecasting, significantly reducing errors compared to other base models like Prophet and SVR. The study also highlights challenges such as limited data sources, insufficient model generalization capabilities, and computational resource constraints, and proposes directions for future improvements. This research confirms the potential of deep learning models in air quality forecasting, providing important references for environmental management and policy-making, thereby enhancing air pollution early warning capabilities.

Keywords: Machine Learning, Statistical Model, Time Series Forecasting, Hybrid Neural Network, Air Pollution Prediction

## 比較多種統計模型與深度學習模型於 PM<sub>2.5</sub> 預測

### 一、緒論

空氣污染問題由於城市的快速工業化和城鎮化，PM<sub>2.5</sub>(粒徑小於 2.5 微米的懸浮顆粒物)是空氣污染中的主要成分之一，來源包括燃煤、汽車尾氣排放、工業生產以及生物質燃燒。由於其微小顆粒可深入呼吸系統，PM<sub>2.5</sub> 被證實對人體健康有顯著的長期危害，如引發哮喘、慢性支氣管炎和心血管疾病。研究動機源於現有 PM<sub>2.5</sub> 預測方法在應對複雜非線性變化和捕捉多重因素影響時具有一定的局限性。

本研究旨在應用並比較多種統計模型與深度學習模型，包括 Prophet 模型、自迴歸綜合移動平均模型(Autoregressive Integrated Moving Average model, ARIMA)、支持向量迴歸(Support Vector Regression, SVR)、迴歸樹(Decision Tree Regressor)、卷積神經網絡(Convolutional Neural Network, CNN)、循環神經網路(Recurrent neural network, RNN)、長短期記憶(Long Short-Term Memory, LSTM)、門控循環單元(Gated Recurrent Unit, GRU)，以及 Transformer 模型，在本研究將這三類分成「統計模型」、「機器學習模型」、「深度學習模型」進行實驗，以實現 PM<sub>2.5</sub> 濃度的精確預測並比較其誤差表現。這些學習技術擁有處理高維度與非線性數據的能力，並具備出色的時間序列建模性能，對於解析和預測空氣污染濃度的時間序列變化尤為有效。

其他章節規劃如下：第二章為文獻探討；第三章為研究方法；第四章為實驗，使用 Song (2017)所提供之北京 PM<sub>2.5</sub> 資料集，建立人工智慧模型，進而評估並比較研究模型之預測效能；第五章結論，探討本研究結果並說明未來研究展望。

### 二、文獻探討

過去對於 PM<sub>2.5</sub> 空氣污染的研究大致可以被分為來源解析及健康影響三個領域，並且透過不同的方法如統計分析、機器學習或深度學習等技術進行研究。本章節先探討 PM<sub>2.5</sub> 研究領域及相關研究方法過往的文獻回顧，後討論每個領域的研究重點及應用。

相關研究顯示：年平均 PM<sub>2.5</sub> 濃度與同期肺癌發病人數之間存在顯著的正相關，相關係數達到 0.87，表明 PM<sub>2.5</sub> 的季度平均濃度與肺癌的發病率密切相關。北京市的 PM<sub>2.5</sub> 年均濃度約為 100  $\mu\text{g}/\text{m}^3$ ，且呈現非正態分佈，長期暴露在這種環境中會增加居民罹患肺癌的風險。PM<sub>2.5</sub> 及鈾 210 等空氣污染物被認為是導致肺癌的重要因素。雖然傳統上吸煙被視為肺癌的主要原因，但近年的臨床數據強調，空氣污染，特別是來自汽車排放、工業廢氣和揮發性化學物質的霧霾，才是更為顯著的影響因素。王鵬等人(2016)調查顯示，重工業城市的肺癌死亡率與空氣中鈾的濃度呈正相關，學術界對大氣污染在肺癌發病中的作用越來越重視。

預言家(Prophet)是由 Meta 團隊於 2017 年發表的時間序列預測模型，利用時

間數據資料中的趨勢、季節、假日及誤差影響項進行預測。預言家模型創建至今七年，研究多用於預測商品、股市等會隨時間變動的資料。

自我迴歸整合移動平均模型(ARIMA)是由 George Box & Gwilym Jenkins 於 1976 年所提出。通過整合自迴歸模型(Autoregressive, AR)和移動平均模型(Moving Average Model, MR)，並加上差分(integrated)的概念使模型能觀察過往的資料與誤差，處理非穩態的資料。作為時間序列預測方法之一，用於多領域中。

原始的支持向量機(SVM)演算法在 1963 年由蘇聯數學家弗拉基米爾·瓦普尼克(Vladimir Vapnik)和亞歷克塞·澤范蘭傑斯(Alexey Chervonenkis)共同發表。機器學習中 SVM 是一種強大的監督式學習模型，被廣泛應用於分類與迴歸分析等領域。它的獨特之處在於能夠有效地處理線性和非線性分類問題，並在許多實際應用中表現出色。SVM 用於分類問題，基於相同原理，SVR 用於迴歸問題，兩者皆著重於找到一個超平面以解決問題，並利用支援向量來定義超平面，使用者可根據不同的任務選擇適合的算法。

