Artificial Intelligence Application in Combustion Modeling

Luke Dillard[1]
Purdue University, West Lafayette, Indiana, 47907, United States of America

Vishal Nandigana[2]
Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India

Jay P. Gore[3]
Purdue University, West Lafayette, Indiana, 47907, United States of America

Abstract: Artificial intelligence (AI) and more specific subsets of AI such as machine learning (ML) and deep learning (DL) have become widely available in recent years. Open source software packages and languages have made it possible to implement complex AI based data analysis and modeling techniques on a wide range of applications. The application of these techniques can expedite existing models or reduce the amount of physical testing required. Two data sets were utilized to examine the effectiveness of multiple ML techniques to estimate experimental outcomes and to serve as a substitute for additional testing. To achieve this complex multi-variant regressions and neural networks were utilized to create estimating models. The first data sets of interest consist of a pool fire experiment that measured the flame spread rate as a function of initial fuel temperature for 8 different fuels, including Jet-A, JP-5, JP-8, HEFA-50, and FT-PK. The second data set consists of hot surface ignition data for 9 fuels including 4 alternative piston engine fuels for which properties were not available. When properties were not available multiple imputation by chained equations (MICE) was utilized to estimate fluid properties. 10 different ML techniques were implemented to analyze the data and R-squared values as high as 92% were achieved.

Keywords: Artificial Intelligence, Machine Learning, Deep Neural Network, Combustion

I. Nomenclature and Abbreviations

AI	= artificial intelligence
ML	= machine learning
CFD	= computational fluid dynamics
ANN	= artificial neural network
GA	= genetic algorithm
QSPR	= quantitative structure−property relationship
MLR	= multiple linear regression
SVM	= support vector machine
AIT	= autoignition temperature
RF	= random forests
DNN	= deep neural networks
CART	= classifier and regression tree
CNN	= convolutional neural networks
DADN	= deep automotive diagnostic network
LSTM	= long short-term memory

1 Graduate Student Research Assistant, Mechanical Engineering, AIAA Student
2 Assistant Professor, Mechanical Engineering
3 Reilly University Chair Professor of Mechanical Engineering, Mechanical Engineering, AIAA Fellow

LSSVM = least squares support vector machine
SVR = support vector regression
GRP = gaussian process regression
FT-IPK = fischer-Tropsch IPK
HEFA = hydro-processed ester fatty acids
SIP = synthetic iso-paraffin
MICE = multiple imputation by chained equations
RBF = radial basis function
MPR = multivariant polynomial regression
BRR = Bayesian Ridge Regression
KNN = K-Nearest Neighbor
MLP = Multi-Layer Perceptron
DANN = Distributed-Artificial-Neural-Network

II. Introduction

Artificial intelligence (AI) is an emerging and powerful data analysis tool that can be utilized in combustion modeling. Current combustion modeling methods focus on the use of either computational fluid dynamics (CFD) and surrogate based kinetics rate models. CFD relies heavily on access to large computing power. The cost is exacerbated in combustion CFD because of the range of time scales (~0.001 to ~ 1 second) and the range of length scales (1 nm to 1 m) of interest. The thousands of chemical species must be identified, and their binary diffusion processes must be computed in addition to their binary and other reaction steps. Furthermore, the boundary conditions needed to define a problem for use in CFD may not be readily available or accurately represent the real world. These lacunae are addressed in many practical applications by empirically informing the surrogate models. The model is constructed using multi species kinetic theory and then incorporated into multi-physics combustion governing equations and solved by CFD methods. The CFD results with empirical models are validated using laboratory combustion experiments. The testing time can have both a high monetary and time cost.

The application of AI can help reduce these costs greatly by utilizing existing data from the literature to fill gaps that would previously require either experimental testing or CFD analysis to fill. Kalogirou et al. in 2003 provided a summary of possible applications of artificial intelligence including boilers, furnaces, and internal combustion engines. Artificial neural networks (ANN), genetic algorithms (GA), fuzzy logic, and hybrid systems of these methods were presented as possible methods for use with combustion systems [1]. Since then, studies have used ML techniques to estimate fundamental properties of fuel based on other available fuel parameters or test data. Wang et al. estimated minimum ignition energy of 61 chemical species using two quantitative structure−property relationship (QSPR) models based on experimental data. One was constructed using multiple linear regression (MLR) and the other was constructed using a support vector machine (SVM). Both models were successful at predicting the MIE. The MLR model was considered easier to implement and had better internal robustness. The SVM model provided a better fit and external robustness [1]. Pan et al. also implemented QSPR solved via a support vector machine SVM to predict the auto ignition temperature of organic compounds. A GA was used to select descriptors that contribute to the autoignition temperature (AIT). The SVM was able to utilize the selected descriptors to solve for the AIT and returned results within the experimental error. This method could be used to solve for AIT of organic compounds based on the theoretical descriptors only and not test data [2]. Shah et al. utilized random forests (RF) and deep neural networks (DNN) to predict the autoignition and flame properties of multi component fuels in homogenous charge compression ignition engines. Both methods outperformed an empirical model which required multiple equations for different fuel blends and equivalence ratios. It was noted that the RF required less user tuning than the DNN [3].

