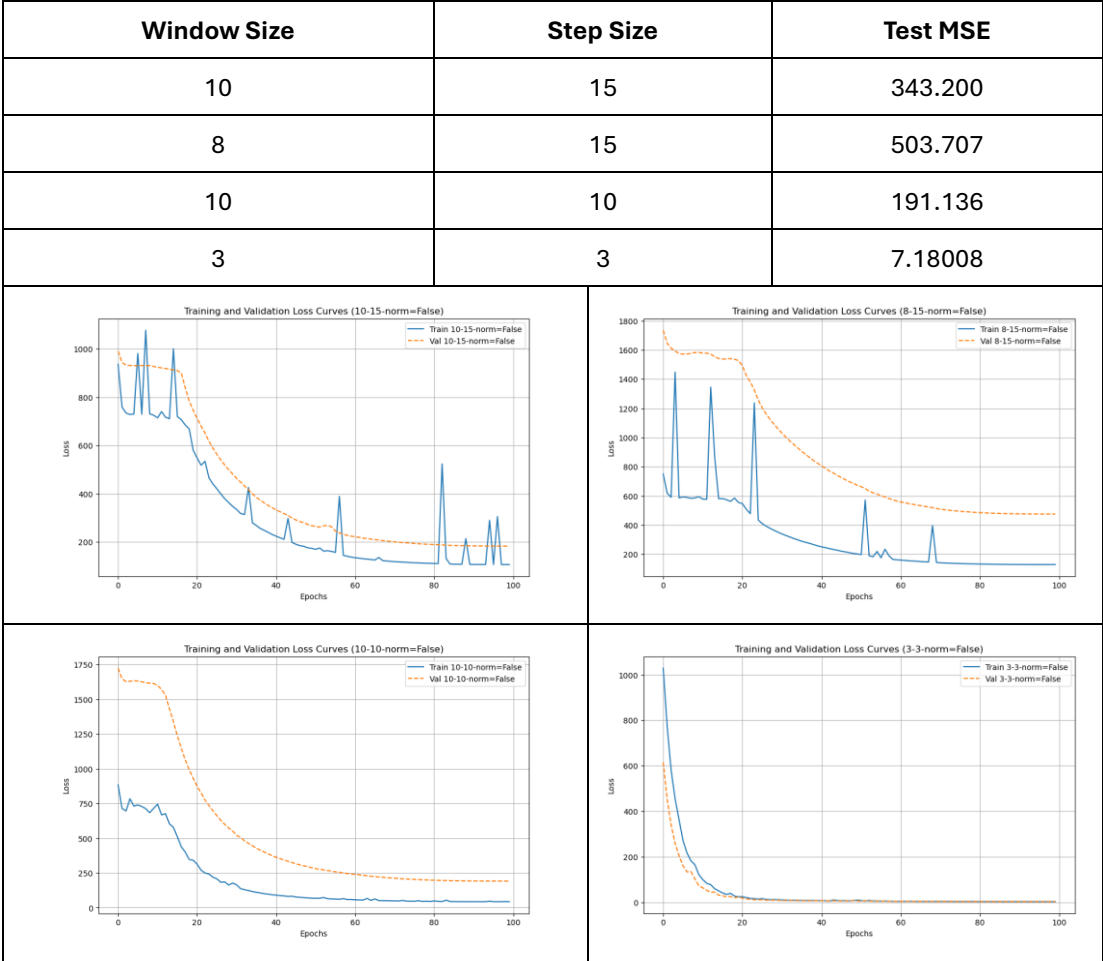


1.在本次作業中，視窗大小與步長分別設定為(10,15)、(8,15)、(10,10)、(3,3)等組合進行訓練。結果顯示，(3,3)組合的 Test MSE 最低，僅為 7.18008，表現最佳。而視窗較大（如 10 或 8）時，MSE 明顯偏高。

由此可知，小視窗、小步長能更精細捕捉時間序列特徵，提升預測準確度；

而視窗過大時，模型難以即時掌握細微變化，導致誤差增加。



2. (1)在本次實驗中，將"Volume"（成交量）納入特徵後，模型整體表現反而明顯下降。使用包含 Volume 的特徵組合時，Test MSE 大幅上升至超過 1000，顯示 Volume 在本資料集中無法有效提供有用資訊。可能原因是成交量數值變異劇烈、且與價格波動關聯性低，增加了模型訓練的難度與不穩定性。

