Data Driven Prediction, Optimization, and Sensitivity Analysis of a Humidifier

Mahyar Abedi

Department of Mechanical Engineering

Motivation & Objectives

- Water and energy shortages is the most critical challenge and by 2025 around 66.7 % of world population would deal with the shortage
- Desalination technologies could act as possible solution
- Humidification-dehumidification (HDH) with packed-bed material exhibited vast potentials due to economically viable desalination, and compact configuration
- In the present work, we aim to predict the behavior of the system under quasi steady state assumption
- A data-set from experimental and well-resolved numerical simulation was used to study three regression machine learning techniques (DNN,GPR, and SVM), optimize DNN, and perform sensitivity analysis

1) Humidifier Mathematical Model

Based on the mathematical model [1], following parameters have impact on the performance:

1. Packed-bed height, diameter, and porosity
2. Temperature, and mass flow for each stream
3. Inlet pressure of the air

$$\frac{\partial T_L}{\partial t} = \frac{L}{\rho_L \alpha_L}\frac{\partial T_L}{\partial z} - \frac{\partial \omega}{\partial z}\frac{G\left(h_{fg} - h_L\right)}{\rho_L \alpha_L C_{P,L}} + \frac{U_L a_w\left(T_{pack} - T_L\right)}{\rho_L \alpha_L C_{P,L}} + \frac{U a_w (T_a - T_L)}{\rho_L \alpha_L C_{P,L}} \tag{1}$$

$$\frac{\partial T_a}{\partial t} = \frac{-G}{\rho_a \alpha_a}\frac{\partial T_a}{\partial z} - \frac{\partial \omega}{\partial z}\frac{G\left(h_{fg,T_L} - h_{v,T_a}\right)}{\rho_a \alpha_a (1+\omega) C_{P,mix}} - \frac{U_G (a - a_w)\left(T_a - T_{pack}\right)}{\rho_a \alpha_a (1+\omega) C_{P,mix}} + \frac{U a_w (T_a - T_L)}{\rho_a \alpha_a (1+\omega) C_{P,mix}} \tag{2}$$

$$\frac{\partial T_{pack}}{\partial t} = \frac{U_G (a - a_w)\left(T_a - T_{pack}\right)}{\rho_{pack}\alpha_{pack} C_{P,pack}} - \frac{U_L a_w\left(T_{pack} - T_L\right)}{\rho_L \alpha_L C_{PL}} \tag{3}$$

$$\frac{\partial \omega}{\partial z} = \frac{\partial T_a}{\partial z}\frac{P}{P - P_{sat,T_a}}\omega(b - 2cT_a + 3dT_a^2) \tag{4}$$

2) Machine Learning Techniques

- Supervised Learning Regressor is the tool that will help us to develop the model
  1. Deep Neural Network
  2. Gaussian Process Regression
  3. Support Vector Machine Regression
- Hyper-parameter optimization
  1. Bayesian Model
- Trained model based on artificial features in the forms of X_i^{-1} and $X_i X_j$
  1. Recursive Feature Elimination (For Linear Regression)
  2. Permutation Feature Importance (For DNN)
  3. Feature Importance (For Random Forest)