迴歸樹是一種廣受歡迎的機器學習方法，由 J. Ross Quinlan 在 1980 年代初期發展，對現代迴歸樹算法的發展產生了深遠影響。Quinlan 的著名算法包括 ID3 和 C4.5，這些算法至今仍是學術研究和實際應用中的重要基石(Quinlan, 1986; Salzberg, 1994)。迴歸樹不僅易於應用，還能在有限的算力下產生良好的結果，使其成為一種廣泛使用的機器學習技術(Layfield & Neri, 2023)。

迴歸樹透過樹狀結構來組織分類規則，其中每個分枝代表決策過程中的一個邏輯判斷，而每個葉節點代表一個分類決策。這種結構使得迴歸樹適合處理具有明顯邏輯分層的數據(Layfield & Neri, 2023)。在銷售預測領域，迴歸樹能有效的處理分類特徵和捕捉非線性關係，這使其能夠有效預測特定市場趨勢(Johannes & Alamsyah, 2021)。然而，迴歸樹存在過擬合的風險，尤其是在樹的深度很大或分枝過多時。為了解決此問題，分枝修剪技術被用來避免模型過度學習特定的訓練數據(Avoiding Overfitting of Decision Tree, 2007)。

卷積神經網路(Convolutional Neural Network, CNN)是由 Yann LeCun 於 1989 年提出的一種深度學習模型，利用局部特徵檢測、權重共享和池化壓縮的方式，使其既能處理大規模數據，又具有較強的泛化能力(LeCun et al., 1989)。

循環神經網路(Recurrent neural network, RNN)是一種強大的人工神經網路，設計用於處理順序資料(Kaur & Mohta, 2019)。與傳統神經網路不同，RNN 包含內部循環，這些循環引入了跨處理元素的遞迴動態和延遲激活依賴性(Hidas i et al., 2013)。這種內部記憶使 RNN 能夠隨著時間的推移維護訊息，對於涉及語音、文字、財務資料、音訊和視訊的任務特別有效(Kanagachidambaresan et al., 2021)。RNN 的運作方式是一次處理一個範例，保留反映長期依賴性的元素(Kaur & Mohta, 2019)。雖然多層感知器(MLP)具有多個層，但 RNN 通常具有定向循環，透過該循環將輸入轉換為輸出(Beysolow, 2017)。與其他演算法相比，這種獨特的結構使 RNN 能夠更深入地理解序列和意義，成為複雜序列資料分析和預測任務的首選(Kanagachidambaresan et al., 2021)。

長短期記憶網絡(Long Short-Term Memory, LSTM)是由 Hochreiter 和 Schmidhuber 在 1997 年提出,並且被設計成能夠捕捉和保留長期資訊,是一種非常適合用於時間序列預測的模型。LSTM 的設計旨在克服傳統循環神經網絡(RNN)所面臨的「梯度消失和梯度爆炸」問題,從而能夠更有效地處理長期依賴性。LSTM 經常作為混合模型的一部分,例如 CNN-LSTM。這類模型結合了 CNN 在特徵提取方面的優勢和 LSTM 在時間序列建模方面的能力,適合用於更為複雜的預測任務,如金融市場預測和環境變化(張曉妮, 2024; 葉國毅, 2024)。

門控循環單元(Gated Recurrent Unit, GRU)由韓國科學家 Kyunghyun Cho 於 2014 年在其博士後研究期間發表,是一種改進版的循環神經網絡(RNN)。GRU 的目的在於解決標準 RNN 中常見的梯度消失與梯度爆炸問題,從而更有效地學習長期依賴。相比於 LSTM, GRU 具有更高的計算效率,克服 LSTM 計算速度較慢的缺點,是一種改良過的深度學習模型。GRU 在解決時序依賴和序列預測問題中展現出極高的靈活性,特別是在涉及長期記憶的任務中尤為高效。由於其結構比 LSTM 更加簡潔, GRU 提供更高的計算效率和更快的訓練速度,使其成為處理各類序列數據時的理想選擇。

在深度學習領域,Transformer 模型因其出色的性能和靈活性而受到廣泛關注。隨著計算需求的增加,研究人員開始尋找提高 Transformer 計算效率的方法,特別是在乘累加運算(MAC)方面。傳統的優化技術如知識蒸餾、剪枝和量化已經被深入探討,但隨著對能耗和計算效率要求的提高,查找-累加(LAC)運算等新方法逐漸成為研究的焦點。Transformer 模型的關鍵在於其多頭自注意力機制,該機制依賴於大量的乘法和加法運算。運用多頭自注意力機制除了可以並行計算、提高計算效率外,更能擴展自注意力,使得透過多個注意力頭來捕捉輸入序列之間的多樣性關係(Vaswani et al., 2017)。

