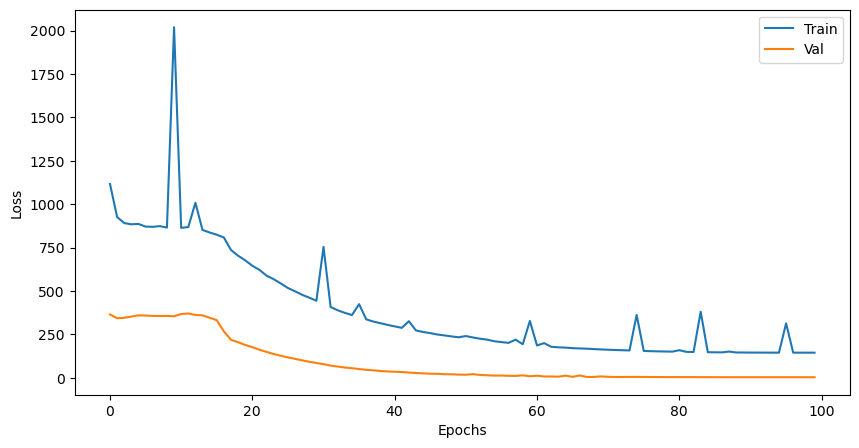
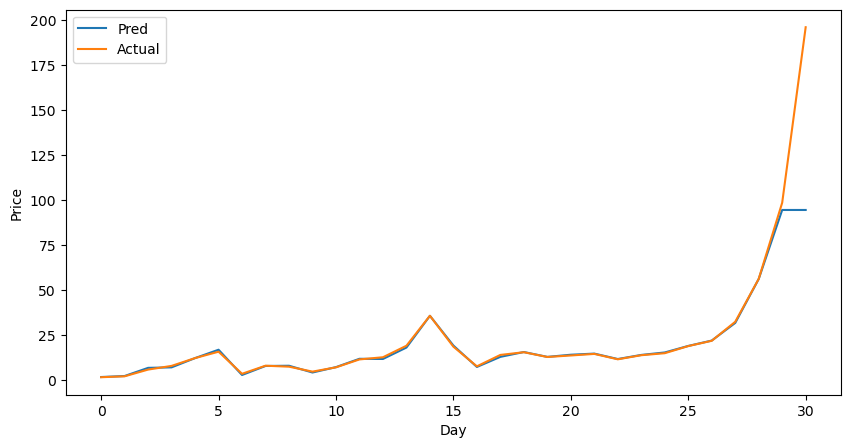
HW4

113034545 黃品諺

1. Original: window size = 10, steps = 15

MSE on test set: 333.88

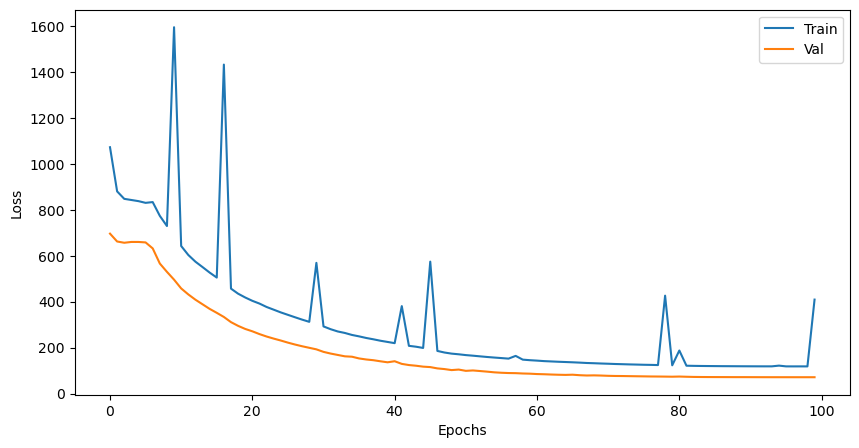
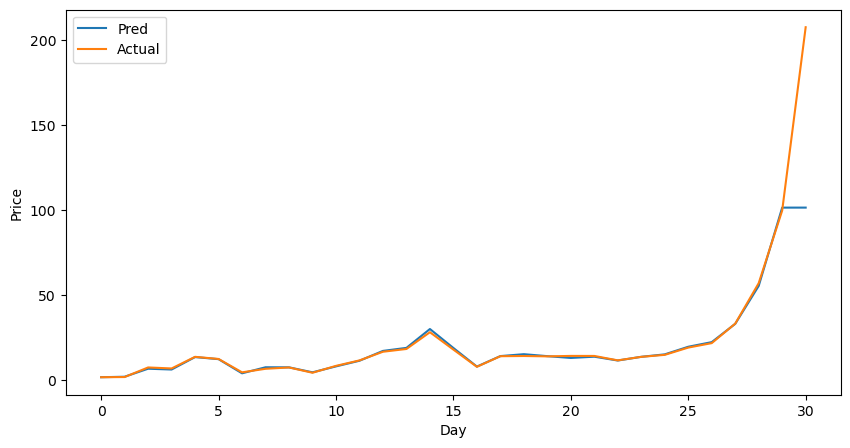
Loss: prediction:

Window size = 30, steps = 15

MSE on test set: 364.44

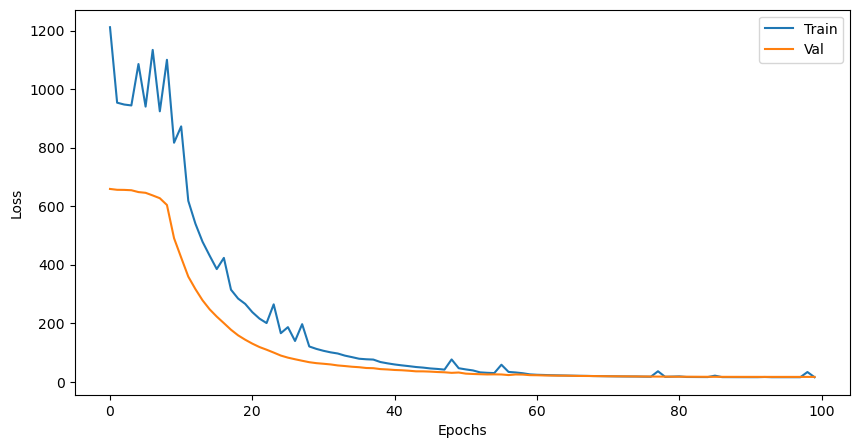
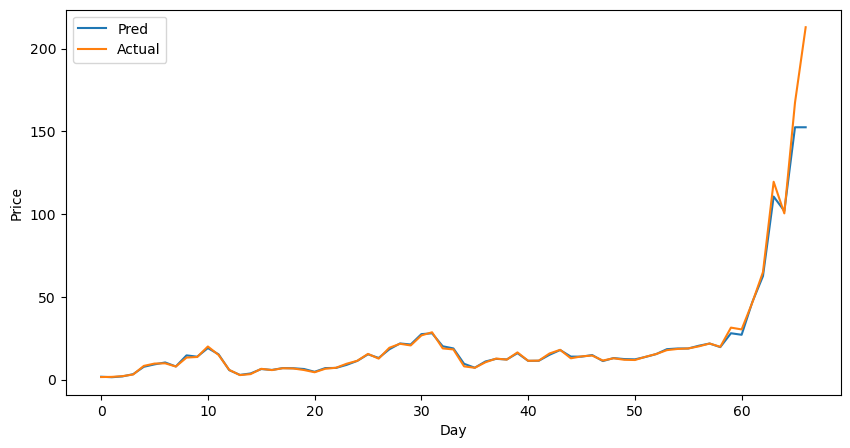
Loss: prediction:

