Application of Machine Learning and Deep Learning models on CFD (Computational Fluid Dynamics) Simulation Data:

Sandip Dhar

Introduction:

Various Machine Learning and Deep Learning algorithms have been applied on CFD (Computational Fluid Dynamics) simulation data for a simple 3D pipe flow problem with heat transfer. It is a simple cfd case with one inlet, outlet and wall. At inlet, we are varying the inlet velocity, temperature, turbulence intensity etc. and after solving the case we are reporting the total heat transfer rate or heat flux at outlet. For generating CFD simulation data, ANSYS-Fluent-ROM (Student version) was used under Workbench framework. In 3D ROM (Reduced Order Model), we created 100 design of experiments or design points and all of our input parameters (inlet velocity, inlet temperature, inlet turbulence intensity) were varied for a specified range and case for each design point was solved in ANSYS-Fluent and we got the value of output parameter (total heat transfer rate or heat flux at outlet). Finally we had the data of 100 design points in a table where first 3 columes were the data for input parameters (independent variables) and the last colume was the data for output parameter (dependent variables). This entire dataset was split into training set and test set where training set = 80% of data (i.e 80 lines of data) and test set = 20% of data (i.e 20 lines of data) for the applications of Machine Learning and Deep Learning models. Purpose of this study is to apply various Machine Learning and Deep Learning models on this dataset and see which model does a good job in terms of predicting the dependent variable (i.e total heat transfer rate or heat flux at outlet) for given independent variables since there is high level of non-linearity in CFD simulation data. Though ML and DL algorithms require good amount of data especially for multi-dimensional complex non-linear problems, here we are using CFD simulation data of 100 design points i.e of 100 lines for demonstration purpose.

Selection of Machine Learning and Deep Learning models:

We chose ANN (Artificial Neural Networks), SVR (Support Vector Regression) with Gaussian rbf kernel, Decision Tree, Random Forest and XGBoost model since the problem is non-linear and there are 3 independent variables. They were applied on that CFD simulation data with proper parameter selections and tuning. ANN is computationally expensive as compared to other models.

Results and Discussions:

Errors and accuracies of predictions with various models are given below:

ANN (Artificial Neural Networks):

Mean Square Error (MSE) of predicted testset results compared to original data = 0.00033578

Mean Square Error (MSE) of predicted trainingset results compared to original data= 0.00139235

SVR:

Mean Square Error (MSE) of testset results compared to original data = 1.97840604
Accuracy of trainingset results compared to original data = 97.42782875 %

Decision Tree:

Mean Square Error (MSE) of testset results compared to original data = 2.09406542
Accuracy of trainingset results compared to original data = 85.64993970 %

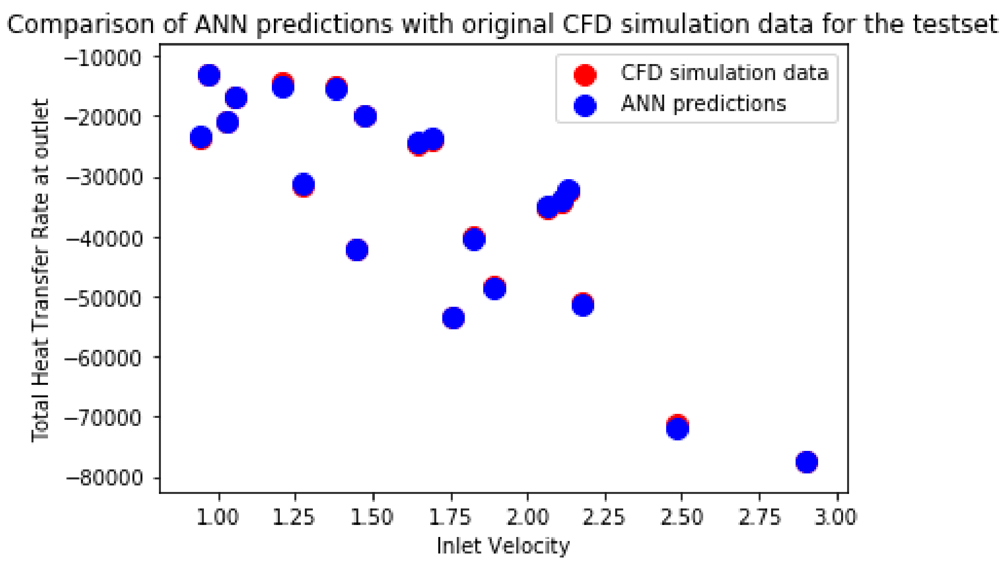
Random Forest:

Mean Square Error (MSE) of testset results compared to original data = 1.91627837
Accuracy of trainingset results compared to original data = 90.88613884 %

