Application of machine learning and deep learning approach to evaluate the thermal efficiency of flex-fuel diesel engine

Vivek Prajapati[a], Shashi Kumar Jain[a], Ankit Goyal[a], Sunil Kumar[b]

[a] *TIT&S, Anand Nagar, Bhopal, 462021, (M.P.), India*
[b] *R.G.P.V., Gandhi Nagar, Bhopal, 462033, (M.P.), India*

Abstract

This work propose the use of machine learning and deep learning models which can predict ethanol mass ratio (EMR) and thermal efficiency at different load factors. Data visualization is done to see the variation of thermal efficiency at different load factors and different EMRs. Viable data of flex-fuel engines at different Ethanol mass ratio (EMR) and load factors has been used. Different machine learning and deep learning algorithms have been used to create and validate different models. This investigation reveals that ANN and XGBoost algorithms gives the best output for the prediction of ethanol mass ratio (EMR) with 2.2 percent and 1.6 percent mean absolute percentage error. The same algorithms are also performing very well for the prediction of thermal efficiency with less than 1 percent error. From this study, it is observed that efficiency variation is highly dependent on EMR but remains in the range of 43 - 52%.

Keywords: Flex-fuel vehicle, Ethanol, Ethanol mass ratio learning, Deep learning, Artificial intelligence

1. Introduction

1.1. Flex-Fuel Vehicles

Alternative to fossil fuels include biodiesel, electricity, ethanol, hydrogen, natural gas, and propane. Worldwide a lot of research is going on to analyze feasibility of ethanol as viable alternative to petroleum-based fuel. Brazil has already switched to ethanol mixed fuel to replace the majority of its alternative fuel requirements. Flex-fuel vehicles (FFVs) in Brazil are powered

Preprint submitted to Simulation modelling practice and theory *March 27, 2023*

by E85 and E100. Brazil, United States, Canada, and European Union, particularly Sweden, have all worked hard on these flex-fuel technologies [1].

The Indian government is monitoring the use of flexible fuel transport facilities in order to encourage the use of biofuels. In the future, India will also be required to replace the fuel for transportation. The Indian government is in touch with various automobile manufacturers to make improvements to their engine technologies so, that they can run on the flex-fuel[2].

Ethanol is currently mixed in petrol in India at a rate of 10%. the government is attempting to boost this ratio by 20% in 2024. So, in India, FFVs are the way of the future [3].

The FFV may run on diesel or on a blend of diesel and ethanol up to 83 percent by volume. According to the region and season, flex-fuel is a diesel-ethanol blend that includes between 51 and 83% ethanol, depending on the blend. Many flex-fuel vehicles (FFVs) have enhanced acceleration capability when using higher ethanol blends, despite the fact that higher ethanol blends impair fuel economy as the engines are primarily designed for diesel as fuel. Ethanol has a high research octane number (RON) as well as a high heat of vaporization [4, 5].

Strategies of flex-fuel (diesel-ethanol) utilization in I.C. engines are as follows[6]:

1. Flex-fuel fumigation into the intake air charge.
2. The complete substitution of commercial flex-fuel via a port injection or direct injection.
3. The use of flex-fuel in a dual-fuel automobile.

To use flex-fuels in I.C. engines several modifications are required. Reverse engineering is one of the growing methods to improve modify engines. One of the new techniques in reverse engineering for analyzing and improving engines is using Information Systems Artificial Intelligence. In this, data is collected from an engine and by using machine learning algorithms prediction models are made, which can be utilized for engine modification purpose.

1.2. Artificial Intelligence

Artificial intelligence enables machines to automate decision making without human intervention. It makes the process cleaner, clearer, faster, and more data-driven. AI is a tool that can be used in various fields such as data science. Data Science is about finding hidden patterns in the data. Data science can be used in various fields, for example in energy demand, machine learning is being used to slow down energy consumption [7]. Even in medical field, machine learning and artificial intelligence are being used to identify major health dieses like - Alzheimer's [8].

Artificial intelligence is becoming an important tool in the fields of materials and mechanical engineering, attributed to its power to predict materials properties, design and discover new mechanisms. It is used in various sectors, for example: manufacturing, production, quality check etc.

