Data information integrated neural network (DINN) algorithm: interpretable machine learning by incorporating correlation information in the model architecture

Waqar Muhammad Ashraf, Vivek Dua

The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, UK

Neural Networks

Pros

- ✓ Neural networks are universal function approximator
- ✓ Neural networks can approximate the ill-defined function space with reasonable accuracy
- ✓ Computational and memory requirements are reasonable

Cons

- ✓ Neural networks are parametric models
- ✓ Neural networks are black-box models by design
- ✓ Interpretation of neural networks' predictions is an open-challenge

Data Information integrated Neural Network

- ✓ Data Information integrated Neural Network (DINN) is a modified version of standard multi-layer perceptron-based algorithm
- ✓ The loss function is customized to include the Pearson Correlation Coefficient (PCC) information to guide the parameters update in the iterative training of DINN:

$$\mathcal{L} = \left(\frac{\lambda}{1+\lambda}\right).\frac{\sum(D-Z)^2}{N} + \left(\frac{1}{1+\lambda}\right).\frac{\sum_{i=1}^{m}\left(r_{X_i|D} - r_{X_i|Z}\right)^2}{N} \tag{1}$$

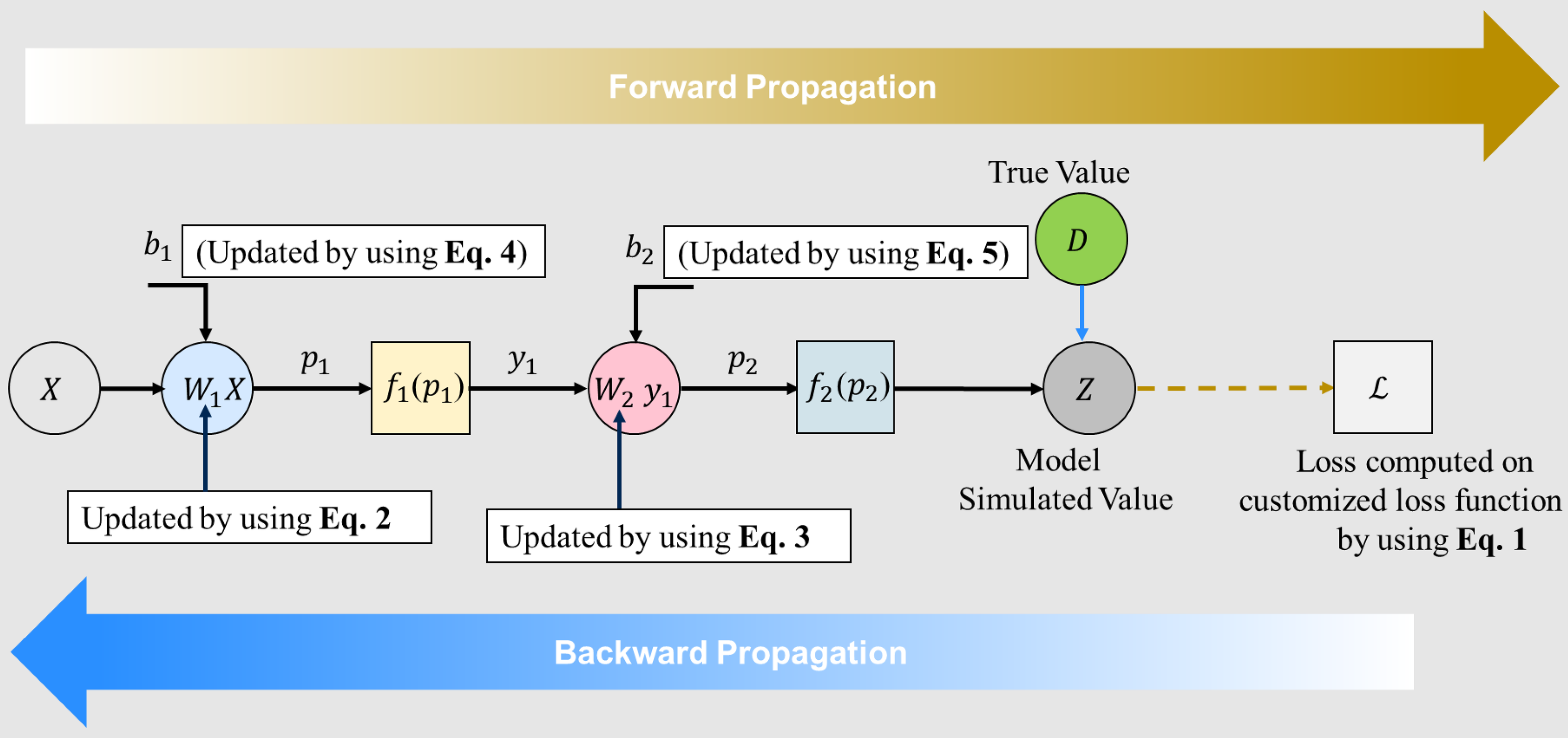
- ✓ The parameters are updated using gradient descent with momentum algorithm:

$$W_1^{new} = W_1 + \eta\left(\beta V_{W_1} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathcal{B}_i} - \frac{Z^{\mu}}{\mathcal{M}}\right)\right)W_2^T(1-y_1^2)X^T\right. \tag{2}$$

$$W_2^{new} = W_2 + \eta\left(\beta V_{W_2} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathcal{B}_i} - \frac{Z^{\mu}}{\mathcal{M}}\right)\right)y_1\right. \tag{3}$$

$$b_1^{new} = b_1 + \eta\left(\beta V_{b_1} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathcal{B}_i} - \frac{Z^{\mu}}{\mathcal{M}}\right)W_2^T(1-y_1^2)\right)\right. \tag{4}$$

$$b_2^{new} = b_2 + \eta\left(\beta V_{b_2} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathcal{B}_i} - \frac{Z^{\mu}}{\mathcal{M}}\right)\right)\right. \tag{5}$$

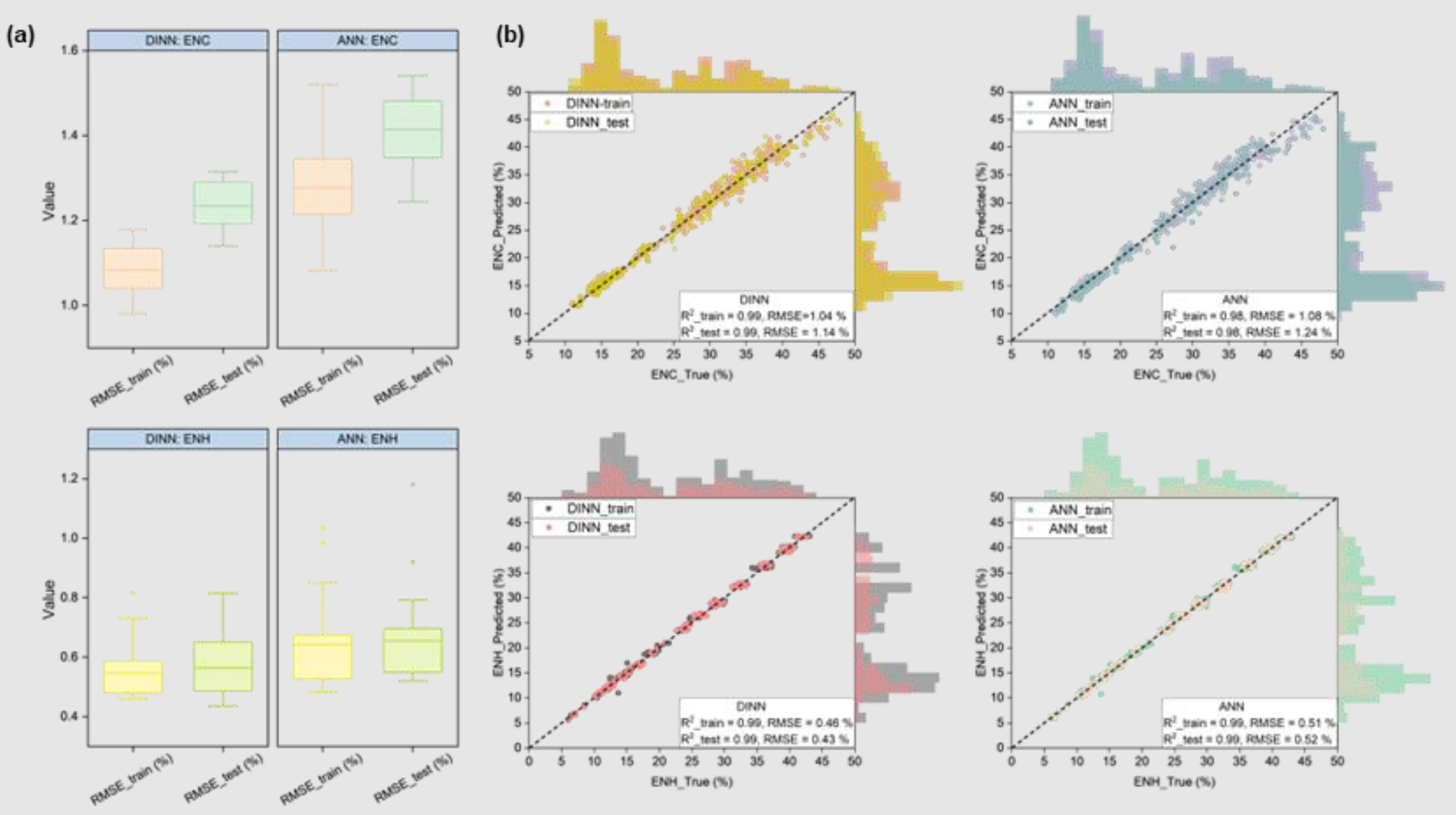


- ✓ η is the learning rate; β is momentum parameter; V_{W_1} is the velocity matrix w.r.t W_1 etc
- ✓ The stopping constraint checks the PCC computed in the model-simulated value after each epoch of training: $\min\left(|r_{X_i|D}| - |r_{X_i|Z}|\right) < \text{goal}$

Case Studies

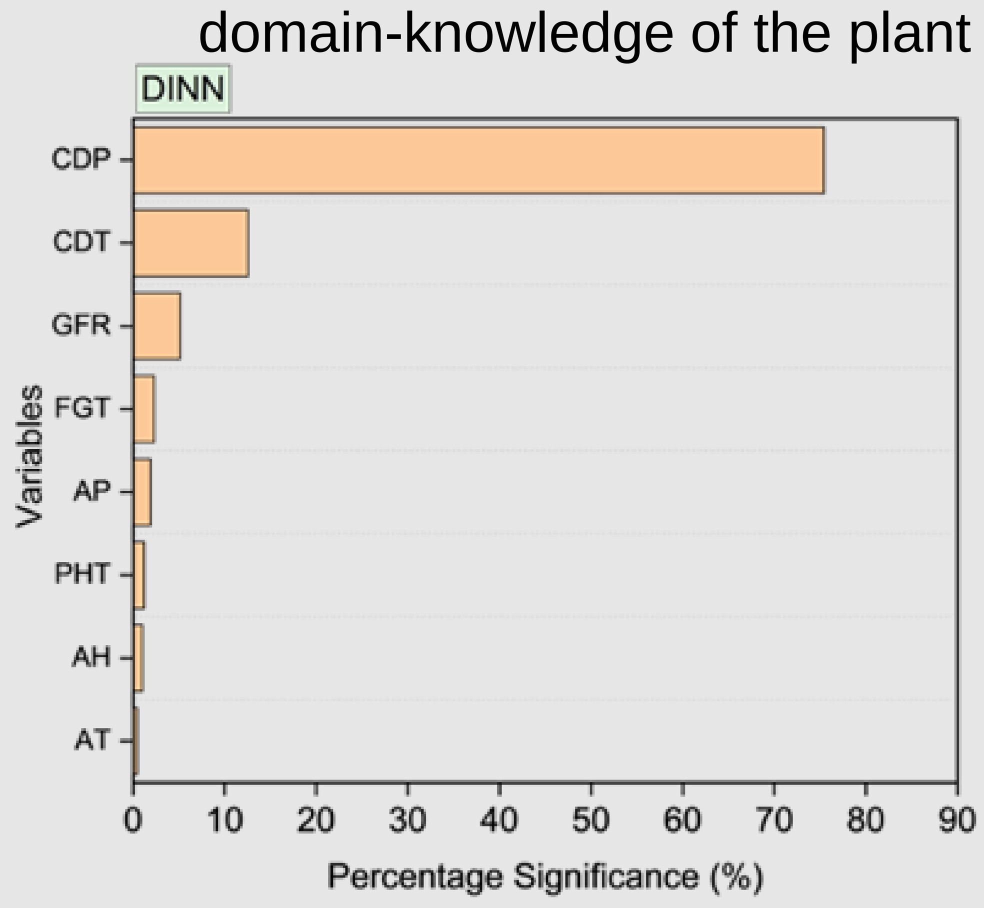
Energy Efficiency Cooling & Heating of Buildings