Window size = 10, steps = 7

MSE on test set: 59.71

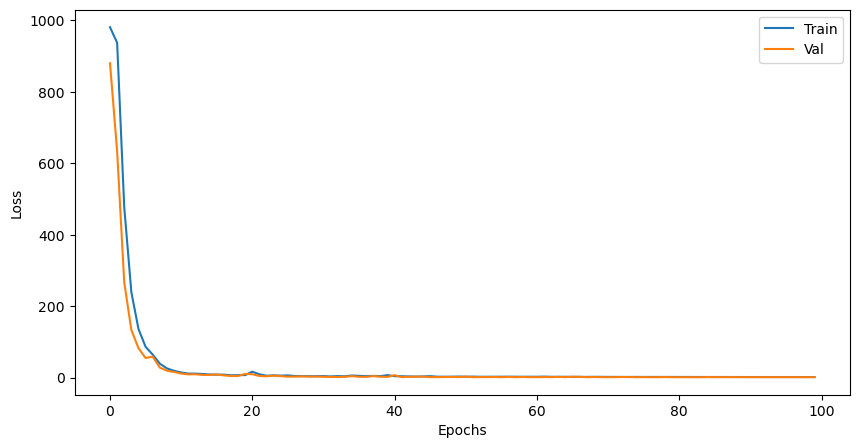
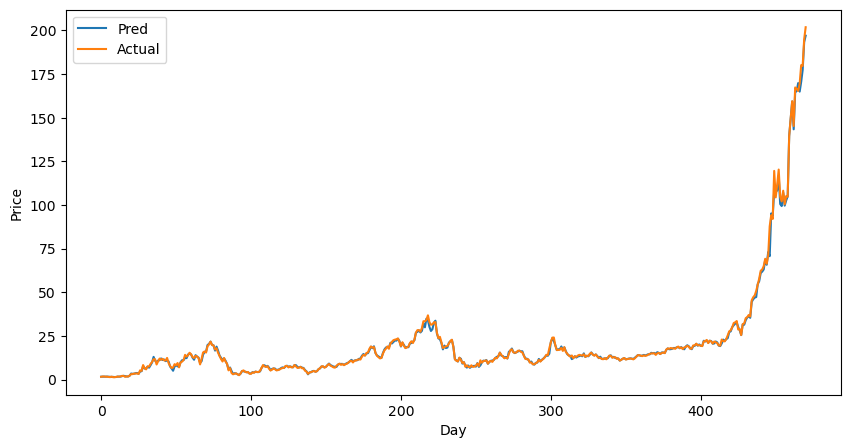
Loss: prediction:

Window size = 10, steps = 1

MSE on test set: 1.89

Loss: prediction:

|  |  |
| --- | --- |
| Window size / step | Test mse |
| 10 / 15 | 333.88 |
| 30 / 15 | 364.44 |
| 10 / 7 | 59.71 |
| 10 / 1 | 1.89 |

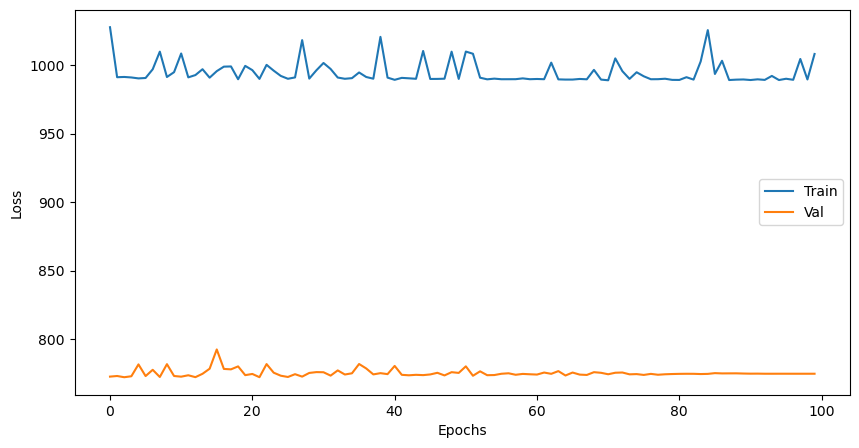
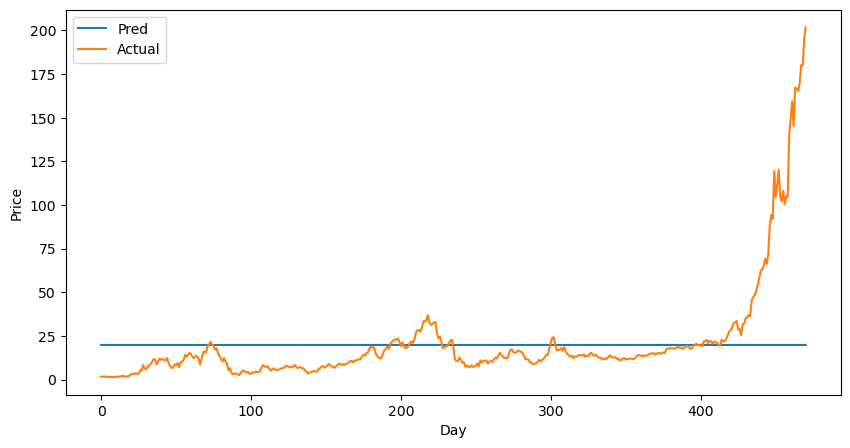
調整window size對於test集的mse影響並不大，且step較大時(約為10)，loss在訓練迭代時會突然上升，但將step調小這個情況回改善；而step越小，test mse也顯著變小，不過訓練時間也會上升。