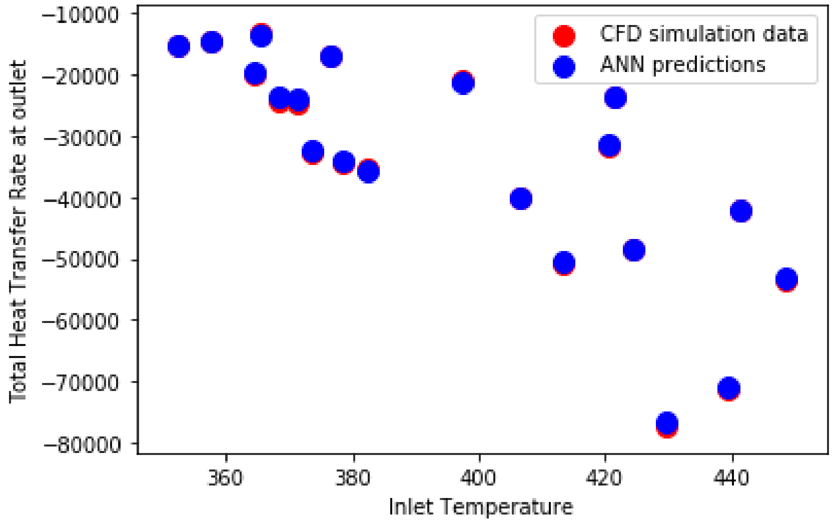
XGBoost:

Mean Square Error (MSE) of testset results compared to original data = 1.95842058
Accuracy of trainingset results compared to original data = 93.39625430 %

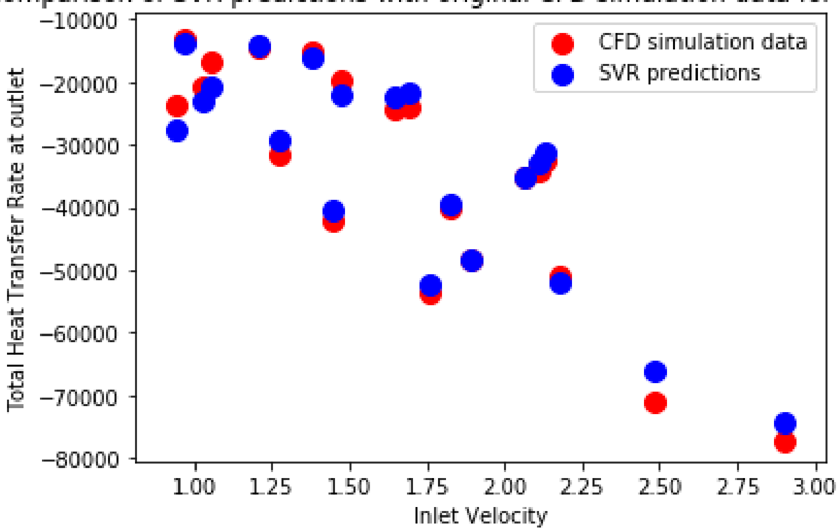
ANN outperformed other models in terms of accuracy of predictions of dependent variable (Total Heat Transfer Rate at outlet). This is quite obvious from the graphical representation of the comparison of predicted results with original CFD Simulation data for all the ML and DL models used here.



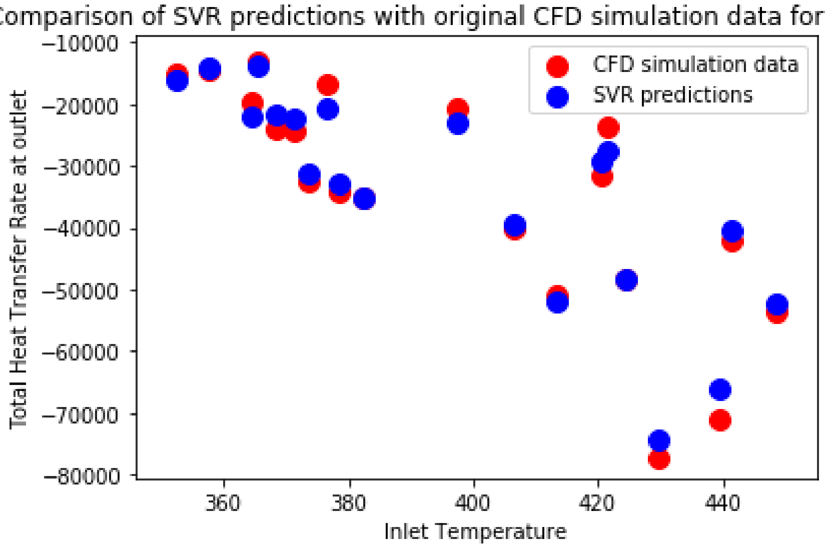
Comparison of ANN predictions with original CFD simulation data for the testset