Use cases of AI in Mechanical Engineering

1. Artificial intelligence is used in modern warfare. Drones powered with artificial intelligence are able to take warfare at next level [9].
2. Artificial intelligence is used in aviation industry. Flight control systems, traffic controlling are some examples of it [10].
3. Artificial intelligence can be used to predict influence of tool geometry and cutting circumstances on grain size in Finite element simulation [11].
4. Railway is also using AI for automation, intelligent development, and better serving the needs of peoples in the new era. In the multi-mode rail transit transfer station, an integrated pedestrian facilities design and personnel assignment problem is solved using a simulation and machine learning based optimization technique [12].

A subfield of artificial intelligence is machine learning. In this field a machine is trained to perform some tasks. Machine learning is subdivided in several fields. Two major fields of machine learning are as follows [13]:

1. Supervised learning
2. Unsupervised learning

There are multiple machine learning models which can be used for prediction. The choice of each model depends on several factors. One of the common techniques is to apply several algorithms depending upon the types of business need and then choose the best one. There are several examples of comparisons of different algorithms [14].

2. Proposed Methodology

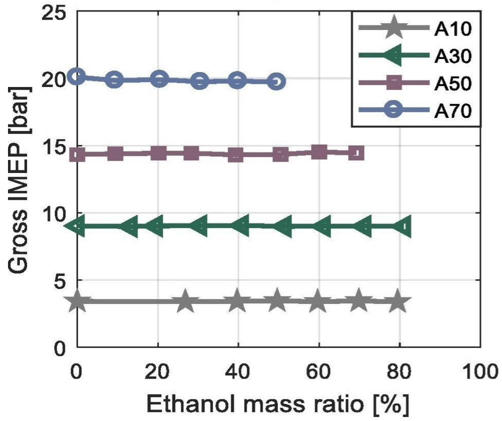
Machine learning in artificial intelligence technology is used to detect and understand the relations between different variables and after establishing these relations it can create a model to predict the dependent variable with respect to other independent variables. This work proposes machine learning and deep learning to predict the ethanol mass ratio and thermal efficiency. In this section, the steps are shown to create a model in machine learning and deep learning from scratch which can also predict the proposed outcome and visualize it. Different steps for processing are as follows:

2.1. Data Collection

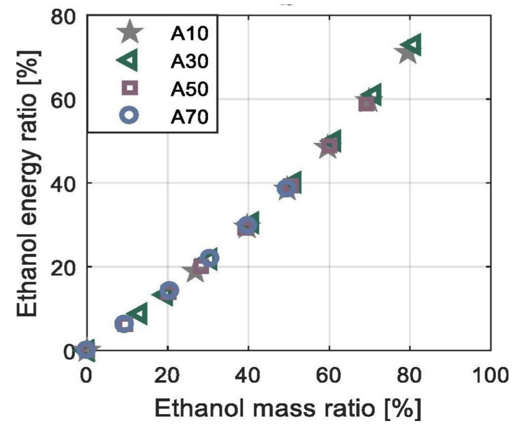
Data has been collected from different sources, literature survey is done and verified data is extracted in the form of graphs of a diesel engine [6]. The engine had run on different load factors with different Ethanol mass ratios. These graphs are shown in Figure 1(a)- 3(b).

Steps: -

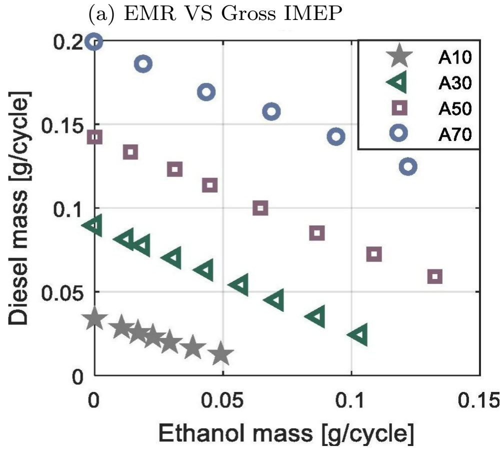
1. Produced CSV data of each graph at A10 load is put it into a folder called A10 data. A10 folder contains 16 data sets, in which each dataset has 2 variables. On the X variable, we have EMR, and the y-axis contains other variables depending on the graph data. Automeris online platform is used for this conversion [15].
2. Now these 16 datasets are merged using outer join on a common variable (EMR). So, we got a new data set (i.e. A10-full.csv). It contains 17 variables.
3. Sort the observations by EMR and reset the index.
4. Dataset contains some null values since data points are near, so backward linear interpolation is used to fill the missing values. After removing lower data point observations where missing values were present, we now have a cleaned dataset (A-10-full.csv) at load A10.
5. Since this dataset corresponds to a 10% load factor, we added a new folder named Load Factor and put a value of 0.1 in each observation.
6. In the same manner, we have created 3 more folders A30, A50, and A70. Repeat steps (1-5) 3 times for A30, A50, and A70.
7. Now we have appended all the 4 data sets (A10-full, A30-full, A50-full, and A70-full) into a new dataset called df-full
8. This data set contains 3800 observations. It is used to train ML model.



