Predictive Modeling of Engine-out Emissions using a Combination of Computational Fluid Dynamics and Machine Learning

Alok Warey [*1] **Jian Gao** [1] **Ronald O. Grover Jr.** [1]

Abstract

Analysis-driven design of Internal Combustion Engines (ICE) is extremely valuable in significantly reducing hardware investments and accelerating development of low Greenhouse Gas (GHG) emitting vehicles compliant with strict emissions regulations. Advanced physics-based engine modeling tools use system-level models coupled with Computational Fluid Dynamics (CFD) simulations to predict engine-out emissions. The success of this methodology largely relies on the accuracy of analytical predictions, especially engine-out emissions. Results show excellent agreement in prediction of engine performance parameters, oxides of Nitrogen (NOx) emissions and combustion noise, while the Carbon Monoxide (CO), Unburned Hydrocarbons (HC) and Smoke emissions predictions remain a challenge even with large chemical kinetics solvers and refined mesh resolution. In this study, a hybrid approach combining CFD analysis with Machine Learning (ML) for prediction of engine-out emissions of CO, HC and Smoke is demonstrated. Input features generated by physics-based CFD simulations and experimentally measured emissions data as labels or targets were used to train a deep Convolutional Neural Network (CNN) model. This approach led to a significant improvement in prediction accuracy of all three emissions species and captured the qualitative trends as well. The ML model could be used to augment the engine modeling toolkit leading to significantly more accurate predictions of engine-out emissions, lower computational costs and reduced turnaround times for engine simulations.

[1]General Motors Global Research and Development, Warren, MI, United States. Correspondence to: Alok Warey <alok.warey@gm.com>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

1. Introduction

Computational Fluid Dynamics (CFD) tools have emerged as an effective means to evaluate numerous designs and ideas prior to hardware build. These tools are increasingly able to resolve complex physics at progressively smaller temporal and spatial scales over multiple engine cycles (Gao et al., 2018a). The ultimate goal is to develop a virtual engine that can capture the behavior of an internal combustion engine (Gao et al., 2018b). The virtual engine model relies on a full three-dimensional (3D) physics-based Computational Fluid Dynamics (CFD) model for engine combustion and emission predictions. The parameters that control engine performance and emissions are used as design factors to generate a Design of Experiment (DoE) matrix. Then, a constrained optimization of critical engine parameters is carried out to minimize fuel consumption, while meeting engine-out emissions requirements. The success of the methodology largely relies on the accuracy of analytical predictions, especially engine-out emissions. However, the effectiveness of CFD simulation tools for in-cylinder engine combustion is often compromised by the complexity, accuracy, and computational overhead of detailed chemical kinetics necessary for combustion calculations. Results show excellent agreement in prediction of engine performance parameters, oxides of Nitrogen (NOx) emissions and combustion noise, while the Carbon Monoxide (CO), Unburned Hydrocarbons (HC) and Smoke emissions predictions remain a challenge even with large chemistry solvers and refined mesh resolution (Gao et al., 2018a;b). The current study differs from previous approaches in that a Convolutional Neural Network (CNN) model was used to predict engine-out emissions of CO, HC and Smoke from in-cylinder contours of multiple scalar fields generated by physics-based CFD simulations. Experimentally measured emissions data were used as labels or targets for training the the CNN model. The overall ML model development workflow is shown in Figure 1.

2. Training and Test Data

Physics-based CFD calculations were performed with a commercial software package, CONVERGE™ 2.3 (CONVERGE, 2016). Images of in-cylinder features were generated over 600 points spanning in-vehicle operating condi-