因此，納入 Volume 未必總是有利，需要根據資料特性判斷。

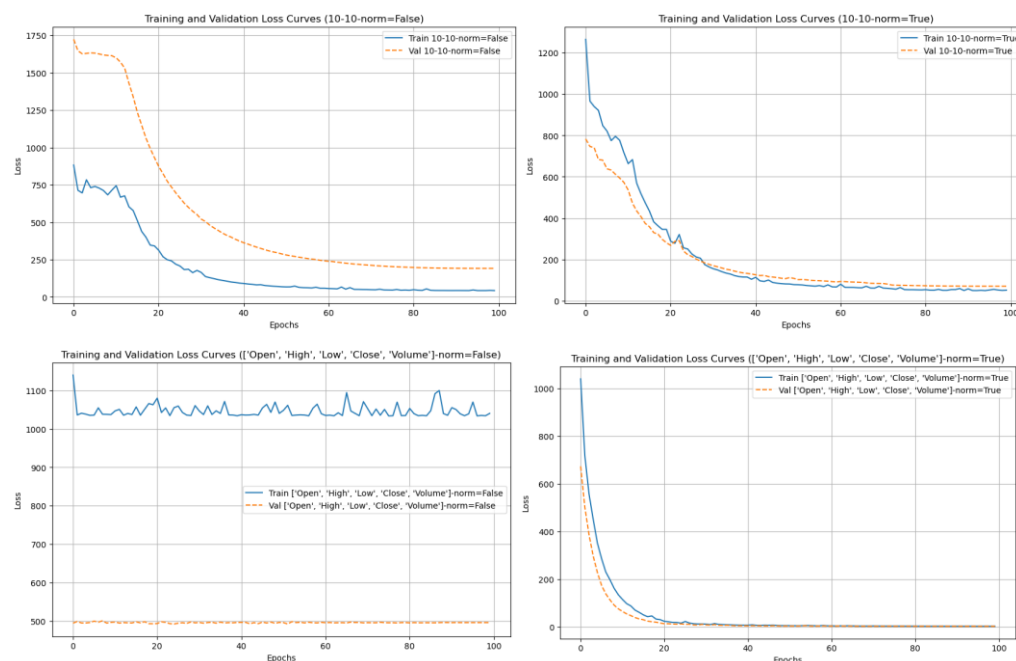
(2) 嘗試多組特徵組合後，發現僅使用['Open', 'High', 'Low', 'Close']四項價格資料的組合效果最佳，Test MSE 僅為 6.47029。選擇這組特徵的理由是，價格本身即反映了市場的主要變化趨勢，且數值範圍穩定，便於模型學習。相較之下，加入 Volume 會引入大量雜訊，降低預測效果。綜合比較後，選擇純價格特徵作為輸入能獲得最佳 MSE 表現。

特徵組合	Test MSE	特徵組合	Test MSE
------	----------	------	----------

['Open', 'High', 'Low', 'Close']	6.47029	['Open', 'Close', 'Volume']	1044.00
['Open', 'High', 'Low', 'Close', 'Volume']	1043.32	['High', 'Volume']	1043.47
['High', 'Low', 'Close']	7.67471		

3. 根據實驗結果，在視窗組合(3,3)時，未標準化的 Test MSE 為 7.18008，標準化後下降至 6.73344，顯示標準化可小幅提升模型精度。在特徵組合['Open', 'High', 'Low', 'Close']時，標準化前 MSE 為 6.47029，標準化後略微上升至 7.2242。整體來看，標準化在含有 Volume 這類尺度變異大的特徵時特別重要，可防止梯度爆炸、穩定模型訓練，在特徵尺度本就接近的情況下，標準化帶來的效益則相對有限。

=== Window/Step Normalization Comparison ===					=== Feature Combination Normalization Comparison ===				
	Window Size	Step Size	Normalized	Test MSE		Features	Normalized	Test MSE	
0	10	15	False	343.2	0	['Open', 'High', 'Low', 'Close']	False	6.47029	
1	10	15	True	340.24	1	['Open', 'High', 'Low', 'Close']	True	7.24242	
2	8	15	False	503.707	2	['Open', 'High', 'Low', 'Close', 'Volume']	False	1043.32	
3	8	15	True	304.46	3	['Open', 'High', 'Low', 'Close', 'Volume']	True	7.15588	
4	10	10	False	191.136	4	['High', 'Low', 'Close']	False	7.67471	
5	10	10	True	150.045	5	['High', 'Low', 'Close']	True	7.20411	
6	3	3	False	7.18008	6	['Open', 'Close', 'Volume']	False	1044	
7	3	3	True	6.73344	7	['Open', 'Close', 'Volume']	True	7.63133	
					8	['High', 'Volume']	False	1043.47	
					9	['High', 'Volume']	True	7.2925	



4. 在 Lab 4 中，將視窗大小設定為小於步長，可以避免相鄰視窗之間產生重疊，減少資料冗餘與重複資訊，從而降低模型過擬合的風險。這樣的設計也能提升特徵多樣性，使模型學習到更多獨立且豐富的模式。此外，減少重疊還能提升訓練與推論的效率，有助於即時應用。我認為，這種設計在需要提取非冗餘特徵的任務中是合理且正確的。

參考文獻：Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*. arXiv preprint arXiv:1803.01271.

5. 時間扭曲（Time Warping）是應用於時間序列資料的有效資料增強方法。

透過對時間軸進行隨機拉伸或壓縮，模擬不同速率下的樣本變化，能增加模型對時間尺度變異的適應能力。

Um 等人（2017）在穿戴式感測器資料上證明，時間扭曲可顯著提升帕金森病監測系統的準確率與模型泛化性，成為常見且有效的資料增強技術。

參考文獻：Um, T. T., Pfister, F. M. J., Pichler, D., et al. (2017). Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. *IJPRAI*. 6.

(i) 卷積模型（Convolution-based models）

卷積模型能接受可變長輸入，但推論時應與訓練時視窗大小保持一致，以維持感受野範圍穩定，確保特徵提取效果一致（Goodfellow et al., 2016）。

(ii) 循環神經網路模型（Recurrent-based models）

循環神經網路（如 LSTM、GRU）在推論時若視窗大小與訓練階段不同，容易出現記憶擴散（memory drift）與性能劣化現象。

Lim 等人（2021）提出，在時間序列預測任務中，應維持推論與訓練時一致的視窗大小，以提高預測準確率與穩定性。

(iii) Transformer 模型（Transformer-based models）

Transformer 模型依賴全序列注意力機制，推論時通常設置固定視窗長度，並結合滑動視窗（sliding window）或遮罩（masking）技術，以控制資源消耗（Vaswani et al., 2017）。

參考文獻：

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Lim, B., et al. (2021). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *International Journal of Forecasting*.

Vaswani, A., et al. (2017). Attention is All You Need. *NeurIPS*.