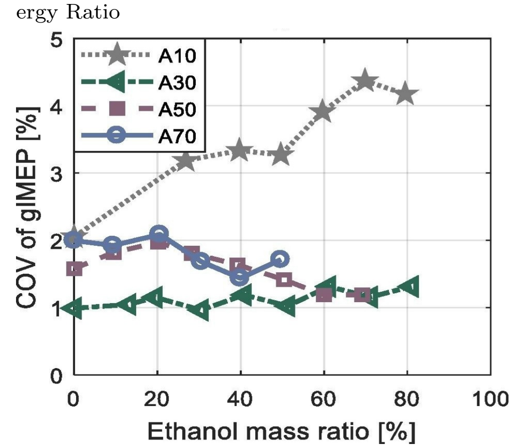
(a) EMR VS Gross IMEP



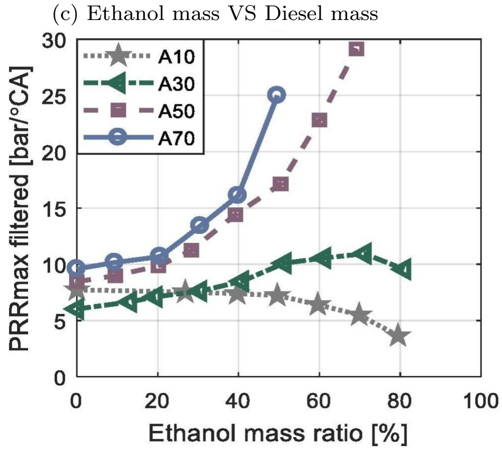
(b) Ethanol mass Ratio VS Ethanol Energy Ratio