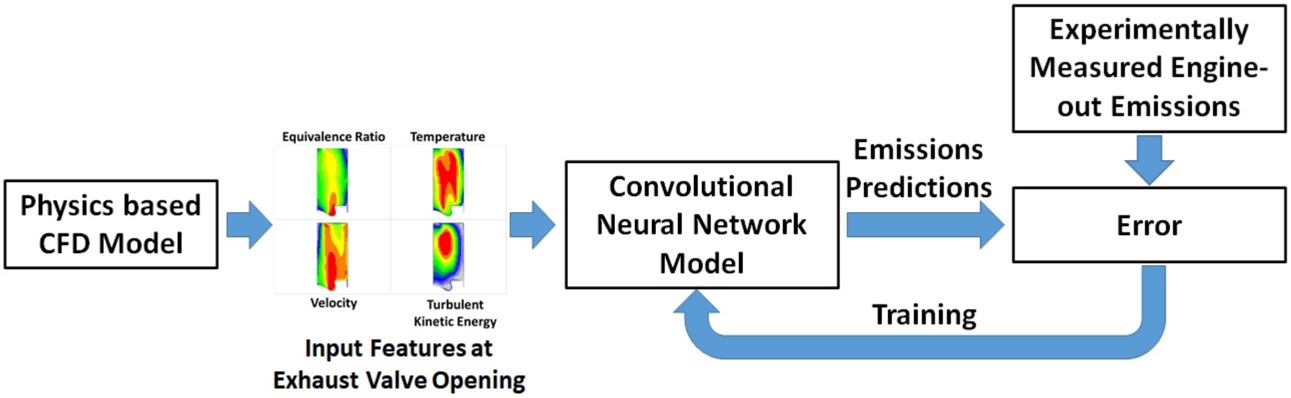


Figure 1. Overall machine learning model development workflow.

tions. Multi-cylinder engine experiments were conducted with a GM 1.6L, 160 HP (119.3 kW) diesel engine for passenger car application. A dataset of 600 samples may seem small, but is quite large for engineering applications due to the expense of experiments and physics-based computations.

Contour plots of multiple scalar fields were processed for each of the 600 cases to create training and test datasets for the CNN model described in Figure 1. A sample processed image is shown in Figure 2. Each processed image consisted of four viewports displaying a single scalar on a vertical cutplane through the center of the cylinder geometry. Only one image was processed per case just prior to exhaust valve opening (EVO). The four scalar fields chosen were equivalence ratio (ratio of the actual fuel/air ratio to the stoichiometric fuel/air ratio), in-cylinder temperature, mean flow velocity magnitude and turbulence kinetic energy (TKE) computed from the Re-Normalization Group Methods (RNG) k-ϵ model. Equivalence ratio and temperature were included based on their well-known correlation to emissions (Akihama et al., 2001), while mean velocity magnitude and TKE were considered to account for flow motion contributing to additional species oxidation effects occurring after EVO that were not directly simulated by the CFD model. A banded color scale was used to clearly demarcate separation between contour levels. Each case used the same absolute scale (i.e., contour values were the same) with red being a 'high' value and gray being a 'low' value. The processing routine was wrapped within a scripted framework to generate a complete set of 1292 pixel x 830 pixel resolution images.

3. Machine Learning Model

The dataset of 600 images, one for each engine operating condition, was randomly split into training and test datasets as follows.

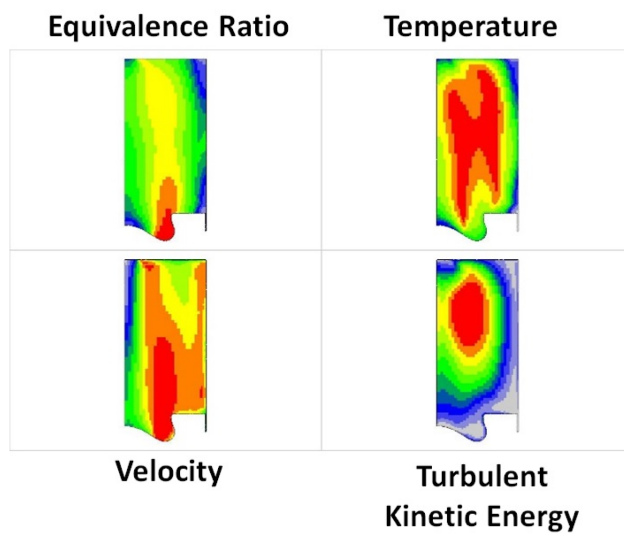


Figure 2. Input image to the CNN model

- Training Set: 500 images.
- Test Set: 100 images.