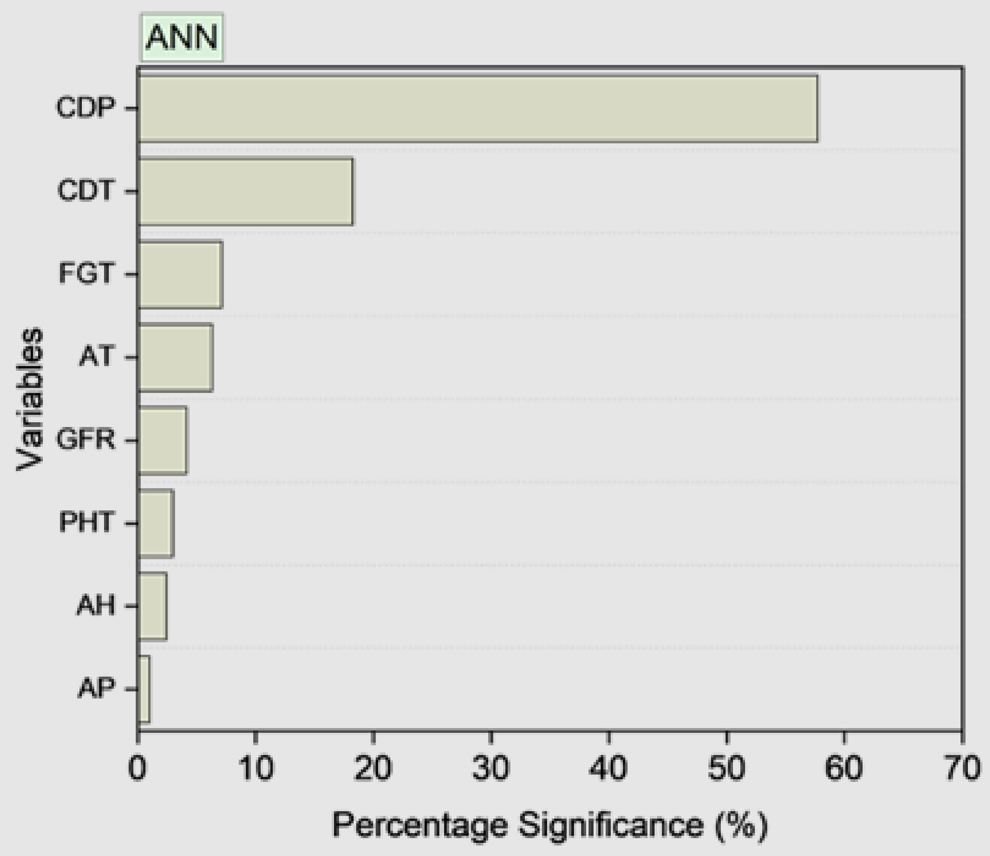
- ✓ The modelling performance of DINN is compared with those of artificial neural network (ANN) for energy efficiency cooling (ENC) and energy efficiency heating (ENH)



Power generation from an industrial 395 MW gas turbine

- ✓ The variables significance order for the power generation from the gas turbine is established for DINN and ANN model by Monte Carlo technique
- ✓ DINN based variables significance order is better compliant with the domain-knowledge of the plant operation





Conclusions

- ✓ DINN presents lower mean RMSE for testing datasets (RMSE_test = 1.23%) in comparison with the ANN model (RMSE_test = 1.41%) and literature (RMSE_test = 1.63%)
- ✓ Better predictive performance of DINN over ANN for modelling power generation from gas turbine (RMSE_DINN = 1.51 MW < RMSE_ANN = 2.75 MW)
- ✓ The PCC information improves the modelling and interpretation performance of DINN

Contact Details

Waqar Muhammad Ashraf: Waqar.ashraf.21@ucl.ac.uk

Vivek Dua: v.dua@ucl.ac.uk

Affiliation

Sargent Centre for Process Systems Engineering

Funding Source