### 三、 研究方法

本研究在「深度學習模型」這類型中導入了多個不同的模型架構作為基礎,並在每個架構中進行了進一步的模型改良與優化,這些優化包括單獨使用或混合使用以下策略:堆疊模型(Stacked Model)、雙向模型(Bidirectional Model)、堆疊雙向模型(Stacked Bidirectional Model)、以及多策略混合模型(Hybrid Model)。

以下分別依序說明本研究所採用的統計模型(Prophet, ARIMA)、機器學習模型(SVR, 迴歸樹)及深度學習模型(CNN, RNN, LSTM, GRU, Transformer):

#### 1. 統計模型

Prophet 是由四個函數所組成的加法模型(Taylor & Letham, 2017):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

在公式 1 中,  $g(t)$  代表趨勢項、 $s(t)$  表季節項、 $h(t)$  表假日項、 $\varepsilon_t$  表誤差項。而在趨勢項中, Prophet 也提供了兩種趨勢函數供使用者使用,當 Linear 用於不飽和的預測時,使用公式 2,相反地,則使用公式 3,其中,  $C(t)$  是不同時間點的

「容量」上界。

$$g(t) = k \cdot t + m \quad (2)$$

or

$$g(t) = \frac{C(t)}{1 + \exp(-k \cdot (t - m))} \quad (3)$$

季節項則應用 Fourier series 來描述，在公式 4 中， $P$  是週期的長度（年的週期是 365.25 天，周的週期是 7 天...），透過控制參數  $a_n, b_n$ ，可以近似出任何一個周期函數。

$$s(t) = \sum_n^N (a_n \cos(\frac{2\pi n t}{P}) + b_n \sin(\frac{2\pi n t}{P})) \quad (4)$$

假日項則假設每種假日  $D_i$  都會帶來不同的影響。

$$h(t) = \sum_i^D \kappa_i \cdot 1(t \in D_i) \quad (5)$$

Prophet 模型即是透過擬合上述四項函數得到時間序列預測值。根據模型特性，Prophet 雖無法透過更多面的資訊預測產品銷售資料(如天氣狀況、商品價格、門市分布情況)，但在時間序列分析上已提供豐富的預測能力。另外值得一提的是，Prophet 還引入了轉折點(change point)的想法，也就是讓趨勢函數在不同的時間區間內，能有不同的增長率  $k$ 。

$$k(t) = k_0 + \sum_j 1(t \leq t_j) \cdot \delta_j \quad (6)$$

在 Prophet 模型建構中，也曾遇到誤差過大、數據過於精細等情況，後續則改使用每日平均作為預測基值，而對於已知資料誤差測量則是採用性能指標(performance\_metric)作為評估標準，其中包含均方根誤差(RMSE)、平均絕對誤差(MAE)、平均絕對百分比誤差(MAPE)。

自我迴歸整合移動平均模型(ARIMA)結合了自迴歸模型(AR)、差分(I)和移動平均模型(MA)三個組件。其中一個重要的組成部分「自迴歸模型(AR)」，是時間分析序列中常見的一種模型。該模型基於時間序列數據，將 AR 模型、MA 模型和 I 結合可得 ARIMA 模型，記成 ARIMA(p, d, q)，可表示公式 4：

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (7)$$

## 2. 機器學習模型

支援向量迴歸(SVR)適用於迴歸問題。與 SVM 類似，SVR 也是通過找到一個超平面，但其目標是使得數據點與該超平面的距離盡可能小，同時保證超平面兩側的誤差都在一個可接受的範圍內(Drucker et al., 1997)。支援向量在 SVR 中

扮演著重要角色，它們是那些與預測函數距離最接近的數據點，用於限制預測誤差。同樣地，SVR 可以處理線性和非線性迴歸問題，並且也可以通過核函數擴展到高維空間中。超平面的一般方程式如下：

$$w \cdot x + b = 0 \quad (8)$$

在公式 3 中， $w$  是法向量(可以不用歸一化)、 $b$  是偏移量。對於每個樣本點  $x_i$ ，其對應的類別標籤  $y_i$  為 1 或 -1。支持向量機的目標是找到這樣一個超平面，使得對於所有的樣本點  $x_i$ ，滿足以下條件：

$$y_i(w \cdot x_i + b) \geq 1 \quad (9)$$

同時最大化  $\frac{2}{\|w\|}$ ，這樣做可以確保超平面到兩個類別中最近的樣本點的距離最大化。

迴歸樹(Decision Tree Regressor)的建構過程包括特徵選擇、樹的建構和樹的剪枝，特徵選擇即為從當前數據的特徵中，找出最佳分割特徵作為結點的劃分標準。此步驟通過評估每個特徵的資訊增益(ID3)，資訊增益率(C4.5)或 Gini 係數(CART)，以衡量分割數據後純度或資訊量的變化。而樹的建構則是從根節點開始選擇特徵，數據集會被分割成兩個或多個子集，這些子集構成當前節點的子節點。接著，透過遞迴的方式，對每個子節點進行特徵選擇和分割，直到滿足停止條件，例如：所有特徵的資訊增益都非常小、無法再進行有效分割，或是達到預定深度等。最後，樹的剪枝意含模型在學習的過程中，為了盡可能正確的分類訓練樣本，不停的對結點進行劃分，這可能會導致分支過多，也就是模型過擬合。為了避免此狀況，可以採用剪枝技術來簡化迴歸樹。剪枝技術有兩種主要形式：預剪枝(停止樹的進一步生長)和後剪枝(移除樹中的部分節點) (Timofeev, 2004)。