In this work, the open source deep learning library TensorFlow (Abadi et al., 2016; Chollet, 2015) was used for developing, training and testing the CNN models. Due to the small dataset, instead of training a custom convolutional neural network, a CNN model was built using the VGG16 architecture (Simonyan & Zisser, 2014) trained on the ImageNet dataset (Russakovsky et al., 2015). The baseline VGG16 architecture was modified for this dataset by retaining the convolutional base, removing the densely connected classifier and adding a fully connected regressor as shown in Figure 3. The learned weights of the VGG16 convolutional base on the ImageNet dataset were retained in this study by freezing the convolutional base during training and hy-

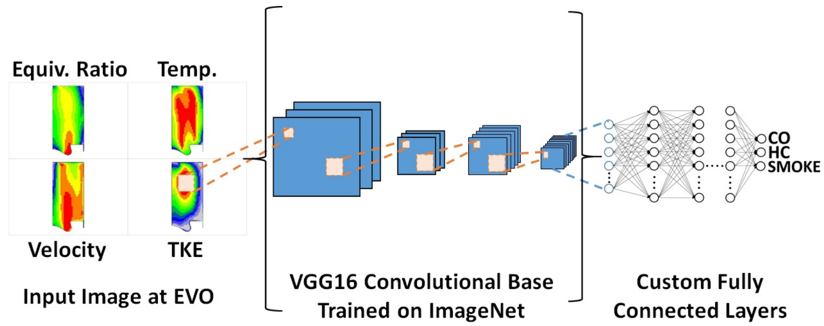


Figure 3. Modified VGG16 model with CFD generated in-cylinder images as input and a fully connected regressor.

perparameter optimization of the fully connected regressor. Instead of using a single model, various ensembles of the top models found during hyperparameter optimization of the regressor were investigated . Ensembles of the top three and five models were investigated using repeated 5-fold cross validation on the training dataset (500 images). A population of model performance scores (MAE on the validation folds) was generated from the repeated cross-validation procedure. Overall the ensembles had better prediction performance, MAE on the validation folds, compared to the single CNN model as shown in Figure 4. Adding more than five CNN models to the ensemble didn't offer any improvement in prediction performance. Hence, the final ensemble was configured to use a uniform average of predictions from five CNN models.

4. Results

The trained CFD + CNN ensemble model was evaluated on the test set (or held-out data) of 100 images. Prediction performance of the CFD + CNN ensemble model and the state-of-the-art CFD model for each emissions species (CO, HC, Smoke) on the test set is given in Table 1. Figure 5 shows the comparison between actual (experimentally measured emissions data) and predicted values of all three emissions species (CO, HC and Smoke) by both the CFD and CFD + CNN ensemble models for the test set. The CFD + CNN ensemble model outperformed the state-of-the-art CFD model in predicting all three emissions species. Significant improvement in prediction accuracy ($R^2 > 0.82$) was observed compared to the state-of-the-art CFD model predictions, especially, for unburnt HC and CO emissions. The CFD + CNN framework only requires a low-fidelity chemistry model to compute the inputs to the CNN ensemble model (scalar fields of equivalence ratio, in-cylinder temperature, mean flow velocity magnitude and turbulence kinetic energy). These computations typically require less than 4 hours to complete for a single operating condition. Alternatively, a high-fidelity combustion model consisting of approximately 1000 species to track emissions requires 50-150 hours to complete. Hence, the CFD + ML approach is believed to have at least a 10X cost benefit compared to a state-of-the-art CFD model, in addition to significantly more accurate emissions predictions.

Table 1. Comparison of the prediction performance of the CFD + CNN ensemble and CFD models on the test set.

CFD + CNN ENSEMBLE MODEL			
	CO	HC	Smoke
MAE [g/h; FSN]	53.5	2.6	0.4
RMSE [g/h; FSN]	114.3	4.9	0.6
CFD MODEL			
	CO	HC	Smoke
MAE [g/h; FSN]	403.0	12.0	1.1
RMSE [g/h; FSN]	712.9	26.3	1.3

Score-Weighted Class Activation Mapping (Score-CAM) (Wang et al., 2019) was used to produce activation heatmaps. Activation heatmaps for two sample images from the test set, when predicting HC emissions, are shown in Figure 6. The activation heatmaps confirm the physical dependence of engine-out emissions on the four scalar fields used in this study.

