Supplementary material

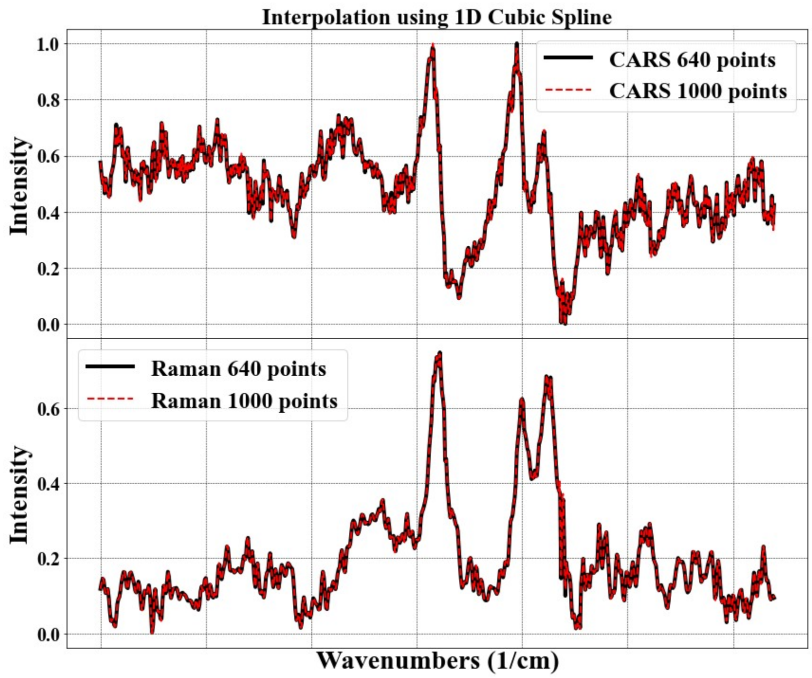

Fig. 1 Interpolation of the CARS and Raman data. The data is interpolated to meet the length requirement of the VECTOR model.

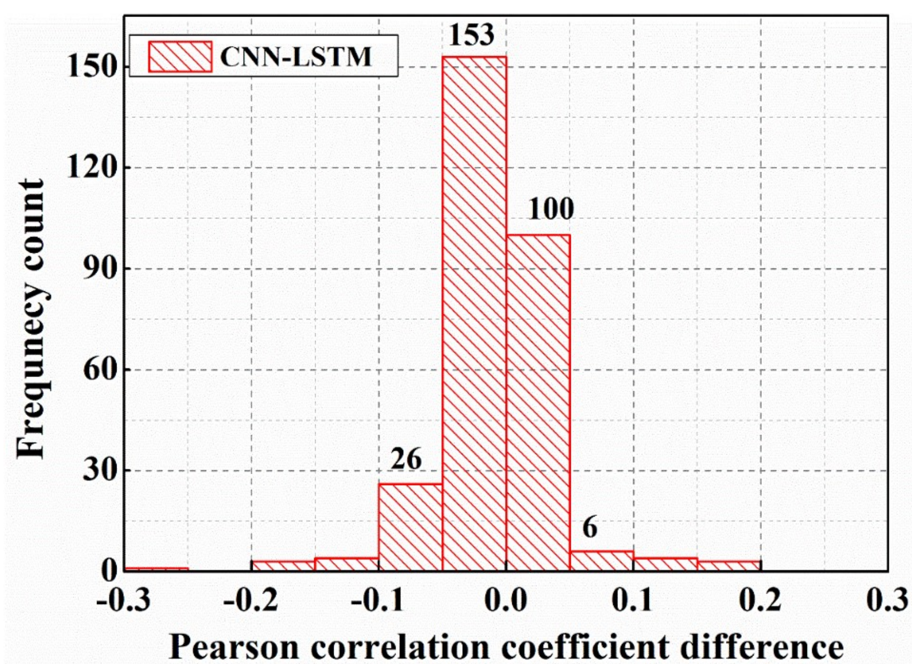

Fig. 2 Histogram plot of the PCC difference obtained from the CNN and LSTM model.