AIT is a fundamental fluid property defined by the ASTM standard E 659 [4]. Due to its well-defined nature and it extensive study in the literature it is good candidate for study when implementing ML learning techniques. Other studies have investigated the use of ML on more complex phenomenon that are not as well defined and are subject to additional physical parameters thus making them not fundamental fluid properties. Blurock et al. used machine learning cluster techniques utilizing the fuzzy logic to study the ignition of ethanol at varied temperatures. The technique was able to identify four phases of ignition for ethanol: the initiation phase, preignition phase, ignition phase, and the post ignition phase. The model was also capable of predicting the ignition delay time and phase structure for other temperatures within the range of the provided training data [5]. Similarly, Jach et al. investigated ignition delay time for various hydrocarbon air mixtures through the use of a DNN. Data was collected from a shock tube. The

results were compared with a detailed reaction mechanism and the DNN produced lower error and required less computational time [7]. Ulcay et al. studied hot surface ignition in a crossflow using both a Classifier and Regression Tree algorithm (CART) and a heat transfer based empirical model based off of experimentally determined constants. The CART algorithm was able to more accurately capture the effects of low air velocity on the hot surface ignition temperature when compared to the empirical model. Elsewhere in the range of velocities studied the two models performed comparably [8].

AI and ML techniques have also found applications in closely related fields such as CFD analysis, early warning systems for engine control, and heat transfer analysis. Sotgiu et al. developed a method for modeling gas turbine blades using tensor representation theory and ML in combination with CFD. The turbine blades are approximated as the Poiseuille flow with heat transfer. ML is used to approximate the turbulent heat fluxes that would normally be solved via the Reynolds-averaged Navier-Stokes equation. The machine learning derived model was implemented in OpenFOAM and shows good agreement with Poiseuille flow at different Reynolds numbers and reference data [9]. Popov et al. modeled a hydrogen jet in air cross flow ignited by laser-induced optical breakdown via CFD using three different methods to reduce the computational cost and to predict the ignition boundary. The first method, a constructed criterion based on radicals near the stoichiometric surface demonstrated 90% accuracy. The other two methods utilized neural networks, one utilized Bayesian regularization-enhanced training the other utilized convolutional neural networks (CNN); both were able to predict 100% of ignitions [10]. Wolf et al. implemented a deep automotive diagnostic network (DADN) using four CNN and two long short-term memory (LSTM) neural networks. Using data from an engine control unit the DADN was able to detect preignition with 90% accuracy. The DADN was found to outperform CNNs, LSTMs, and traditional neural networks when applied by themselves [11]. Baghban et al. used three methods including multi-layer perceptron artificial neural network , adaptive neuro-fuzzy inference system , and least squares support vector machine (LSSVM) to model Nusselt number as a function of multiple test parameters for carbon nano tube and water nano fluid flows through coils. It was found that LSSVM was the most accurate methodology used [12]. Wei et al. used a set of trustable data to train using three different methods, the support vector regression (SVR), Gaussian process regression (GPR), and CNN to predict the effective thermal conductivity of composite materials. It was found that all three methods used outperformed the Maxwell-Eucken model and the Bruggeman model in terms of accuracy. The computational cost of SVR, GPR, and CNN are far smaller than that of solving the partial differential equations and thus can be used as a fast prediction tool and can supplement physic based models [13].

The available literature shows a wide range of applications for AI and ML techniques. Additionally, the literature shows that there is a wide range of algorithms and methodologies that can be employed. Each application may be suited for one type of ML and not another. In order to understand which style of algorithm is appropriate for a data set it is beneficial to perform a multitude of ML techniques in order to understand which method works best. Applying several techniques can be easily accomplished with the mass adoption of software packages such as SciKit Learn, Tensor Flow, and Keras. Furthermore, processing power can now be delegated to server systems such as Google Colab which allow user to utilize high speed GPU's to process and solve the algorithms at no monetary cost to the user. Due to the adoption and availability of these techniques and resources, it is clear that AI will continue to play a larger role in future data analysis as well as finding new uses for pre-existing data sets.

III. Methodology

The current study investigates two data sets. The first data set used training comes from Ref. [14], [15], and [15].These studies investigated the flame spread rates for aviation fuels as a function of fuel temperature over a large pool of fuel. The bulk of the data comes from Ref. [14] in which the experiments of interest consist of 180 cm long x 20 cm wide x 2.5cm deep rectangular pans containing Jet-A and three alternative fuels Fischer-Tropsch IPK (FT-IPK), hydro-processed ester fatty acids (HEFA) and synthetic iso-paraffin (SIP). Each fuel was heated with a recirculatory heating system to maintain test fuel temperatures and ensure temperature uniformity within the test pans. A torch ignition method as well as a Nd:YAG laser were utilized. The transient liquid-phase and gas-phase temperatures and the arrival of the flame front were captured with 16 K-type thermocouples. Ref. [14]provides fuel composition information for the fuels including a detailed list of alkanes and aromatics with each of the fuels test. Similar fuel properties were gathered for the fuels used in Ref. [15] and [15] were gathered from publicly available sources. The fuel properties including fuel type, AIT, cetane number, molecular weight, density, flash point, boiling point, and weight percentages of paraffins, cycloparaffins, and aromatics along with test parameters such as initial

fuel temperature and flame speed were utilized for training and testing the ML techniques used in this study. The total data set consist of 162 data points.