Contact Information

- Email: abedimah@msu.edu

3) Train/Test Data-set

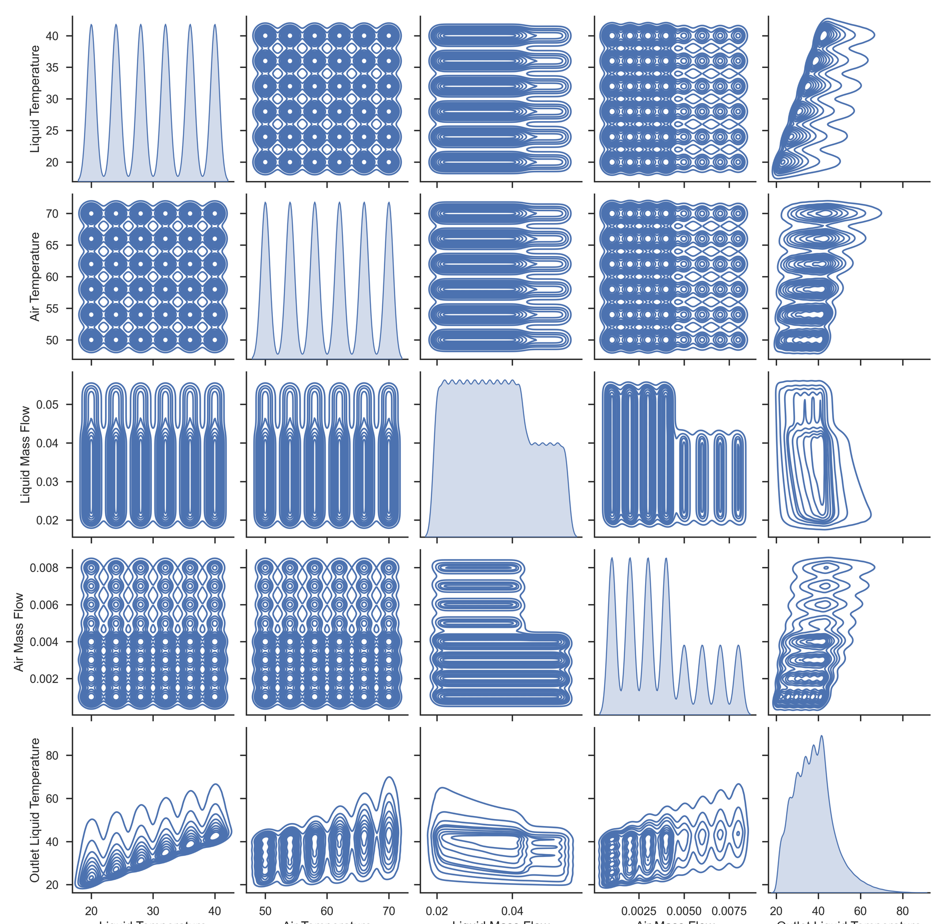
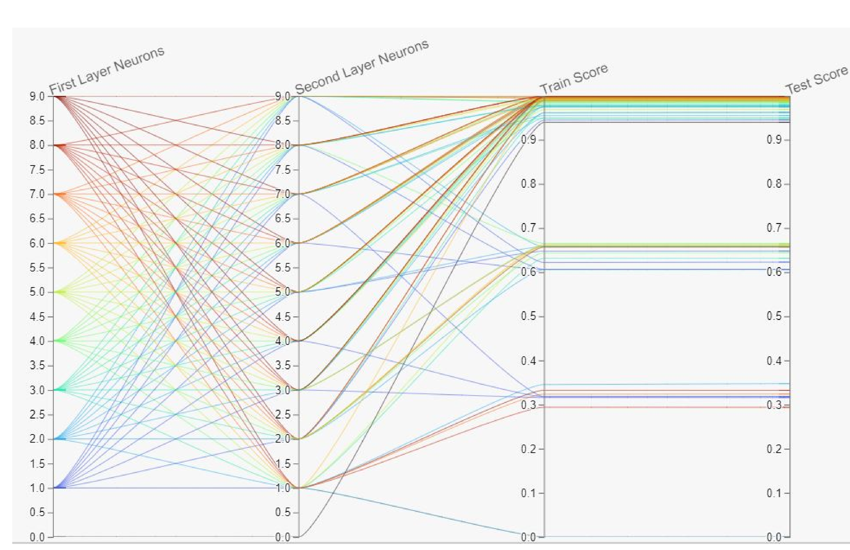
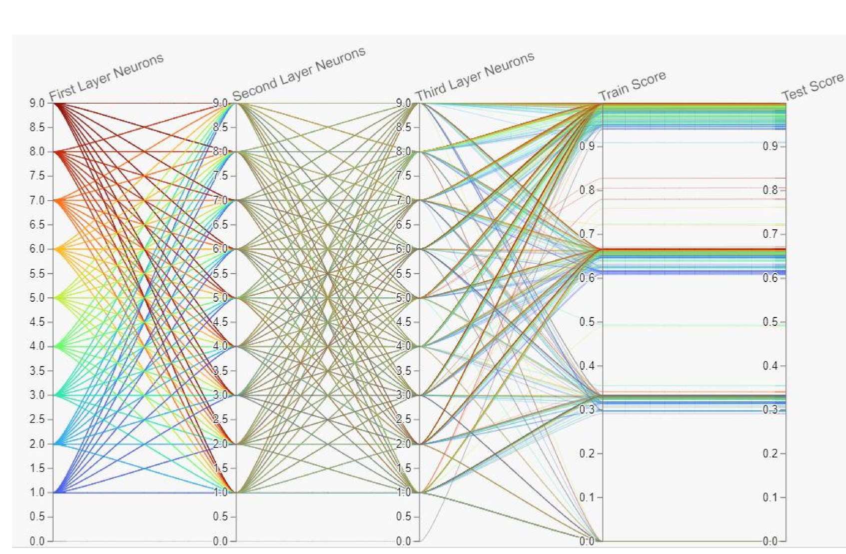