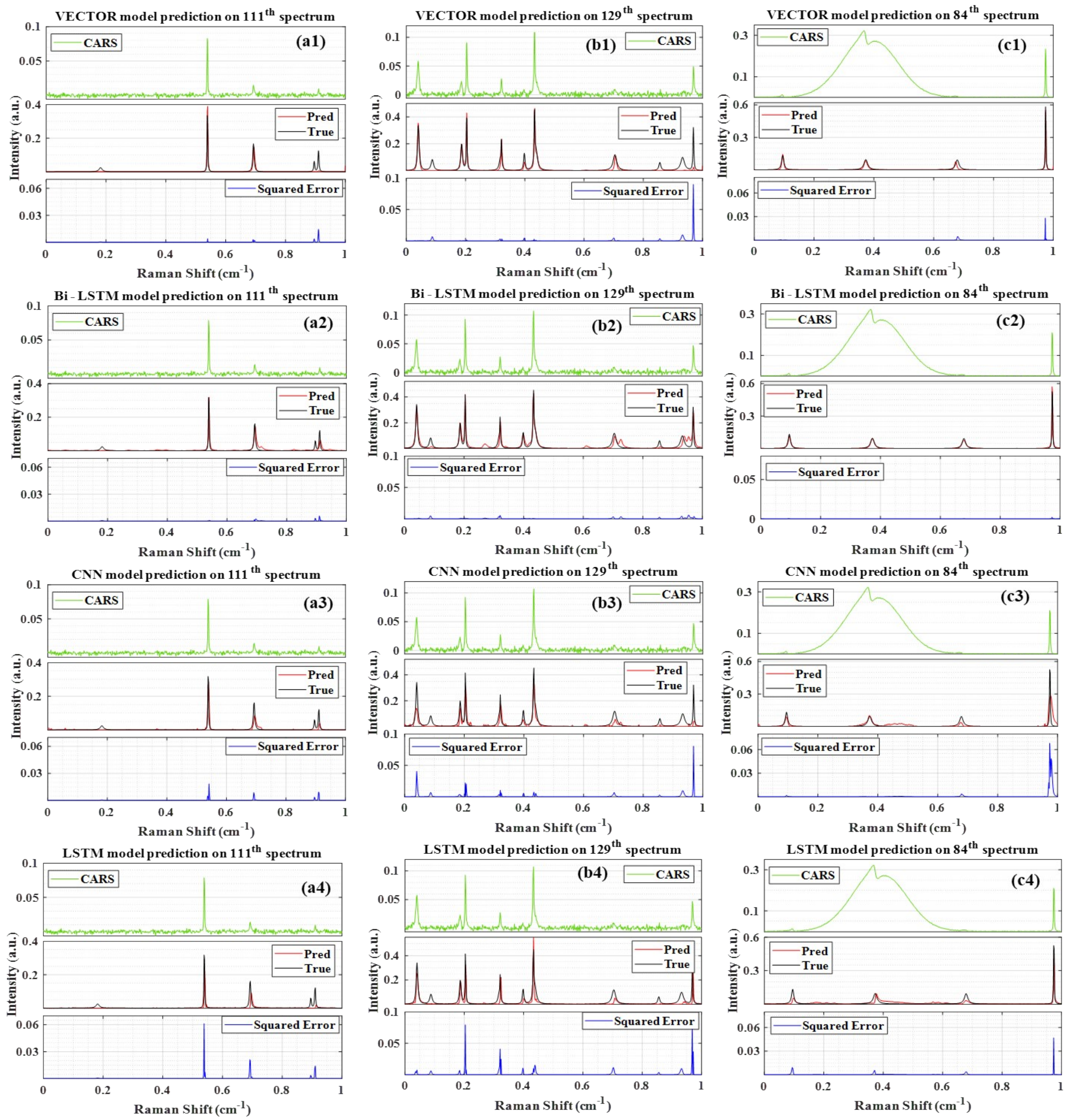

Fig. 3 Histogram plot of the PCC difference obtained from the CNN and LSTM model. Comparison of the results obtained from the four models. (a1-a2) Raman signal extracted from the 111th test spectra using VECTOR, Bi-LSTM, CNN, and LSTM models respectively, (b1-b4) Results of 129th spectrum, (c1-c4) Results of the 84th spectrum. Pred is the predicted Raman signal, and True represents the actual Raman signal. Squared error corresponds to their difference.