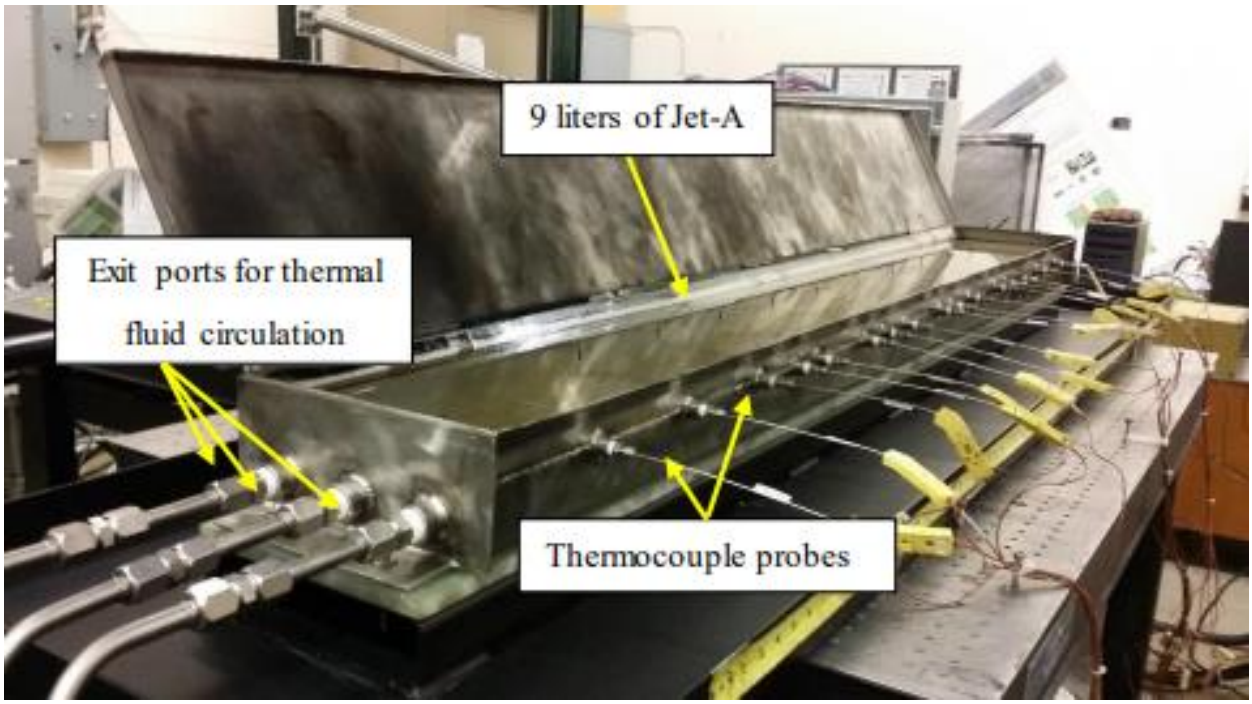

Figure 1: Test apparatus for flame speed measurement [14].

The second set of data investigated also come from Ref. [14] and pertains to a hot surface ignition test. Test fuels were deposited from a varied height in single drops from a programmable syringe onto a stainless-steel plate or round cylinder that was electrically heated to within ±5°C of a target temperature. High speed and infrared photography were used to detect the presence of ignition kernels. The surface temperature was varied to construct logistics curves to map the ignition probability as a function of surface temperature. The fuels studied include n-heptane, Jet-A, JP-8, JP-5, Avgas 100LL, and 4 experimental fuels for which fuel properties were not available. Fuel type, density, AIT, flash point, relative vapor density, boiling point, surface type, injection height, probability, and surface temperature were used as variables for training and testing the ML models. The complete data set consists of 221 points spread over all the permutations of the variable of interest. For fuel properties that were not available, multiple imputation by chained equations (MICE) was utilized to fill the gaps in the data set. Imputation is a process by which missing data is filled in via statistical methods. The most basic method of which would be to utilize the mean as a stand in value. This is done when removing the data point would greatly diminish model accuracy. Due to the small number of available data points imputation is a valuable tool. MICE utilized multiple chained imputations often with a specified regression method to create a more complex and accurate result based on the other available variables in the data set [18].