5. Summary

The ability to accurately predict engine-out emissions of Carbon Monoxide (CO), Unburned Hydrocarbons (HC) and Smoke is critical in accelerating development of high-efficiency engines compliant with emission regulations. This study demonstrates that a combination of physics-based and machine learning models can be used to accurately predict engine-out emissions. The methodology outlined in this paper can be applied to both gasoline and diesel fueled engines. The hybrid approach presented in this study could lead to significantly more accurate predictions of engine-out emissions, lower computational costs and reduced turnaround times for engine simulations.

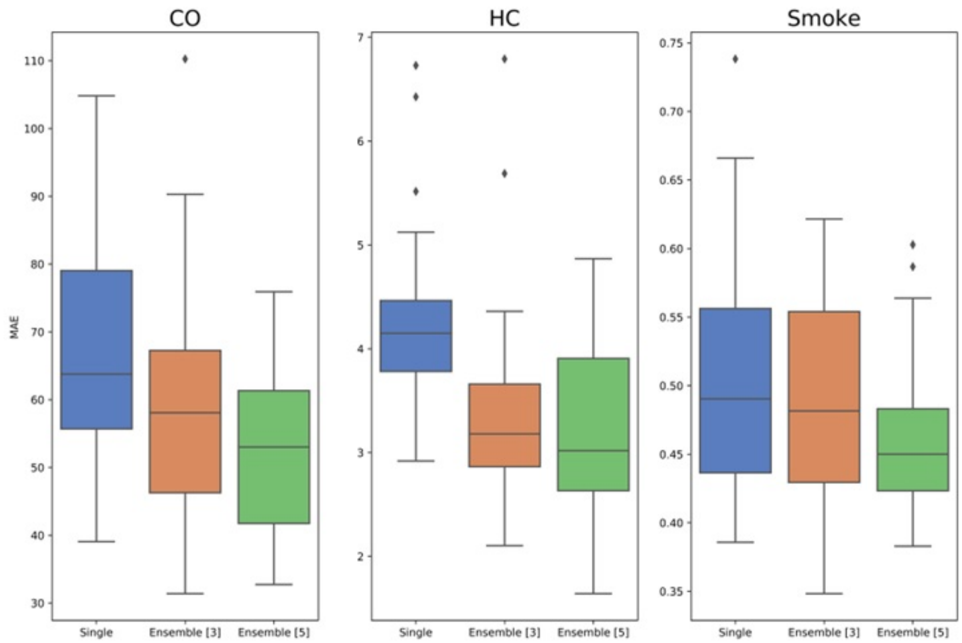


Figure 4. Performance of the ensemble models vs. a single optimized model with repeated cross-validation on the training dataset (500 images).

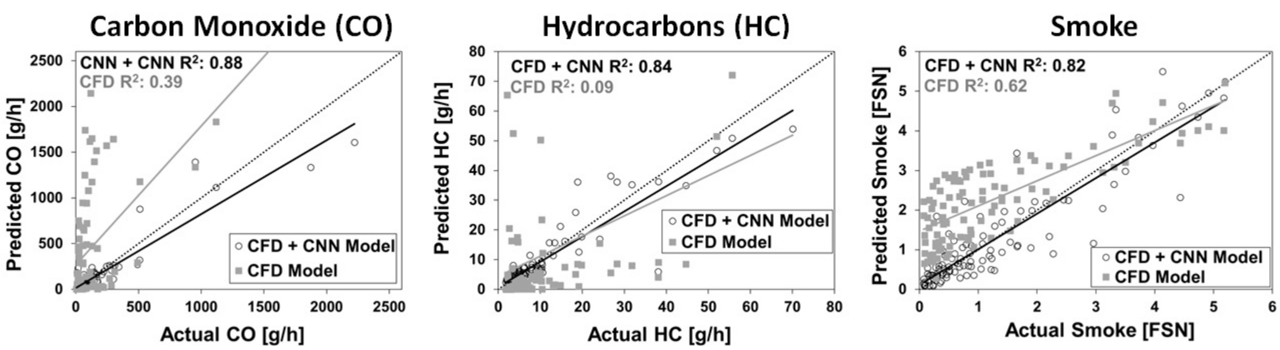
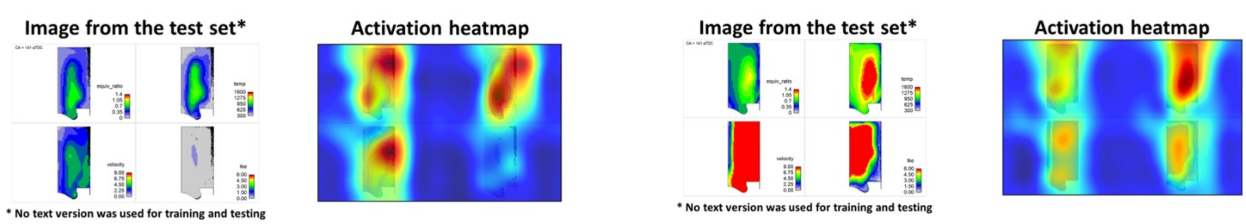