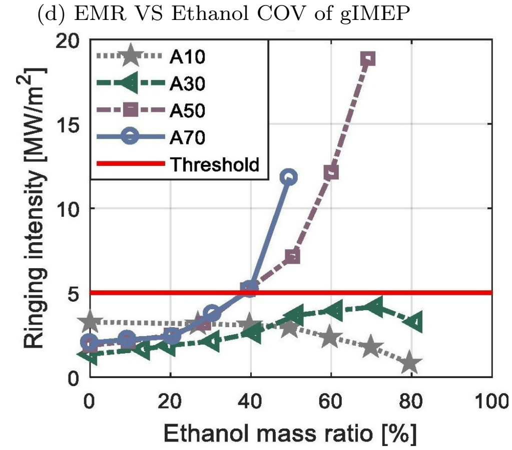
(c) Ethanol mass VS Diesel mass



(d) EMR VS Ethanol COV of gIMEP

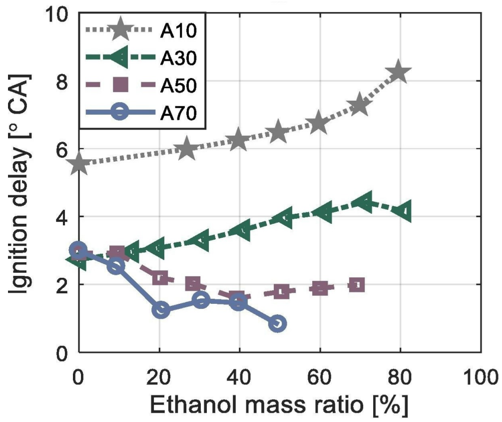


(e) EMR VS PRRmax filtered

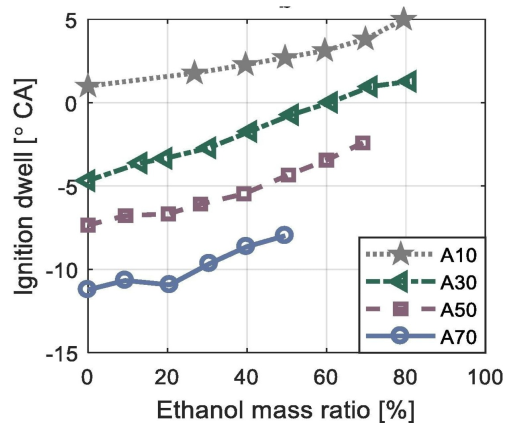


(f) EMR VS Ringing intensity

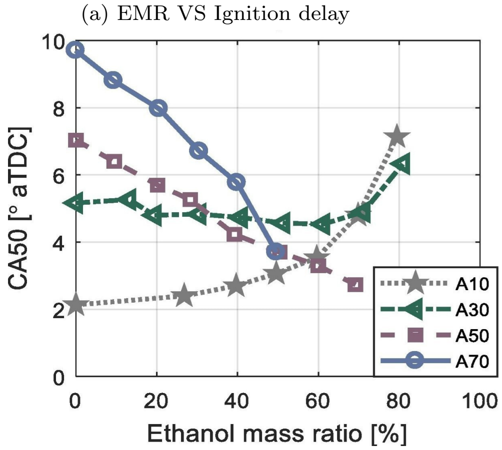
Figure 1: Variation of EMR with different properties