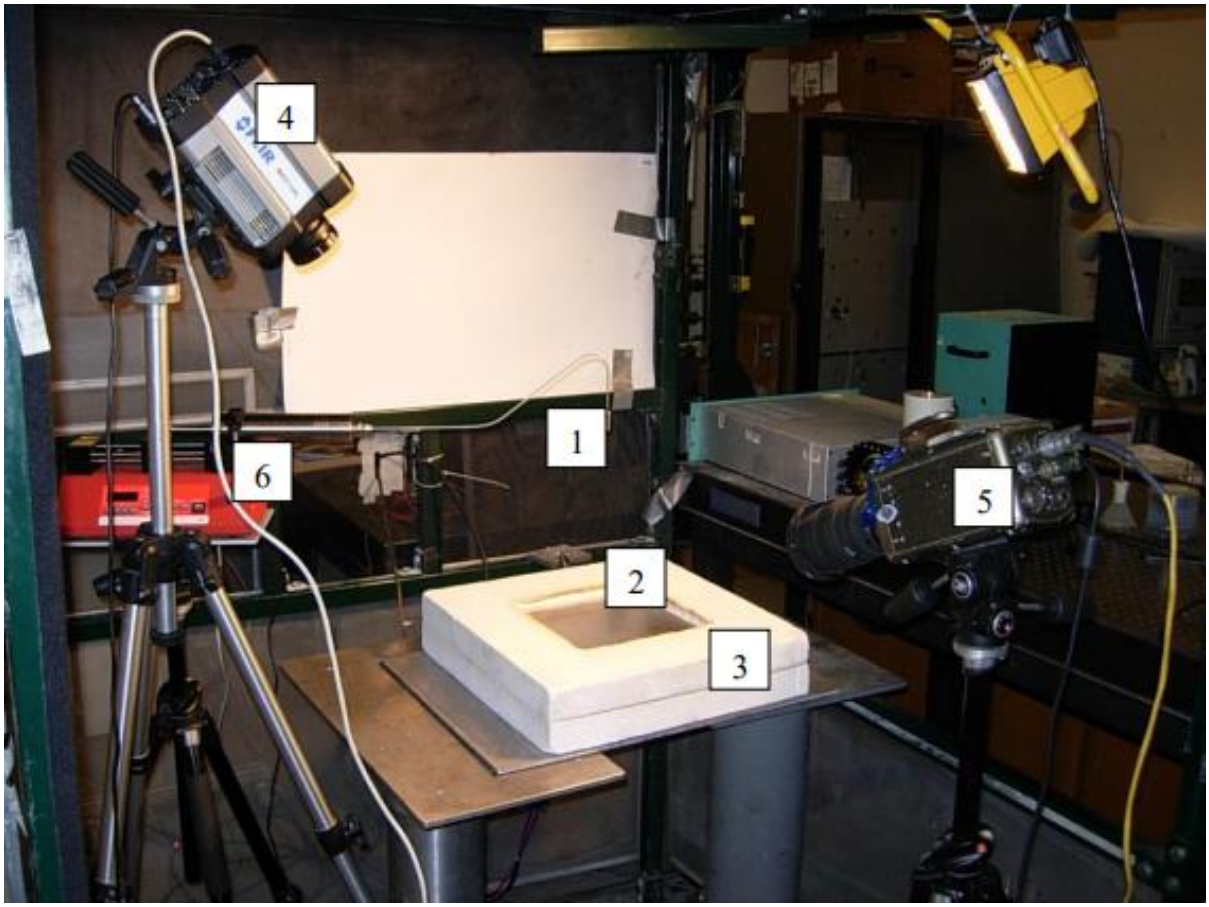

Figure 2: Experimental apparatus for hot surface ignition [14], 1) fuel nozzle, 2) hot surface plate, 3) ceramic insulation, 4) infrared camera, 5) high-speed camera, and 6) programable syringe.

Data handling and the implementation of the ML techniques was performed using the python language. Once data was passed into the software it was plotted to determine if a relationship between variables exists. In figure 3. the relationships for the hot surface ignition data are shown. Data that shows a linear or clumped shaped indicate that the variables have some relation between them. If two variables are unrelated then there would be no discernable trend between them. In the data shown below there are clear trends that can be seen between each of the variables. This points to this data set being a good candidate for ML technique application. Note that the diagonal axis shows a graph that represents the frequency of data points. In the data set that was investigated there is a strong preference for one fuel type as such the frequency graphs show a large spike around the value that represents that fuel. Traditional data analysis techniques may struggle with a data set that is so heavily skewed to, but ML is able to utilize this data to extrapolate the information provided by one subset of the data to another subset. This allows the generated models to more accurately predict outcomes for inputs for which there is limited data available. The significance of the ML or deep learning techniques lies in the fact of how well the model accurately predicts when the data is less related. The challenge of ML or deep learning techniques is to predict the outcomes and a clear deep learning technique winner can then be arrived at when high accuracies can be obtained for this case.

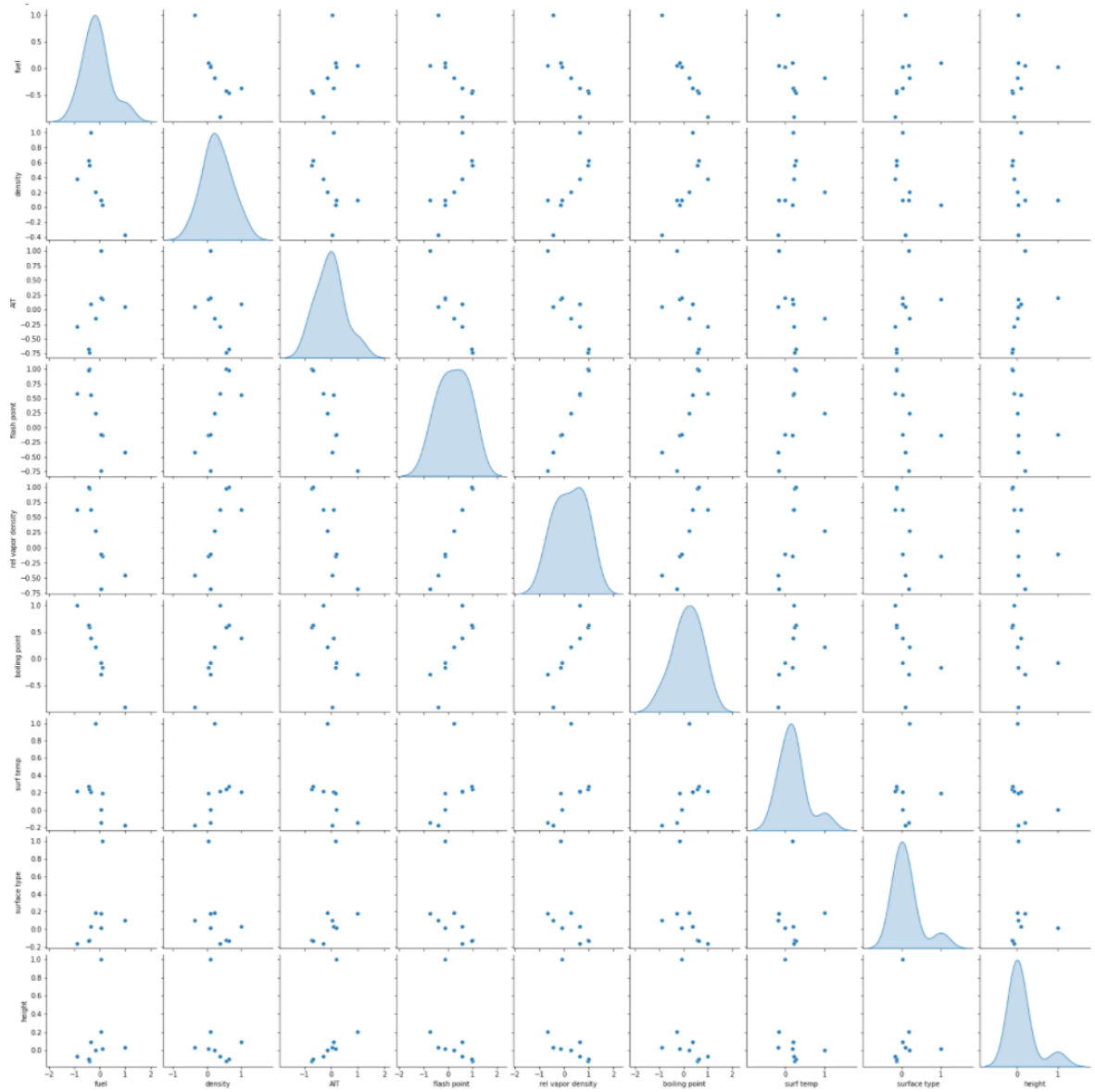

Figure 3: Plot of data relationships for hot surface ignition data with frequency plotted on the diagonal axis.