Figure 5. Comparison between actual (experiment) and predicted values of CO, HC and Smoke emissions by both the CFD + CNN ensemble and CFD models on the test set of 100 images.



Predictions for this Image

Actual HC [g/h]	CFD + CNN Prediction [g/h]	CFD Prediction [g/h]
2.31	3.7	3.9

Predictions for this Image

Actual HC [g/h]	CFD + CNN Prediction [g/h]	CFD Prediction [g/h]
4.83	5.24	0.099

Figure 6. Activation heatmaps for two sample images from the test set when predicting HC emissions.

Impact Statement

The transport of goods and people accounts for about 20% of the total global primary energy consumed, around 23% of CO_2 emissions (Kalghatgi, 2018). Currently, transport is almost entirely ($> 99.9\%$) powered by internal combustion engines (ICE). Even with significant expansion of electrification, a significant portion of transport will still be powered by internal combustion engines for many years to come, either deriving all or some portion of the energy from the engine (hybrid powertrains) (Kalghatgi, 2018). Therefore, it is essential to continue improving the fuel efficiency of the internal combustion engine, while complying with the strict emissions regulations. Physics-based models perform well in prediction of engine performance parameters, oxides of Nitrogen (NOx) emissions and combustion noise, while Carbon Monoxide (CO), Unburned Hydrocarbons (HC) and Smoke emissions predictions remain a challenge. The hybrid approach presented in this study could lead to significantly more accurate predictions of engine-out emissions, lower computational costs and reduced turnaround times for engine simulations (Warey et al., 2021a;b).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Tensorflow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, 2016.

Akihama, K., Takatori, Y., Inagaki, K., Sasaki, S., and Dean, A. Mechanism of the smokeless rich diesel combustion by reducing temperature. *SAE Technical Paper 2001-01-0655*, 2001.

Chollet, F. *Keras, GitHub repository*. 2015.

CONVERGE. *CONVERGE CFD 2.3 Manual*. 2016.

Gao, J., Gopalakrishnan, V., Diwakar, R., Elwasif, W., Edwards, K., Finney, C. E., and Whitesides, R. Steady-state calibration of a diesel engine in computational fluid dynamics using a graphical processing unit-based chemistry solver. *Journal of Engineering for Gas Turbines and Power*, 140(10):102802–1–5, 2018a.

Gao, J., Gopalakrishnan, V., Diwakar, R., Warey, A., Rodio, A., and Ardanese, R. Virtual engine model for analysis-driven pre-calibration. *THIESEL 2018 Conference on Thermo- and Fluid Dynamic Processes in Direct Injection Engines*, 2018b.

Kalghatgi, G. Is it really the end of internal combustion engines and petroleum in transport? *Applied Energy*, 225: 965–974, 2018.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

Simonyan, K. and Zisser, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint, arXiv:1409.1556*, 2014.

Wang, H., Wang, Z., Du, M., F. Yang, Z. Z., Ding, S., Mardziel, P., and Hu, X. Score-cam: Score-weighted visual explanations for convolutional neural networks. *arXiv preprint, arXiv:1910.01279*, 2019.

Warey, A., Gao, J., and Grover, R. Prediction of engine-out emissions using deep convolutional neural networks. *SAE International Journal of Advances and Current Practices in Mobility*, 3(6):2863–2871, 2021a.

Warey, A., Gao, J., and Grover, R. Generalization performance of a deep learning based engine-out emissions model. *Energy and AI*, 5:100080, 2021b.