Figure 1: Pair-plots for some of the features in the data-set. All of the features were scaled prior to ML. Data-set sample number=89856

4) DNN Regression



(a) DNN with 2 layers



(b) DNN with 3 layers

Figure 2: Architecture investigation for DNN

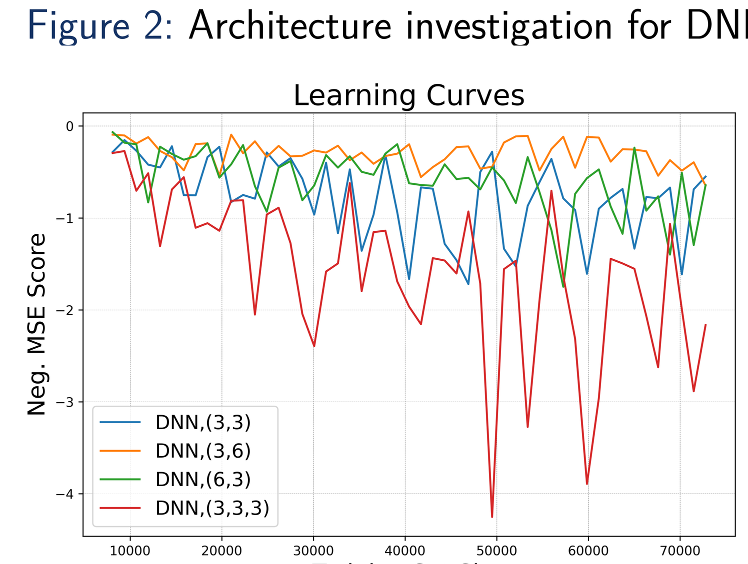


Figure 3: Learning curve for DNN with respect to training size based on mean squared error

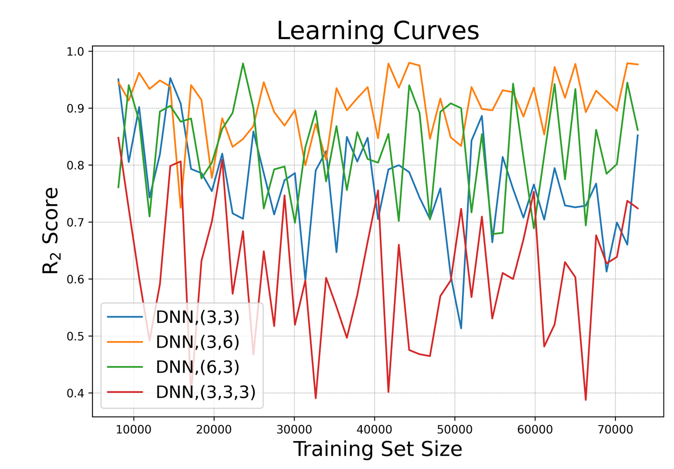
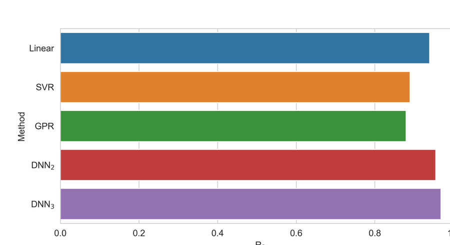
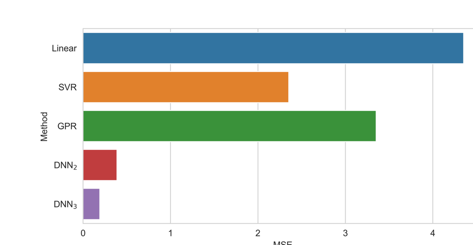