Once the data sets have been investigated for anomalies and pre-processed, they are split into training and testing sets for model creation and validation. Both sets of data are normalized to values between -1 and 1. This normalization greatly assists in model accuracy as many of the techniques implemented here are sensitive to data sets that contain multiple orders of magnitude. In order to have a baseline to compare with a simple and widely implemented MLR was used as the first method for generating an estimator model.

The second set of ML methods used were SVR models. This is similar to an SVM model, but the focus is to generate a regression instead of classification. Both SVR and SVM utilize the creation of hyperplanes in the multi-dimensional data space to maximize the margin between the data points and to minimize the error. There are several different kernels that can be used, each of which take the input data and transforms to be used for the creation of the hyperplanes. The present work uses the linear, polynomial (4 degrees of freedom) , and radial basis function (RBF) kernels. The degrees of freedom for the polynomial kernel were varied to find the best result. The next method used was a multivariant polynomial regression (MPR). The number of degrees of freedom were iterated over to determine the best result.

A Bayesian Ridge Regression (BRR) was also utilized. A BRR generates a linear fit based on probability distributors instead of a point estimate like a traditional linear regression would. A BRR pulls the output of the model from the generated probability distribution. The advantage of this method is that it allows for the fitting of models that contain insufficient data or poorly distributed data.

The next method investigated is a regression-based decision tree classifier algorithm or CART algorithm based on an Iterative Dichotomiser 3 developed in 1986 [19].In CART, a binary tree is grown using features and thresholds that yield the largest information gain at each node. The grown tree is pruned to improve model accuracy by removing pre-conditions at appropriate nodes. Figure 4 shows a pruned decision tree. While the example shown only utilizes two variables or test parameters, X_1 and X_2, the method can be extended to a multitude of variables. A set of variables X_i defined by the number of test parameters of interest is utilized within a decision tree. Threshold values, t_i, of the variables, X_i, are rapidly iterated over with a selected optimizer to identify the regions of interest, R_i. Once a tree has been constructed the branches of the tree that yield less information based on the selected optimizer are pruned away to leave a final decision tree that creates a decision boundary to separate the original test parameters into regions of interest that have unique characteristics. This final decision boundary is known as a classifier. Applying the above procedures leads to complex multi-dimensional decision boundaries that imitate the underlying physical phenomena that would otherwise take large amounts of computations and associated computer power to simulate and resolve with CFD.

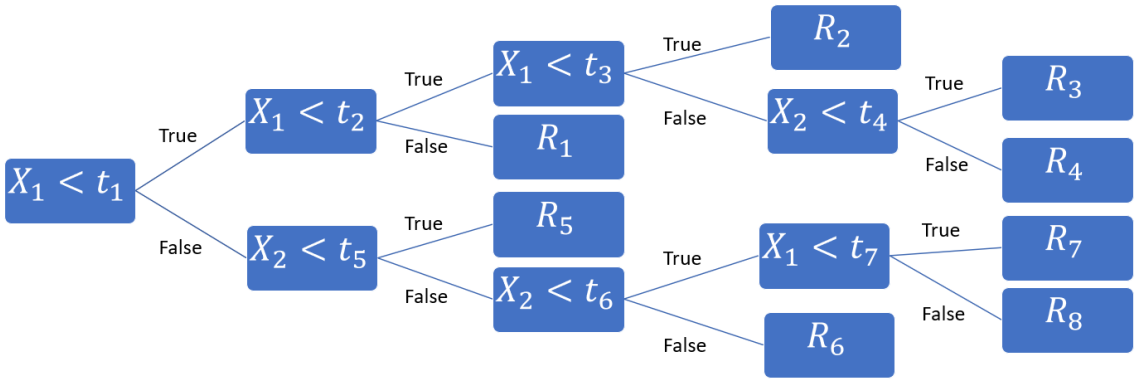

Figure 4: Flow chart diagram of CART model post pruning

A K-Nearest Neighbor (KNN) method was also implemented utilizing the 5 nearest neighbors. In this method the distance between parameters in one plane of the multi-dimensional data space is calculated. The algorithm predicts that neighbors in the same plane will share similar values in other intersecting planes of the data space. The model uses the *k* closest neighbors in the plane to estimate the desired out of plane value. The distance to the *k* nearest neighbors was calculated three different ways. Using the Euclidean, Manhattan, and Hammering distances. The algorithm automatically selects the distance that returns the lowest mean squared error.

The next method used is the first neural network utilized. A Multi-Layer Perceptron (MLP) with 100 neurons and 3 hidden layers was created. The hidden layers use the rectified linear activation function and the output layer uses backpropagation with no activation function. The learning rate is set to 0.001, the L2 penalty parameter is 0.001, the stochastic gradient-based optimizer (adam) is used for the solver, and the squared-loss is optimized for. In this method the data inputs are given in the input layer and the relationships between each of the parameters is investigated over several iterations. The activation function acts as a transform for the input of the neuron and defines the output. The algorithm seeks to minimize the loss of the model by taking the partial derivates with respect to the model parameters. The MLP model used here is contained in a software package and thus is limited in its customizability.