1. (i)

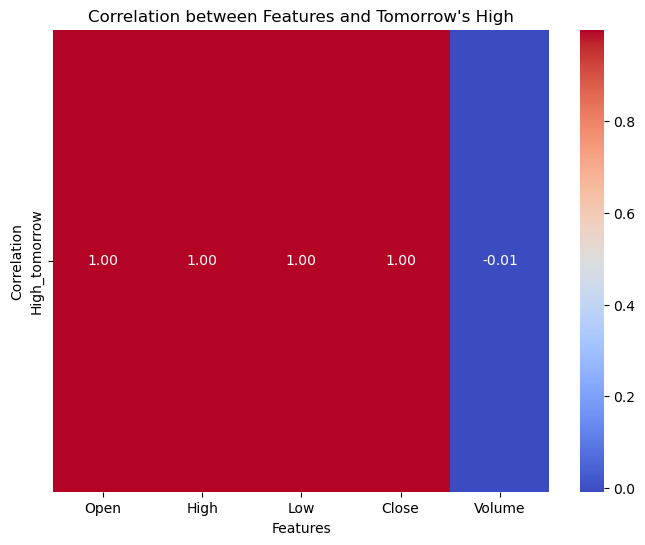
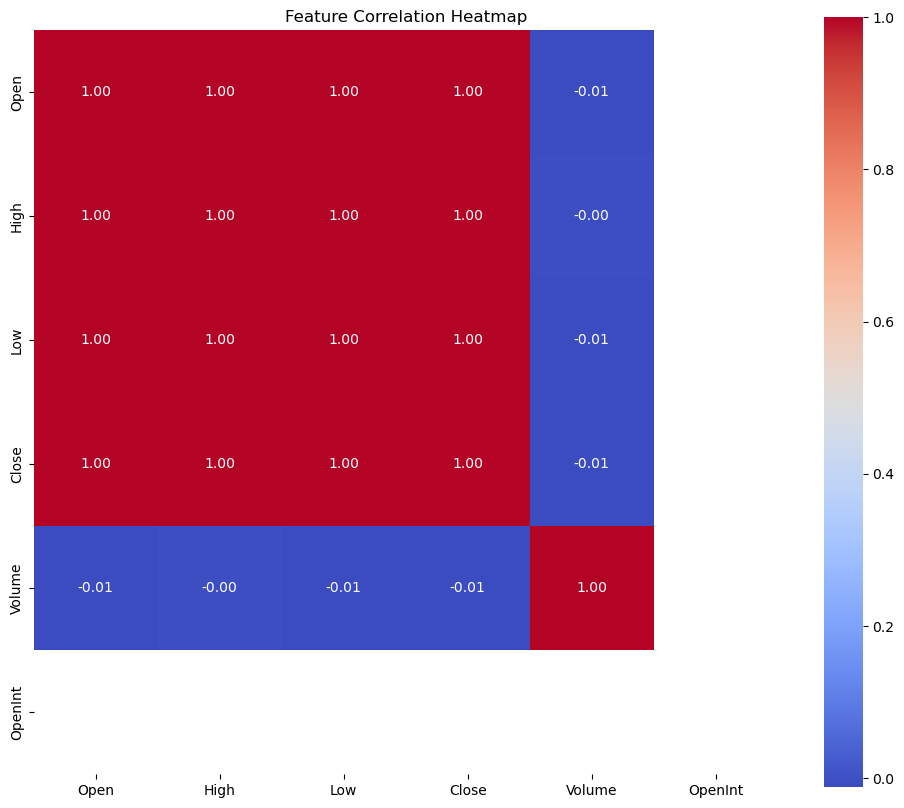
加上volume，window size = 10, step = 1