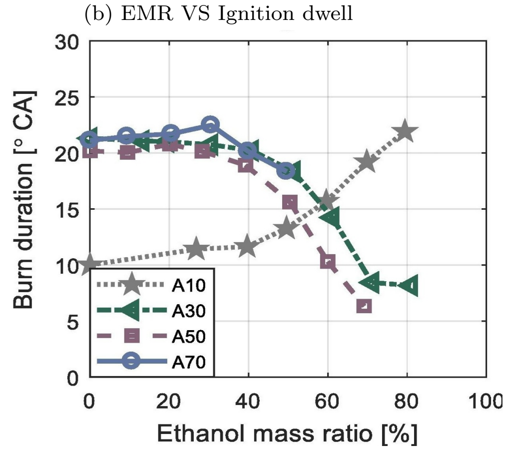
(a) EMR VS Ignition delay



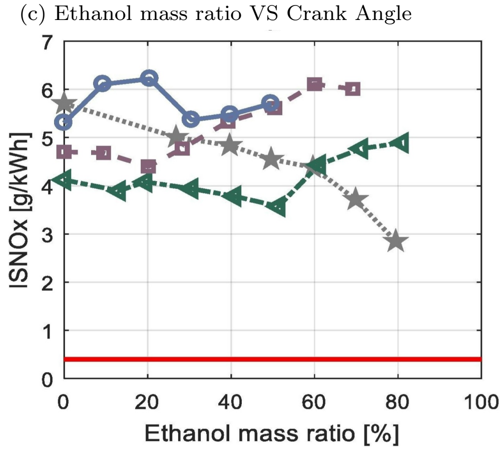
(b) EMR VS Ignition dwell



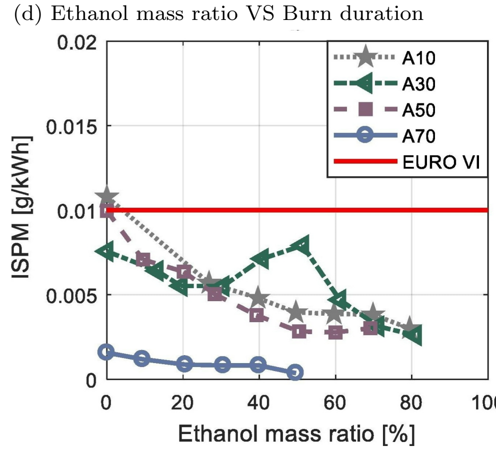
(c) Ethanol mass ratio VS Crank Angle



(d) Ethanol mass ratio VS Burn duration

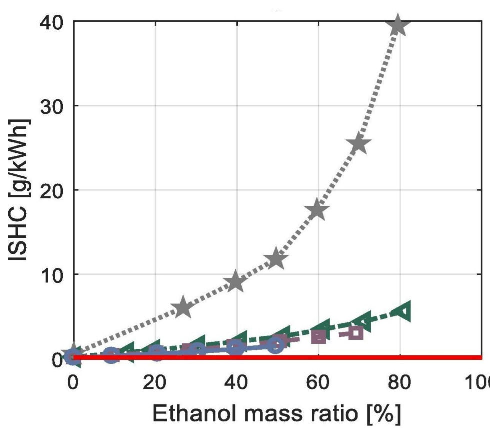


(e) Ethanol mass ratio VS NOx emission

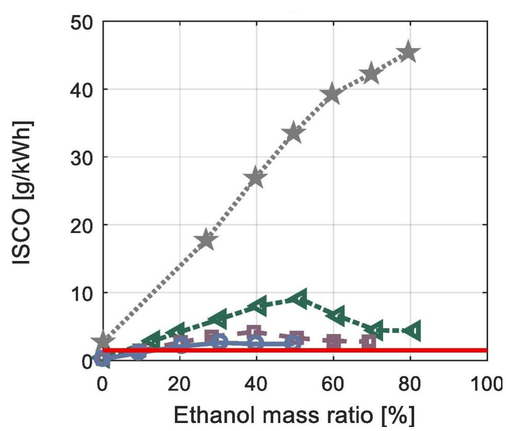


(f) Ethanol mass ratio VS PM emission

Figure 2: Variation of EMR with different properties



(a) Ethanol mass ratio VS hydro-carbon emission



(b) Ethanol mass ratio VS CO emission

Figure 3: Variation of EMR with different properties

2.2. Data Analysis

Data analysis can be done through various tools like- Excel, Power-BI, Tableau, R, Python. Here we have used Python as coding language, Jupyter Notebook as an IDE (integrated development environment). Data visualization is done at 3 levels, namely -

1. Univariate Analysis: It shows the distribution of each variable. It is the most fundamental type of data analysis.
2. Bi-Variate Analysis: Bivariate analysis is done for EMR and thermal efficiency (dependent variable) with other independent variables.
3. Multi-Variate Analysis: it looks at numerous variables (more than two) to see if there is any conceivable link between them. The correlation ranges from (-1) to (+1).

2.3. Feature Selection

The data attributes used to train machine learning models significantly impact model's accuracy. So, feature selection is done to remove high dimensionality. It can be done either automatically or manually.

2.4. Dividing the data into train test

Train/Test is a technique for determining model's correctness. Split the data set into two sets: a training set and a testing set, it's termed Train/Test split. The training set is used to train the model and the testing set is used to test the model. Data is split into 80:20 ratio for train and test purpose.

2.5. Feature Engineering

Feature engineering (scaling) is done on independent variables. It is also known as data normalization in data processing. Standard scaler is used to scale down datapoints.

2.6. ML Algorithms

Regression machine learning algorithms are used for prediction of Ethanol mass ratio (EMR) and thermal efficiency.

2.6.1. Linear Regression

As a reason of its simplicity, linear regression is frequently used. It is a prevalent and helpful approach of generating predictions when the target vector is a quantitative value. In order to construct accurate prediction models, a number of linear regression approaches are used.
It is assumed that the connection between features and target vectors is linear. Thus, the characteristics (also known as coefficient, weight or parameter) that impacts the goal vector remains constant. The basic equation of linear regression is as follows-

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{B}_1\mathbf{X}_1 + \mathbf{B}_2\mathbf{X}_2 + \text{€}$$

The model in basic linear regression is trained to minimize the sum of squared error, also known as the residual sum of squares (RSS) between the true values (Yi) and predicted values (ŷ):

$$RSS = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

Where ŷ is target variable, x is the independent variable, and the coefficients identified after fitting the model are B0, B1, B2, and €. After we fit our model, we can view the value of each parameter. For example, bias or intercept [16].

2.6.2. Support Vector Regression

Support vector regression works on the concept of hyperplanes. A hyperplane is a subspace in an n-dimensional space which can be used to differentiate 2 different classes.

$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i K\left(x_i, x_{i'}\right)$$

$$K\left(x_i, x_{i'}\right) = \sum_{j=1}^{p} x_{ij}, x_{i'j}$$

For instance, use of a one-dimensional hyperplane to partition a two-dimensional space. A two-dimensional hyperplane would be used to partition a three-dimensional space. Support vector regression is merely an n-dimensional extension of this idea. Support vector regressions categorize data by determining the optimal hyperplane [17].