### 3. 深度學習模型

CNN 在提取空間特徵方面具有優勢，尤其適合處理高維數據或具有強烈局部相關性的資料。透過堆疊與混用策略，CNN 的深層結構(如圖 2 所示)可以捕捉更抽象的特徵，提升模型的解釋能力。CNN 包含卷積層(Convolution)、池化層(Pooling)及全連接層：卷積層會使用過濾器(filter)在數據上滑動並進行卷積計算，逐步提取特徵並生成特徵圖；池化層用於減少特徵圖的大小，並保留重要重要特徵，常使用最大池化(Max Pooling)和平均池化(Average Pooling)；全連接層整合從卷積層和池化層學到的特徵，並將這些特徵轉換成最終的分類結果(LeCun et al., 1998; Krizhevsky et al., 2012)。

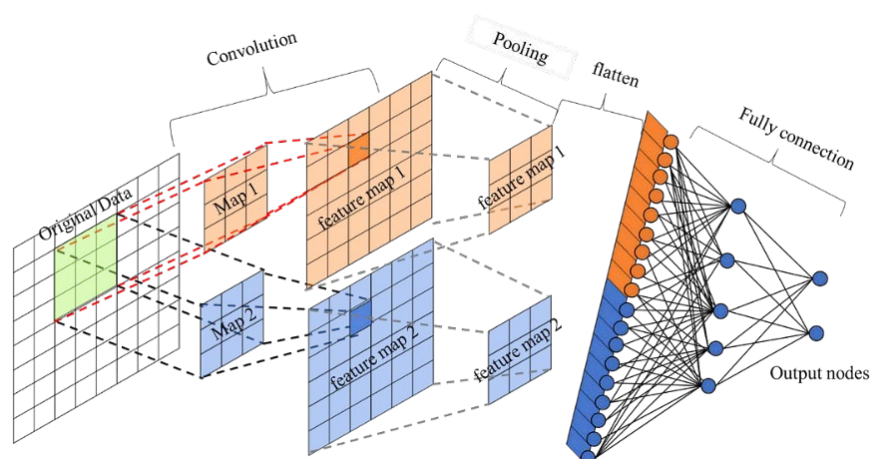


圖 1：CNN 示意圖

RNN 在序列數據的處理上表現出良好的基本能力，但其記憶長序列的效果受到梯度消失問題的限制。雙向與堆疊策略對 RNN 的改進效果顯著，有助於捕捉前後依賴性更強的特徵。但可惜的是記憶長序列的效果仍受到梯度消失問題的限制，其架構共三層：編碼器、解碼器以及雙向讀取(IBM, 2021)。

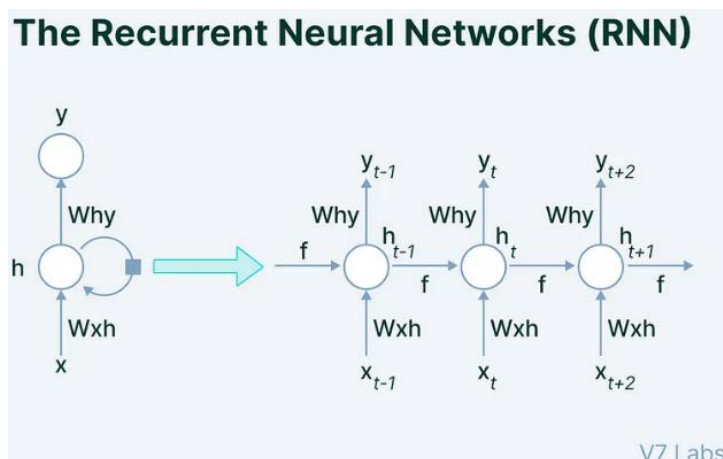


圖 2：RNN 示意圖(Mohsen, 2023)

LSTM 是一種非線性模型，改進了 RNN 在處理長序列時的記憶能力，對時間序列特徵具有更強的解釋度，但通常對於間隔與延遲非常長的重要事件較為拿手。LSTM 普遍用於自主語言辨識中，特別是在堆疊雙向設計中，LSTM 能夠同時捕捉資料的遠程依賴性和上下文特徵，對提升模型性能有明顯作用，也可用於構造更大型的深度神經網路，其構造則由細胞單元(cell)、輸入閥(input gate)、輸出閥(output gate)、遺忘閥(forget gate)組成。LSTM 的關鍵在於透過遺忘閥控制過去資訊的保留與丟棄，輸入閥決定新資訊是否被記錄，輸出閥則影響最終輸出的資訊內容。這種結構使 LSTM 能夠有效避免標準 RNN 常見的梯度消失(Vanishing Gradient)問題，確保長時間依賴關係能夠被保留(Hochreiter & Schmidhuber, 1997)。