Figure 4: Learning curve for DNN with respect to training size based on R_2 error

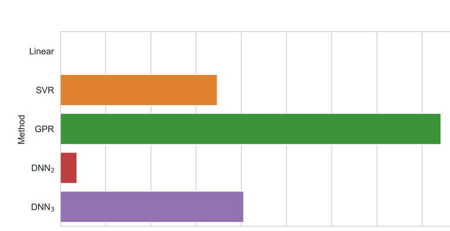
5) Comparison of ML Training Method



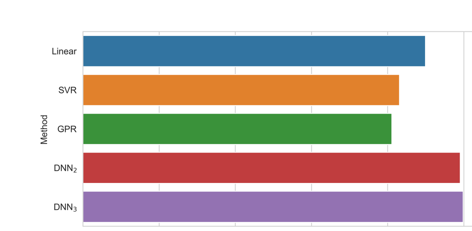
(a) R_2 error



(b) Mean squared error



(c) Execution time



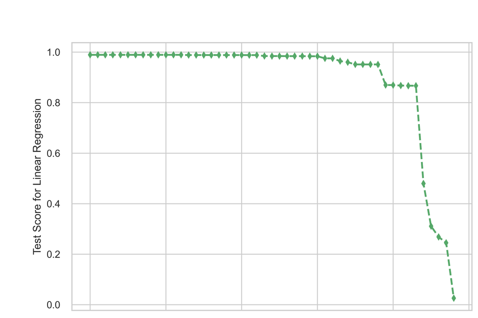
(d) Score for testing data

Figure 5: Performance comparison between different ML models

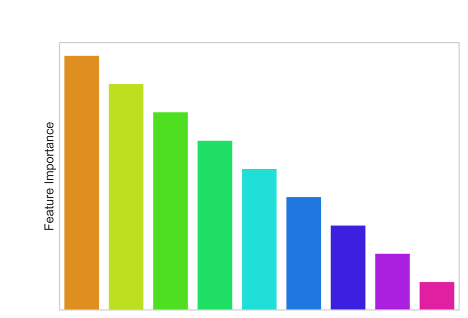
6) Sensitivity Analysis

X_1:T_L	X_2:T_A	X_3:P_A	X_4:H
X_5:D	X_6:$\dot{M}_L$	X_7:$\dot{M}_A$	X_8:ε

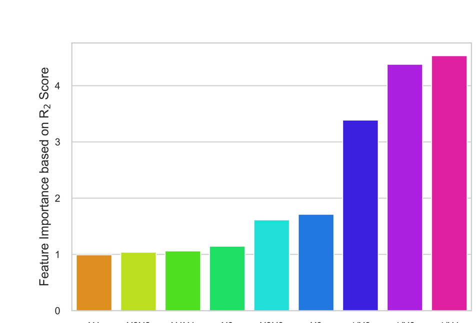
Table 1: Physical interpretation of different features



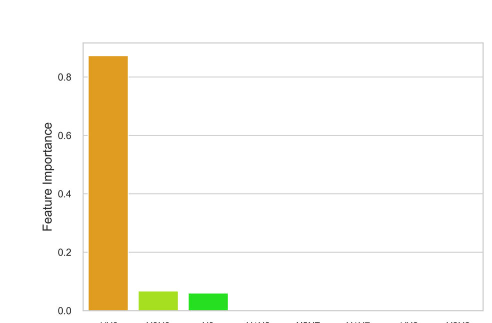
(a) Test score for linear model



(b) Linear model important features



(c) DNN important features (R_2)