Computational times:

The key characteristics of the models were kept the same as in the original articles, and they are shown in Figure 1 in the manuscript. The models were trained and tested on a computer with an i5-11320H @ 3.20GHz, 12.0 GB RAM. The training times of the models differ significantly, as shown in below Table 1. Bi-LSTM and LSTM have long training times, although they contain much less parameters than the CNN and VECTOR models. In an LSTM architecture, each node is composed of 4 gates, including forget gate, input gate, output gate, and cell state, which all have their computation processes. Therefore, more calculations are needed in each node when compared with the regular ANN nodes. In a Bi-LSTM model, input data flows in both directions through a bidirectional LSTM layer, so the computing time is increased. Also, the architecture contains an order of magnitude more parameters than the standard LSTM. In the case of VECTOR architecture, the large number of parameters in the encoders and decoders affect the training time. The inference times per sample are short, and the differences may be relevant in selected applications.

Table 1. Details of the computational parameters and times. Means square error (MSE) and mean absolute error (MAE) are loss functions. Stochastic gradient descent (SDG) is an optimizer for the VECTOR model.

S. No	Model	Loss Function	Optimizer	Parameters	Training time [h]	Testing time [ms]
1	CNN	MSE	Adam	6,016,932	0.8	77.1 ± 14.4
2	LSTM	MAE	Adam	3,871	5.6	83.7 ± 18.2
3	Bi-LSTM	MSE	Nadam	51,421	29.8	238.7 ± 22.5
4	VECTOR	MAE	SGD	178,942,720	12.6	51.6 ± 12.3

Model learning curves:

The learning curves of the four deep learning models are shown in Figure 4. It should be noted that the magnitudes of the loss value in the plots are different. The training loss values are the batch loss averages within each epoch. In contrast, the validation loss values are the loss averages of the validation set using the current model. Therefore, in most cases, the validation loss is smaller than the training loss. To complement, dropout is used with the CNN and Bi-LSTM models affecting the training but not inference. The models' training processes converge differently. It is evident that with an appropriate criterion, early stopping of the training process could be adopted without affecting the model performance. This would also affect the training time, especially in the case of VECTOR, but also with the CNN and Bi-LSTM models.