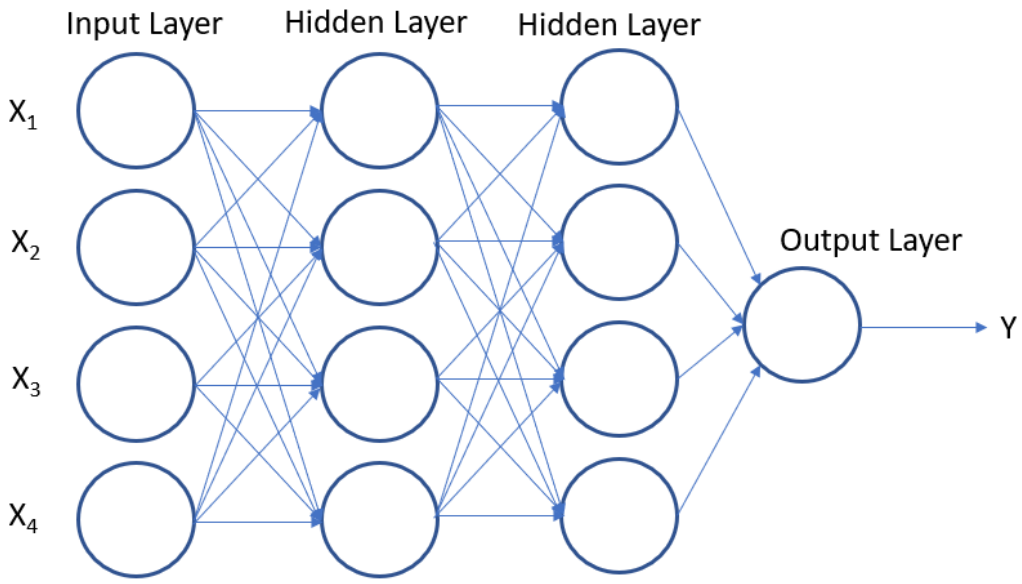

Figure 5: Flow chart of simplified MLP model

To alleviate the limited customization of the MLP a Distributed-Artificial-Neural-Network (DANN) algorithm is used. The DANN consist of several dense layers of neurons using the rectified linear activation function. Each dense layer is followed by a dropout layer that removes a set percentage of the next layer inputs. This is done to regularize the inputs of the following layer. In effect this prevents overfitting of the model. The final output layer of the DANN tailored to the data set. In the case of the hot surface ignition data a sigmoid activation function greatly improved model result. The rectified linear activation function produced the best result for the pool fire data. The model learning rate is set to 0.00001 and the mean squared error is used as the loss. The model is able to stop training based on a user defined parameter to expedite the data analysis process.

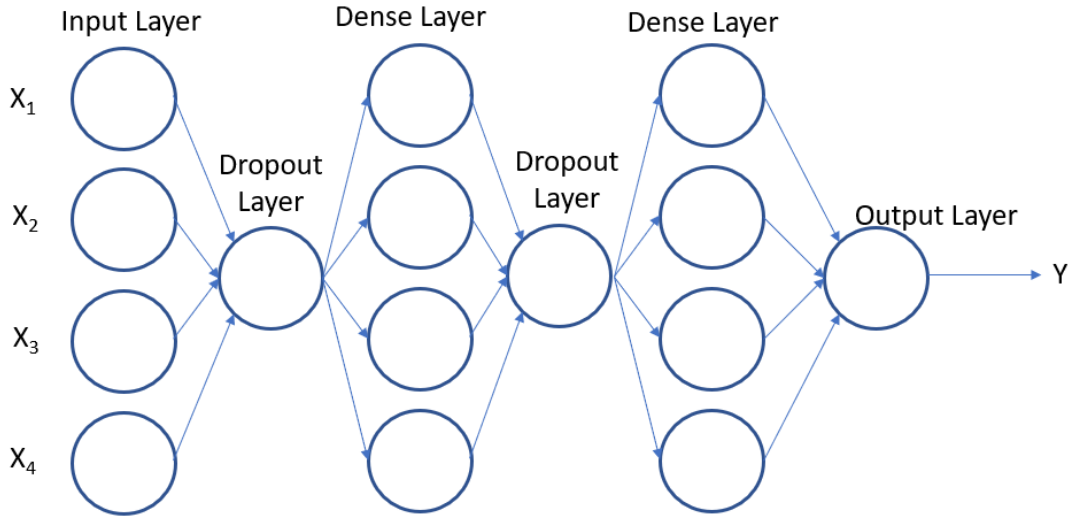

Figure 6: Flow chart of simplified DANN model

The training and testing data sets were randomly selected among the grouped test conditions. For both data sets the spilt is 75/25 where 75 percent of the data is used for training. The accuracy of each method was assessed by evaluating the R-squared of each of the regressions. Where possible the training and validation values for R-squared were compared to ensure overfitting was not occurring. If the R-squared values between the training and testing data sets are significantly different this indicates that overfitting may be occurring.