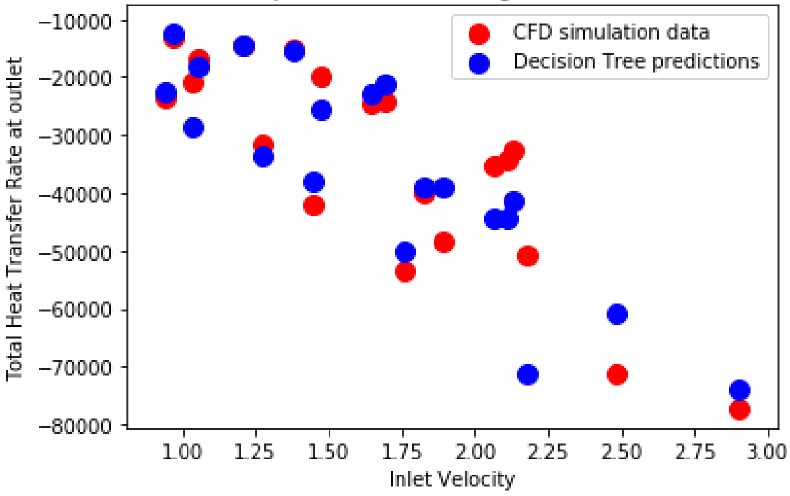
Comparison of SVR predictions with original CFD simulation data for the testset



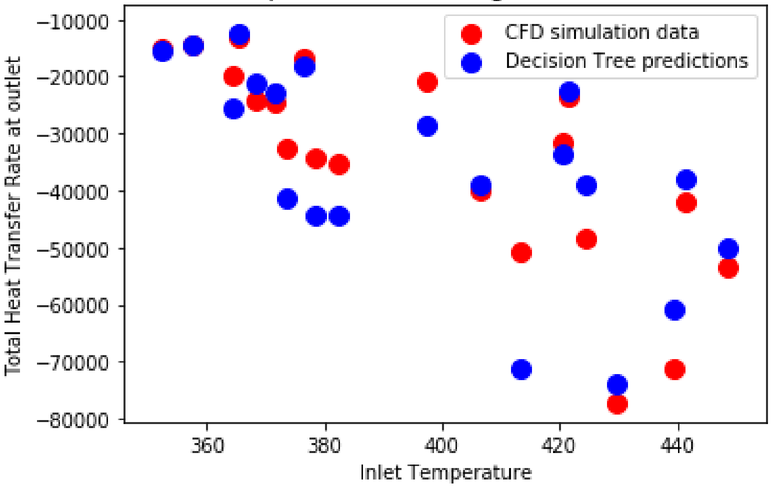
Comparison of SVR predictions with original CFD simulation data for the testset



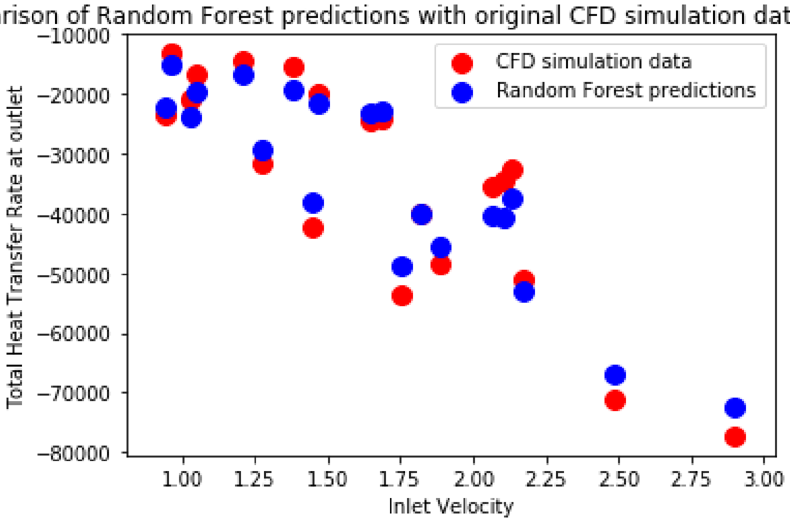
Comparison of Decision Tree predictions with original CFD simulation data for the testset