。



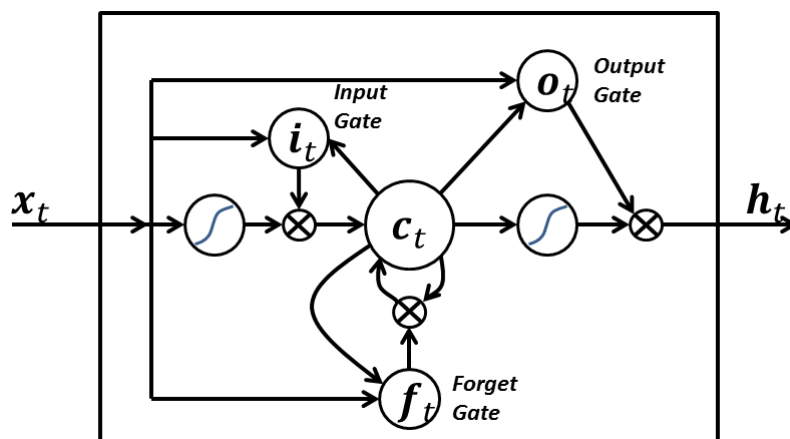


圖 3：LSTM 示意圖

GRU 相較於 LSTM，結構更為簡潔，計算效率更高。在處理短序列數據時，GRU 表現出更好的效率，而其性能與 LSTM 在多數實驗場景中表現相當。本研究模型利用 GRU 擅長捕捉序列時間依賴的特性，並通過全連接層進一步擴展學習能力，以提高對 PM<sub>2.5</sub> 時間序列的預測準確性。為了更進一步提升模型對 PM<sub>2.5</sub> 時間序列的預測能力，在 GRU 的基礎上加入了全連接層作為輸出層，藉此整合並學習更豐富的特徵表示。GRU 所擅長的序列依賴捕捉能夠有效將過去時刻的關鍵資訊保留下來，而全連接層則能在隱藏層特徵的基礎上進行更深層次的非線性映射，進一步優化預測結果。最終，將 GRU 隱藏層的輸出連接至全連接層，並透過誤差反向傳播進行權重更新，持續優化對 PM<sub>2.5</sub> 變化趨勢的捕捉能力，達成更精準的預測表現(Chung et al., 2014)。

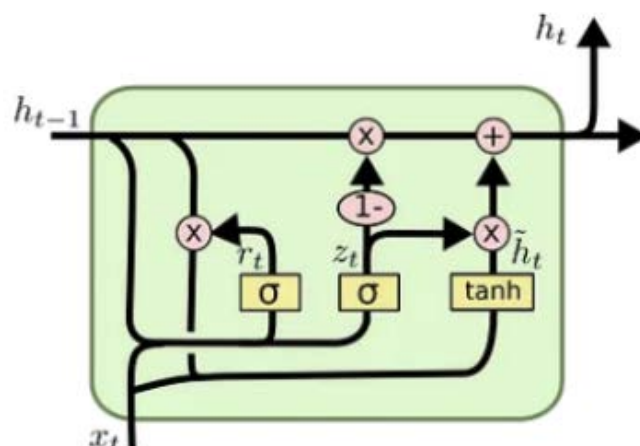


圖 4：GRU 示意圖

Transformer 是一種基於注意力機制的架構，擅長處理長距離依賴和多維特徵。透過混用策略結合堆疊與注意力機制，顯著提升其對於特定資料集的解釋能力。尤其在高維數據與大規模資料下，Transformer 展現極高的準確率與穩定性。為了進一步防止過擬合，在模型訓練過程中使用 Layer Normalization 與 Dropout 技術，以提升模型的泛化能力。相較於傳統 RNN 和 GRU，Transformer 能夠更有效地處理長距離時間依賴，並且透過平行運算顯著提升計算效率。在本實驗結果

中，Transformer 在 PM<sub>2.5</sub> 時間序列預測任務上表現優異，尤其在長期趨勢預測方面顯示出更高的準確性與穩定性(Vaswani et al., 2017)。