MSE on test set: 898.02

Loss: prediction:

(ii)

把Date刪去後將所有的feature做heatmap，觀察得出high和open, low, close的correlation都是1，且open, high, low, close與high的隔天數據的correlation也是1，實驗window size = 10, step = 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 刪open | 刪high | 刪low | 刪close |
| Test Mse | 2.06 | 1.65 | 1.81 | 1.63 |
|  | 剩open | 剩high | 剩low | 剩close |
| Test Mse | 1.92 | 2.28 | 2.03 | 1.96 |

透過實驗刪去某一個feature，發現變化不大，且實驗只留下一個與high的correlation為1的feature也發現mse只有小幅度下降，根據correlation推論使用open, high, low, close等特徵的差異不大。而透過實驗結果得出mse最低的最佳特徵組合為open, low, close

1. 以第一題最佳的實驗組合Window size = 10, step = 1

|  |  |  |
| --- | --- | --- |
|  | 有normalized | 無normalized |
| Test mse | 1.69 | 1.89 |
| Training loss |  |  |
| prediction |  |  |
| Training time | 38.3 s | 36.6 s |

Window size = 10, step = 7

|  |  |  |
| --- | --- | --- |
|  | 有normalized | 無normalized |
| Test mse | 55.2 | 70.98 |
| Training loss |  |  |
| prediction |  |  |
| Training time | 7.1 s | 7 s |

在step較小時，有normalized的結果mse會小幅度改善；而當step較大時，有normalized的mse會改善且training loss的收斂速度較快。

1. 在本實驗中window size比step size 大，因此會有重疊，幫助模型捕捉更多細節變化；如果window size < step size，沒重疊 導致訊息量少一點，容易失去某些資料動態

。但根據Understanding the effect of window length and overlap for assessing sEMG in dynamic fatiguing contractions: A non-linear dimensionality reduction and clustering文章的說內容，小overlap(0%、25%)會產生較小的變異性，代表資料比較穩定。高overlap(75%、90%)反而會讓資料變得較不穩定，更高的變異性，因此window size和step的大小沒有固定的關係。

1. 另一種適用於時間序列資料的資料增強方法是time warping。指在時間軸上以隨機方式拉伸或壓縮時間序列的局部區段，從而模擬不同速度下的變化特徵。這種方法可以提升模型對時序變異的適應性。根據研究，時間扭曲有效提高了神經網路在可穿戴裝置時間序列分類任務上的表現。Ref: T. T. Um, D. T. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Demircan, and D. Lee, "Data Augmentation of Wearable Sensor Data for Parkinson’s Disease Monitoring Using Convolutional Neural Networks," *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1-6, IEEE, 2017
2. (i)

Convolution-based models  
可接受任意長度的輸入，因此推論時可選擇固定或滑動視窗處理。常見做法是設定固定window size，使卷積核能有效捕捉區域性特徵，若輸入長度變動，可padding。

(ii)

Recurrent-based models  
RNN、LSTM等模型能處理長度可變動序列，因此推論時window size可依任務自由調整。若需要長期依賴資訊，可適當延長視窗；若僅需短期預測，可使用較小視窗。

(iii)

Transformer-based models  
Transformer對序列長度敏感，因自注意力機制計算量與序列長度平方成長。推論時常設定適中的window size，如固定長度切片，或使用改良結構（如sliding window Self-Attention）以處理長序列。