IV. Results

For the hot surface ignition data, the model was asked to predict the probability of ignition based on the provided test conditions. The results for each of the models can be seen in table 1. As expected, the MLR model struggles to capture the trends seen in the data. This makes intuitive sense. It is difficult to draw a straight line through a set of data that exists in a multivariant data space unless it is tightly clustered. The SVR model with linear kernel is similarly effects. The poly and RBF kernels return more accurate results and the training and test values are similar which indicates that the model is not over fitting. The MPR model also struggles to capture the phenomenon being investigated and at higher degrees of freedom the model quickly begins to greatly overfit. The BRR model is still creating a linear model and as such the fit is poor. It does outperform the MLR model which is the intended purpose of the BRR model when data is limited. The CART model returns a 0.749 R-squared value, but the difference between the R-squared values of the training and data set is large (0.19) which indicates overfitting. This is a common trait of CART models. They produce high levels of accuracy but are prone to overfitting. The KNN model produces an R-squared value of 0.574, but similarly to the CART model the difference between the training and test values for R-squared is large (0.22) indicating over fitting. The MLP model returns a result of 0.715 and the difference between the training and test values is 0.07 which is much smaller than the CART and KNN models. The DANN model produced the best result at 0.886 and the difference between the training and testing data sets was only 0.015. This indicates a good fit. Figure 7 shows the results decay of the training and validation loss over the epochs. The noisy structure of the lines shown indicates the model could use further tuning with a decrease in the learning rate.

Hot Surface Ignition Data		
Model type	R-squared with imputed data	R-squared without imputed data
MLR	0.166	0.158
SVR(linear)	0.11	0.144
SVR(poly)	0.515	0.658
SVR(RBF)	0.73	0.793
MPR	0.166	0.441
BRR	0.181	0.169
CART	0.749	0.64
KNN	0.574	0.621
MLP	0.715	0.697
DANN	0.886	0.897

Table 1: Results for hot surface ignition data.

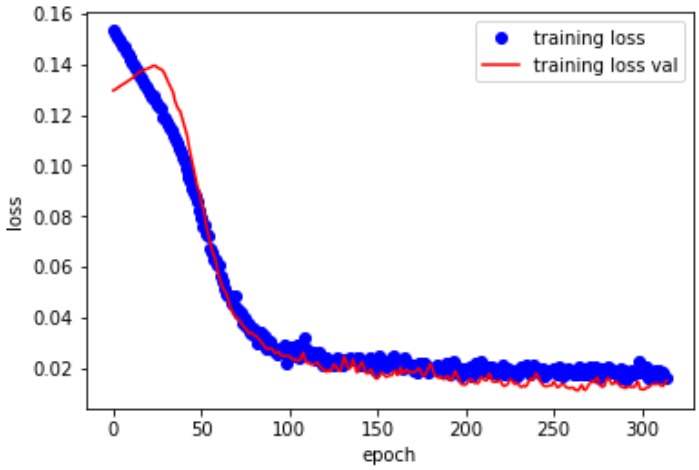

Figure 7: Loss for the DANN model over the training period consisting of 353 epochs.

When comparing the data that had been treated with the MICE algorithm with the alternative method of removing the parameters that contained missing data, in this case the AIT, relative vapor density, and flash point, there is little difference for the majority of the models created. The models utilizing polynomial methods saw the largest increase in performance from 0.515 to 0.658 for the SVR (poly) and from 0.166 to 0.411 for the MPR models. Both models still had a large spread in the values for the training and test data sets indicating overfitting. The CART algorithm saw a decrease in accuracy but maintained a large difference in the R-squared values for the training and test data sets. All other models only shifted by small amounts. The MICE algorithm can be further improved via tuning which may demonstrate larger gains in model accuracy.

For the pool fire data, the model was asked to estimate the flame spread rate for the given set of test parameters the results of which can be seen in table 2. For this data set the linear based models consisting of MLR, SVR (linear), and BRR all performed at higher levels of accuracy when compared to the hot surface ignition data. All of which had differences between the training and testing R-squared values less than 0.01. This indicates that the model is not over or underfitting and is performing as best it can. The SVR (poly) and SVR (RBF) also performed well and had differences between the training and testing R-squared values less than 0.01. The CART algorithm performed best of the non-neural network methods and had a difference between the training and testing R-squared values less than 0.03. This means that in this case the CART algorithm did not overfit. The MLP and DANN model preformed best overall with R-squared values of 0.905 and 0.92 respectively. In both cases the differences between the training and testing R-squared values less than 0.01.

Pool Fire Data	
Model type	R-squared
MLR	0.717
SVR(linear)	0.71
SVR(poly)	0.794
SVR(RBF)	0.819
MPR	0.819
BRR	0.753
CART	0.827
KNN	0.798
MLP	0.905
DANN	0.92

Table 2: Results for pool fire data.

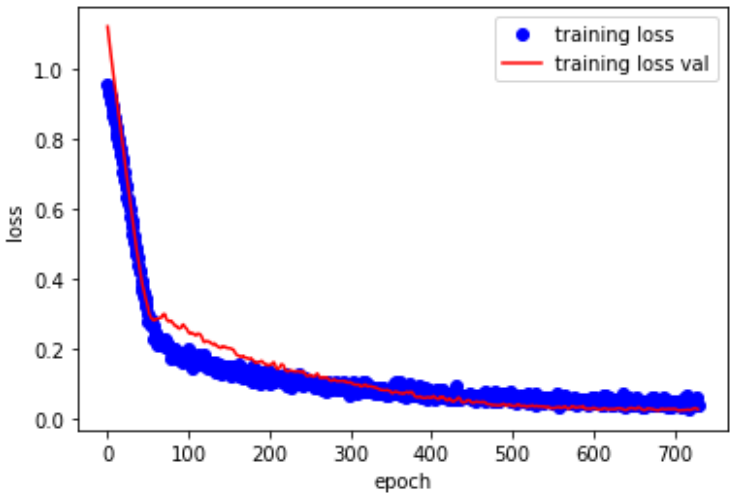

Figure 8: Loss for the DANN model over the training period consisting of 738 epochs.