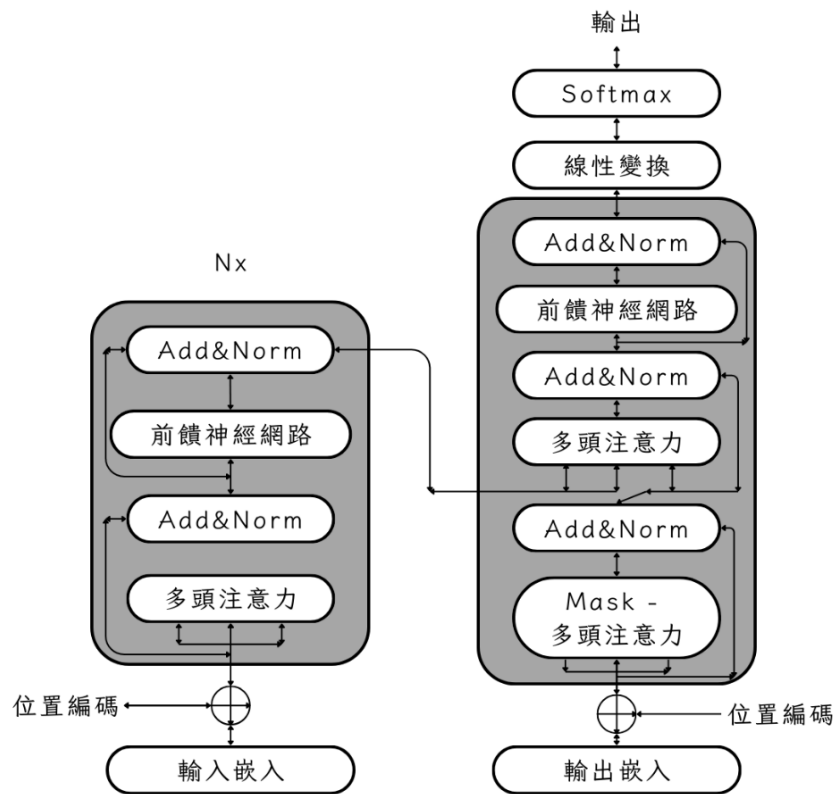


圖 5：Transformer 模型

四、實驗結果

1. 衡量指標

本研究為了評估及比較模型的預測性能與準確性，選用平均絕對誤差(Mean Absolute Error, MAE)、均方根誤差(Root Mean Squared Error, RMSE)與平均絕對百分比誤差( Mean Absolute Percentage Error, MAPE )這三項誤差值作為主要的衡量指標。這三項指標在時間序列預測和迴歸問題中常被廣泛應用，被普遍認為能夠提供對於模型準確性及穩定性的全面評價以便使用者觀察及運用(Vandeput, 2019.; Data Science Stack Exchange, 2018)。

表 1 指標意義及公式比較

指標	意義	公式
平均絕對誤差 (MAE)	計算了預測值與實際觀測值之間的平均偏差，顯示模型整體預測的精準度，反映模型在常見情況下的穩定性。	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $

均 方 根 誤 差 (RMSE)	計算了預測值與實際值偏差的平方平均後的平方根，揭示模型在處理異常數據或急劇變化時的能力與不足	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
平均絕對百分比誤差(MAPE)	用於評估模型預測準確性的指標且可用於不同單位的數據集比較的原因	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right  \times 100\%$

2. 參數與超參數設置

在時間序列預測中，ARIMA 模型是一種傳統的統計方法，需要進行平穩性檢驗，若平穩性檢驗顯示不平穩，則需要進行差分及參數估計處理以確保模型的有效性。然而，在本研究中，正規化並未顯著提升模型的性能，其決定係數最高僅為 0.01，遠低於平穩性檢驗的臨界值標準 0.05，且後續於差分與參數估計處理後仍無明顯變化。

在迴歸樹的實驗中，最基本的模型即表現出不俗的解釋能力，甚至超越了上述已提到的模型，同樣地，迴歸樹也遇到與 ARIMA 相同的問題——標準化後解釋能力無顯著變化，其數值始終在 0.58 附近徘徊。經過持續實驗，我們透過特徵篩選與最大深度的變化使決定係數提高至 0.75。這表明迴歸樹在特徵選擇和模型調整後能夠取得更好的預測效果。

在 RNN 系列研究中，最基礎的 RNN 模型取得了最佳結果，具體參數包括隱藏層神經元數為 64、批次大小數為 8、訓練週期數為 50，MAE 為 13.020。同時，LSTM 模型的最佳結果則為出現在堆疊 LSTM，具體參數為 LSTM 層神經元數分別為 64 及 128、批次大小數為 64、訓練週期數為 30，MAE 為 14.426。

最後，GRU 模型的最佳結果出現在 GRU 層神經元數為 256、隱藏層數分別為 64 及 2、批次大小為 64、訓練週期為 30，MAE 為 13.896。值得注意的是，由於雙向 LSTM 即為基於變換器的雙向編碼器表示技術(Bidirectional Encoder Representations from Transformers, BERT)，但此種特色在於 BERT 僅使用編碼器部分，而這將會造成模型可能無法預測未來資料而只能將過去空缺資料補齊，因此本次實驗中的變化僅實驗了堆疊 Transformer 模型，而其具體參數為隱藏層神經元數為 128、Dense 層神經元數為 1、批次大小數為 2、dropout 設置為 0.01、訓練週期數為 1、隱藏層維度為 8、key\_dim 數為 64、注意力頭數為 2，MAE 為 19.936。