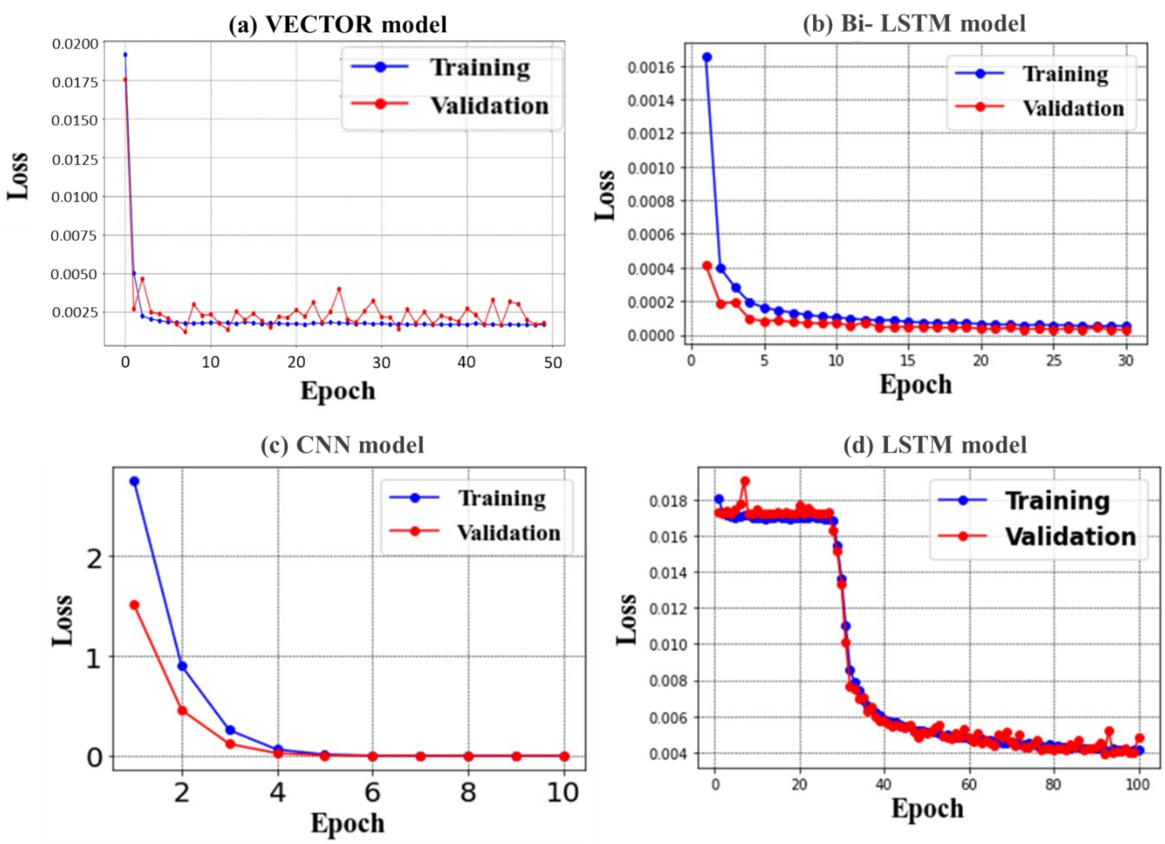

Fig. 4 (a-d) Represent the learning curves of VECTOR, Bi-LSTM, CNN, and LSTM models respectively.

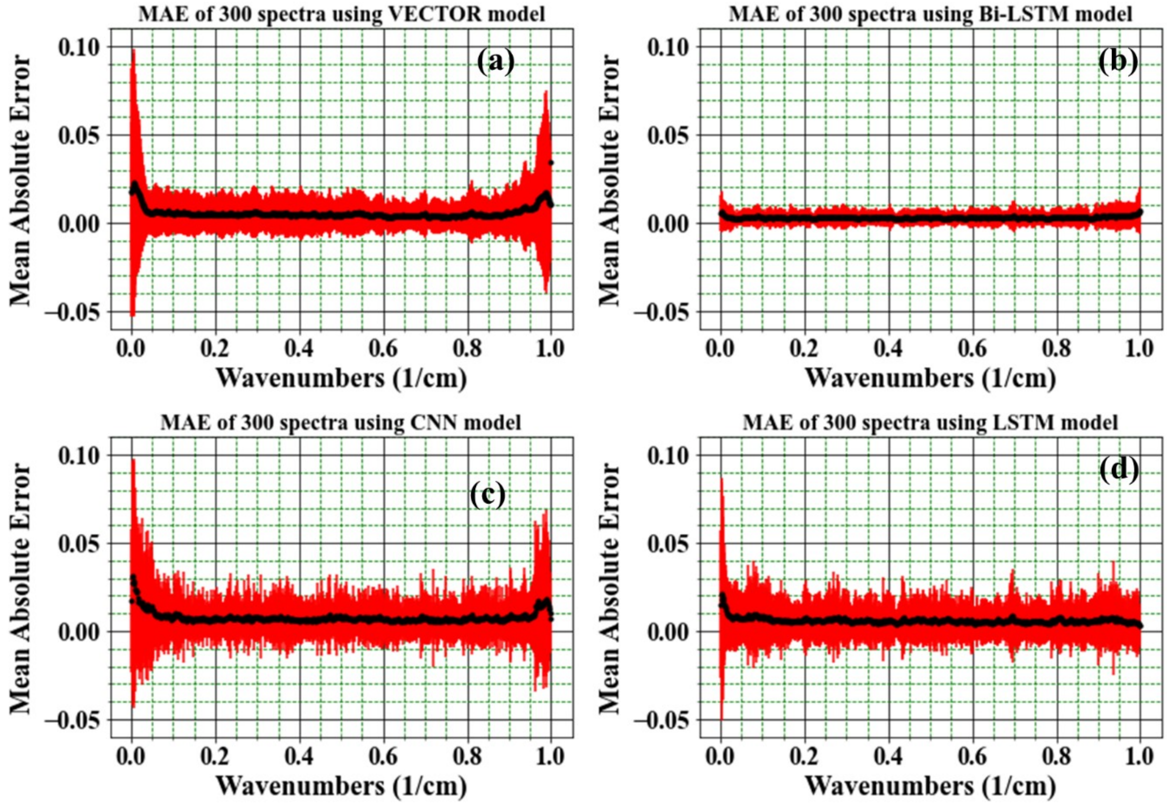

Fig 5. (a-d) Represent the mean absolute error (MAE) obtained from 300 test spectra for VECTOR, Bi-LSTM, CNN, and LSTM models respectively.