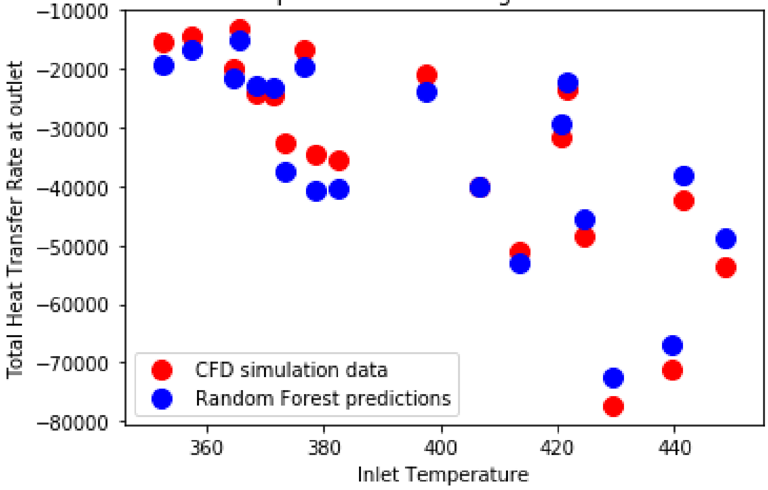
Comparison of Decision Tree predictions with original CFD simulation data for the testset



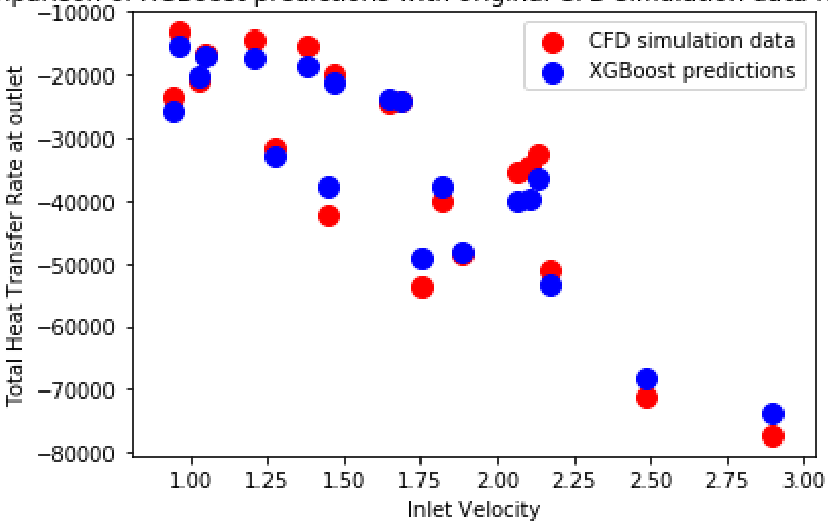
Comparison of Random Forest predictions with original CFD simulation data for the testset



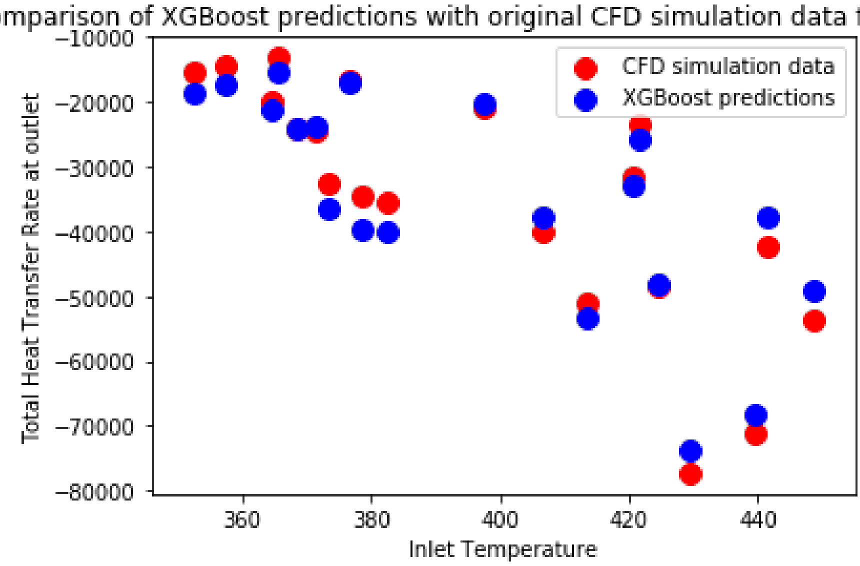
Comparison of Random Forest predictions with original CFD simulation data for the testset



Comparison of XGBoost predictions with original CFD simulation data for the testset



Comparison of XGBoost predictions with original CFD simulation data for the testset



Conclusions:

Multi-dimensional CFD simulation data even for a simple case like pipe flow are quite complex and non-linear. Artificial Neural Networks (ANN) predicted the results quite well as compared to other models though it is computationally expensive. ANN clearly outperformed other models in terms of accuracy of predictions. Though we had a dataset of just 100 design points i.e 100 lines, it would be good to repeat this study with more data (say 500 lines) for better predictions. But, generating CFD simulation data for 500 design points is quite time and memory consuming. Anyway, present study gives us some direction that ANN can be used for the predictive analysis of engineering simulation data with good accuracy.