3. 實驗比較

本研究使用北京 PM<sub>2.5</sub> 作為訓練資料，旨在分析其變化趨勢與可能的影響因素。本研究選取近年來的歷史資料，結合多種機器學習模型進行訓練，探討風向、氣溫、氣壓等變量與 PM<sub>2.5</sub> 濃度之間的關聯性。訓練資料使用 2010/01/01 至

2014/12/31 期間內的細懸浮顆粒資料。以去掉最後七天的資料作為訓練資料集，最後七天作為測試資料集。每筆訓練資料包含標準化後的各項天氣細項指標以及對應的細懸浮顆粒指數，並依照日期、時間序列依序編制為訓練資料。

為了驗證本次提出的模型架構是否能充分說明資料，除了深度學習的神經網路模型，例：RNN、LSTM、GRU、Transformer，本研究以此為基礎做出Bidirectional、Stacked、混用等變化。而後將整體的訓練資料依照日期、時間序列輸入個別模型後，計算各模型的 MAE、RMSE、MAPE，驗證此模型架構與其他基礎模型相比較下是否有更好的效果，實驗結果如表 2。

表 2 所有模型指標比較

模型架構	MAE	RMSE	MAPE
Prophet	58.74	79.73	85.42%
ARIMA	69.42	94.21	92.37%
SVR	57.04	88.62	78.81%
Decision Tree Regressor	46.27	26.05	65.29%
<b><u>RNN</u></b>	<b><u>13.020</u></b>	<b><u>26.634</u></b>	<b><u>29.85%</u></b>
B – RNN	14.620	24.972	24.31%
S – RNN	14.268	24.482	25.07%
SB – RNN	17.061	28.774	28.48%
RNN – CNN	18.548	26.390	55.19%
RNN – LSTM	14.833	25.174	24.66%
RNN – GRU	15.109	26.102	25.13%
LSTM	14.481	24.219	24.82%
B – LSTM	14.663	25.161	24.51%
<b><u>S – LSTM</u></b>	<b><u>14.410</u></b>	<b><u>24.574</u></b>	<b><u>24.02%</u></b>
SB – LSTM	14.546	24.791	24.09%
LSTM – CNN	14.956	25.622	29.20%
LSTM – RNN	14.749	24.706	26.39%
LSTM – GRU	14.689	24.763	24.51%
<b><u>GRU</u></b>	<b><u>13.896</u></b>	<b><u>23.832</u></b>	<b><u>23.20%</u></b>
B – GRU	14.072	24.501	24.24%
S – GRU	14.068	24.809	25.03%
SB – GRU	14.277	24.409	23.62%
GRU – CNN	16.460	25.400	39.13%
GRU – RNN	14.806	24.827	27.64%
GRU – LSTM	14.528	24.724	24.70%
Transformer	20.406	33.442	29.77%
<b><u>S – Transformer</u></b>	<b><u>19.936</u></b>	<b><u>35.586</u></b>	<b><u>22.39%</u></b>

註：B – ：Bidirectional、S – ：Stacked、SB – ：Stacked Bidirectional

從上表中可以得知 RNN 的表現相對於其他基礎模型比較之下有較好的效果，顯示出以 RNN 作為預測本數據集會得到最小誤差的預測，而當撇除掉數據集的限制下，S-Transformer 的程式設計與貼合度則是最佳的。以此可知，在使用這些模型做為未來預測中可能取得與現實將相符的結果。

## 五、 結論

本研究基於多種深度學習模型對北京 PM<sub>2.5</sub> 濃度進行了詳細的預測分析，結果顯示以 RNN 為基礎的變化模型(如 B-RNN、S-RNN)在精度與穩定性上表現優異，特別是在 MAE、RMSE 與 MAPE 等評估指標中均顯示出相對較低的誤差，證明其在細懸浮顆粒濃度預測中的潛力。相比之下，傳統模型(如 Prophet、ARIMA)在處理複雜時間序列與高非線性數據時表現不如深度學習模型。本研究不僅驗證了深度學習模型在 PM<sub>2.5</sub> 預測中的應用價值，亦為未來改進相關模型與方法提供了參考。隨著研究的深入，這些模型有望在空氣品質預測與管理中發揮更大的作用，為環境保護與公共健康政策提供有力支持。

關於研究限制部分，本研究仍受到部分限制，例如數據來源的局限性、模型泛化能力的不足及計算資源限制對模型優化的影響。首先，數據來源主要集中於北京地區，可能無法完全反映其他地區的污染情況，限制了模型的泛化能力。此外，數據中可能存在缺失值與異常值，對模型的準確性造成影響，特別是在長期預測中，誤差可能進一步累積。此外，模型超參數調整的範圍受計算資源限制，未能窮舉所有可能的組合，而這將影響模型性能的最優化。同時，特徵選擇僅限於可獲取的氣象與污染數據，未能考慮交通流量、工業活動或政策變化等潛在因素。此外，部分模型(如 Transformer)對計算資源需求較高，導致其性能可能未完全發揮。本研究主要聚焦於短期預測，對長期趨勢的捕捉能力相對有限。然而，這些限制為未來研究提供了方向，包括引入更多元的數據來源、融合多種模型架構、優化超參數調整策略以及提升模型的解釋性與資源高效性。