Overall, the pool fire data was a much easier data set to model. This is clear from the higher level of accuracy from the models, but also clear when one looks at the data. The hot surface ignition data is capturing a stochastic process which is statistical in nature. There are often multiple expected values for one given set of input conditions. This reduces the accuracy of the regression models. To combat this either more data would be needed or a more through data preprocessing method would be needed to eliminate these inconsistencies. For both data sets the neural network methods outperformed the other ML techniques. The customizability of the DANN model allowed for it to be more easily tailored to the data set making it a more versatile model. Both the MLP and DANN models were also the most computationally expensive requiring noticeable solving time to converge on a result 5-10 minutes. The other ML techniques solved in less than 1 second meaning that they were far less computationally expensive, but at the cost of accuracy.

V. Conclusion

AI and ML data analysis techniques are rapidly becoming widespread tools for data analysis and modeling. The best methods and algorithms are often data set specific and as such need to be experimented with in order to determine the best set of ML techniques to provide the best result. Two data sets were investigated, a pool fire data set investigated the flame spread rate of 8 different fuels including alternative aviation fuels and a hot surface ignition data set that studies the ignition probability as a function of surface temperature for 9 fuels including 4 experimental fuels for which fluid properties were not available. The MICE algorithm was used to fill the missing fluid properties and was compared with models derived from a data set where the missing properties were removed as variables. 10 different ML techniques were used to compare the effectiveness of different methodologies. For both data sets neural networks provided the highest level of accuracy (0.92) with the least amount of over and under fitting. This high accuracy came at a higher computation cost when compared to other ML techniques, but at a substantially lower cost than CFD derived results. The specific method for developing an accurate estimation model was shown to be data specific and the best results came from models that could be highly tailored to the data set in question over several repetitions. Future work for the data sets would include the application of deep multi-layer-perceptron DMLP model to a variety of experimental combustion data. Several databases in use today date back over 30 years and DANN model may offer new analysis methods for extracting additional value from these older data bases.

VI.References

[1] S. A.Kalogirou, "Artificial intelligence for the modeling and control of combustion processes: a review," *Progress in Energy and Combustion Science*, vol. 29, no. 6, pp. 515-566, 2003.

[2] B. Wang, L. Zhou, K. Xu and Q. Wang, "Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure–Property Relationship (QSPR) Models," *Industrial & Engineering Chemistry Research*, vol. 56, no. 1, pp. 47-51, 2017.

[3] Y. Pan, J. Jiang, R. Wang, H. Cao and Y. Cui, "Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine," *Journal of Hazardous Materials*, pp. 1242-1249, 2009.

[4] N. Z. P. D. D. a. G. H. Shah, "Prediction of Autoignition and Flame Properties for Multicomponent Fuels Using Machine Learning Techniques," *SAE Technical Paper*, p. 8, 2019.

[5] A. S. E659-15, "Standard Test Method for Autoignition Temperature of Chemicals," ASTM International, 1978.

[6] E. S. Blurock, "Automatic characterization of ignition processes with machine learning clustering techniques," *International Journal of Chemical Kinetics*, pp. 621-633, 2006.

[7] A. Jach, M. Zbikowski, K. Malik and A. Teodorczyk, "Ignition delay time model based on a deep neural network," in *27th ICDERS*, Beijing, China, 2019.

[8] M. S. Ulcay, L. N. Dillard, J. P. Gore and P. C. Sweeney, "Machine Learning based Calculations of Minimum Hot Surface Fluid Ignition Temperature," in *Spring Technical Meeting Central States Section of the Combustion Institute*, Huntsville, 2020.

[9] C. Sotgiu, B. Weigand and K. Semmler, "A method for modeling gas turbine blades using tensor representation theory and machine learning in combination with CFD is implemented. The turbine blades are approximated as the Poiseuille flow with heat transfer. Machine learning is used to approximate," *International Communications in Heat and Mass Transfer*, vol. 95, pp. 74-79, 2018.

[10] Popov, Pavel P, David A Buchta, Michael J Anderson, Luca Massa, Jesse Capecelatro, Daniel J Bodony, and Jonathan B Freund., "Machine Learning-assisted Early Ignition Prediction in a Complex Flow," *Combustion and Flame*, vol. 206.C, pp. 451-66, 2019.

[11] A. M. T. T. N. B. B. a. S. G. P. Wolf, "Pre-ignition Detection Using Deep Neural Networks: A Step Towards Data-driven Automotive Diagnostics," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, 2018.

[12] M. K. M. A. N. M. H. A. W.-M. Y. Alireza Baghban, "Sensitivity analysis and application of machine learning methods to predict the heat transfer performance of CNT/water nanofluid flows through coils," *International Journal of Heat and Mass Transfer*, vol. 128, pp. 825-835, 2019.

[13] H. Wei, S. Zhao, Q. Rong and H. Bao, "Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods," *International Journal of Heat and Mass Transfer*, vol. 127, pp. 908-916, 2018.

[14] V. Goyal and J. P. Gore, "INVESTIGATION OF FIRE SAFETY CHARACTERISTICS OF ALTERNATIVE AVIATION FUELS," West Lafayette, 2019.

[15] D. B. C. L. F. C. &. L. J. White, "Flame spread on aviation fuels," *Fire safety journal*, vol. 28, pp. 1-31, 1997.

[16] M. H. L. S. X. G. J. &. T. K. L. Li, "Study on Flame Spread over Aviation Kerosene and Diesel," *Advanced Materials Research*, vol. 1016, pp. 587-591, 2014.

[17] V. Goyal, A. B. Carayon, S. Meyer and J. P. Gore, "Hot Surface Ignition Temperatures of Hydrocarbon Fuels," in *AIAA*, Grapevine, TX, 2017.

[18] S. Van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, vol. 45, no. 3, 2011.

[19] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, 1992.