(d) RF important features (R_2)

Figure 6: Features importance investigation for DNN and linear model

7) Optimization

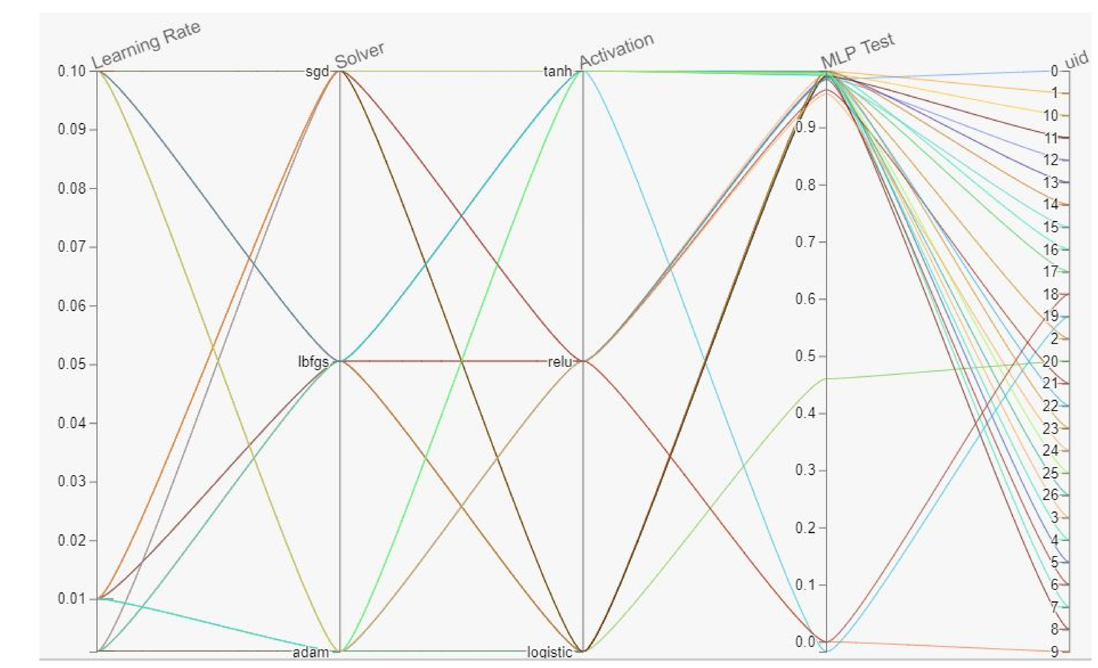


Figure 7: DNN Hyper-parameter Optimization

Iteration	Score	Parameter	Iteration	Score	Parameter
1	0.9785	0.9671	8	0.9890	0.7153
2	0.0101	0.5477	9	0.9980	0.9725
3	0.9889	0.9727	10	0.9897	0.6985
4	0.9897	0.7151	11	0.9883	0.9723
5	0.9870	0.6980	12	0.9860	0.9669
6	0.9271	0.7151	13	0.9860	0.6987
7	0.9890	0.6983	14	0.9988	0.9666

Table 2: Implementation of Bayesian Optimization for learning rate of DNN

Conclusion & Future Works

- Comparison of ML techniques showed that DNN model is an accurate (99% prediction accuracy) and required smaller amount of time (and resources)
- Since GPR creates a matrix based on features, with larger dataset run time would be much higher and demand more computation resources (for this dataset, 128 GB of RAM is needed)
- Features ranking for sensitivity analysis was unexpected because based on the numerical investigation, X_3 or pressure doesn't have considerable impact on different variables but according to linear regression and random forest, it's one of the most effective features

References

[1] F. Alnaimat, J. F. Klausner, and R. Mei, "Transient analysis of direct contact evaporation and condensation within packed beds," *International journal of heat and mass transfer*, vol. 54, no. 15-16, pp. 3381–3393, 2011.

Acknowledgements

The dataset for this poster is prepared with the support of ARPA-E (Advanced Research Projects Agency-Energy) of the department of energy under award no DE-AR0001000.