關於未來研究部分可在以下幾方面展開：首先，引入多地區、多樣化的數據，結合數據增強技術(如合成數據生成)，提升模型的適用性與泛化能力。其次，探索模型融合策略(如 Transformer - CNN 或 Transformer - LSTM 混合架構)與自適應超參數調整技術(如貝葉斯優化或遺傳算法)，進一步優化模型性能。為提升模型的解釋性，可結合 SHAP 或 LIME 等方法分析模型對不同特徵的依賴性，盡可能降低 MAPE 以提高模型解釋度。此外，未來可融入社會經濟指標、交通流量與政策變化等非傳統數據，增強模型對多維度因素的綜合預測能力。在資源高效化方面，可設計輕量級模型，實現模型在邊緣設備上的實時部署，滿足即時預測需求。針對長期預測，可設計多步預測框架，並引入循環更新策略或情境假設，以改善長期趨勢的預測性能。最後，應將模型應用於實際空氣品質管理場景中，如輔助政府制定減排政策，或作為即時警報系統的一部分，以提升研究對實際問題的影響力。

## 參考文獻

- 王鵬、陳武義、林招鵬、賴律翰，2016，『北京地區霧霾之微懸浮粒子 PM<sub>2.5</sub> 污染與肺癌患病率關係分析』，台灣應用輻射與同位素雜誌，第十二卷・第四期：1405~1410 頁。
- 張曉妮，民 113，應用 CNN-LSTM 深度學習模型於預測台灣二氧化碳排放量，元智大學工業工程與管理學所碩士論文
- 葉國毅，民 113，使用多頭注意力機制的多重輸入 LSTM 股價預測模型，國立政治大學資訊科學系碩士在職專班碩士論文
- Baha M. Mohsen, "Developments of digital technologies related to supply chain management," *Procedia Computer Science*, vol. 220, pp. 788-795, 2023.
- Bay, J. 2024. "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information* (15:9), pp. 517–540.
- Bay, J., and Li, F. 2024. "Comparative Effectiveness of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) Architectures for Radiology Text Report Classification," PubMed Database.
- Box, G. E. P., and Jenkins, G. M. 1970. "Time series analysis: Forecasting and control," Holden-Day.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. 2014. "Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. 2014. "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS 2014 Workshop on Deep Learning*, December 2014.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., and Vapnik, V. 1997. "Support Vector Regression Machines," *Advances in Neural Information Processing Systems* (9), pp. 155–161.
- Hidasi, B., Karatzoglou, A., Baltrunas, L., and Tikk, D. 2016. "Session-Based Recommendations with Recurrent Neural Networks," *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*.
- Hochreiter, S., and Schmidhuber, J. 1997. "Long Short-Term Memory," *Neural Computation* (9:8), pp. 1735–1780.
- Jain, L. C., and Medsker, L. R. 1999. "Recurrent Neural Networks: Design and Applications," *Proceedings of the International Conference on Neural Networks*.
- Kaur, M., and Mohta, A. 2019. "A Review of Deep Learning with Recurrent Neural Network," *Proceedings of the International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 460–465.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems* (25), pp. 1097–1105.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. 1989. "Backpropagation Applied to Handwritten ZIP Code Recognition," *Neural Computation* (1:4), pp. 541–551.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE* (86:11), pp. 2278–2324.
- PubMed Team. 2024. "Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research," PubMed Database.
- Salehinejad, H., Baarbe, J., Sankar, S., Barfett, J., Colak, E., and Valaee, S. 2017. "Recent Advances in Recurrent Neural Networks," *arXiv preprint arXiv:1801.01078*.
- Sharma, S., Sharma, S., and Athaiya, A. 2020. "A Review of Deep Learning with Recurrent Neural Networks," *IEEE Access* (8), pp. 48750–48773.
- Taylor, S. J., and Letham, B. 2018. "Forecasting at Scale," *The American Statistician* (72:1), pp. 37–45.
- Timofeev, R. 2004. "Classification and Regression Trees (CART) Theory and Applications," Humboldt University, Berlin, pp. 1–40.
- Vapnik, V., Golowich, S. E., and Smola, A. 1997. "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," *Advances in Neural Information Processing Systems* (9), pp. 281–287.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. 2017. "Attention Is All You Need," *arXiv preprint arXiv:1706.03762*.
- Yamashita, K., and Suzuki, T. 2020. "Developing a Novel Recurrent Neural Network Architecture with Fewer Parameters and Good Learning Performance," *bioRxiv* (July), pp. 1–20.