2.6.3. XG-Boost Regression

XG-Boost is a tree-based algorithm. It comes under ensemble method. It uses gradient boosting technique to find efficient output. Multiple models (known as base learners) are trained and merged to give a single prediction in the XGBoost technique. A regularisation function and a loss function are included in the objective function. It describes the gap between actual and expected values, or how close the model outputs are to the actual values. Regularization is used to reduce overfitting [18].

2.7. Model Creation

There are a variety of machine learning models to choose from, the task is to determine best one. One of the methods is to test several models and fine-tune parameters to get the most accuracy possible.
We have used various machine learning algorithms linear regression, Support vector regression, XG Boost and artificial neural network.

2.7.1. EMR Predictor Model

Training different models on X-train and y-train. then predicting the outcome on X-test. So, we get y predicted (y-pred). Now to check the performance of the model, used y-test and y-pred. Here we are checking root mean squared error (RSME) and mean absolute percentage error (MAPE).

a. Training Linear Regression model

```
# fit linear-Regression model on training data
linear_regressor = LinearRegression()
linear_regressor.fit(X_train, y_train)
y_pred_linear = linear_regressor.predict(X_test)
rmse_linear = np.sqrt(mean_squared_error(y_test,
    y_pred_linear))
print("RMSE: %f" % (rmse_linear))
# Using MAPE error metrics to check for the error rate and
    accuracy level
MAPE_linear= MAPE(y_test,y_pred_linear)
print("linear MAPE: ",MAPE_linear)
```

Listing 1: Linear Regression Model

b. Training SVR model

```
#### fit support vector Regression model on training data
    ####
svr_regressor = SVR(kernel = 'poly')
svr_regressor.fit(X_train, y_train)
y_pred_svr = svr_regressor.predict(X_test)
# evaluate predictions
rmse_SVR = np.sqrt(mean_squared_error(y_test, y_pred_svr))
print("RMSE: %f" % (rmse_SVR))
# Using MAPE error metrics to check for the error rate and
    accuracy level
MAPE_SVR = MAPE(y_test,y_pred_svr)
print("SVR MAPE: ",MAPE_SVR)
```

Listing 2: SVR Model

c. Training XG boost model

```
xgb_model = XGBRegressor()
xgb_model.fit(X_train, y_train)
y_pred_xgb = xgb_model.predict(X_test)
# evaluate predictions
rmse_XGB = np.sqrt(mean_squared_error(y_test, y_pred_xgb))
print(" RMSE: %f" % (rmse_XGB))
# Using MAPE error metrics to check for the error rate and
    accuracy level
MAPE_XGB= MAPE(y_test,y_pred_xgb)
```

```
print("XG Boost MAPE: ",MAPE_XGB)
```

Listing 3: XGBoost Model

d. Training ANN model

```
#### fit ANN model on training data ####
### Initializing the ANN
ann = tf.keras.models.Sequential()
### Adding the input layer and the first hidden layer"""
ann.add(tf.keras.layers.Dense(units= 139, activation='relu'))
### Adding the second hidden layer
ann.add(tf.keras.layers.Dense(units= 139, activation='relu'))
### Adding the third hidden layer
ann.add(tf.keras.layers.Dense(units= 139, activation='relu'))
### Adding the output layer
ann.add(tf.keras.layers.Dense(units=1, activation='relu'))
### Compiling the ANN###
ann.compile(optimizer = "Adam", loss = "MAPE", metrics = ['
    accuracy','MAPE','mean_squared_error'])
### Training the ANN model on the Training set
ann.fit(X_train, y_train, batch_size = 4, epochs = 400)
### Prediction on X_test
y_pred_ANN = ann.predict(X_test)
rmse_ANN = np.sqrt(mean_squared_error (y_test.reshape(777,),
    y_pred_ANN.reshape(777,)))
print("RMSE: %f" % (rmse_ANN))
# Using MAPE error metrics to check for accuracy level
MAPE_ANN = MAPE(y_test.reshape(777,),y_pred_ANN.reshape(777,)
    )
print("ANN MAPE: ",MAPE_ANN)
```

Listing 4: ANN Model

2.7.2. Efficiency Predictor Model

In a similar manner, we can create a Thermal efficiency predictor model. None of the above codes will change. Only X and Y sets will be changed. Which we have already described in section (2.7).

2.8. Model Selection

There may be numerous conflicting factors beyond model performance when it comes to model selection, such as complexity, maintainability, and available resources.

3. Results and Discussions

3.1. Accuracy of EMR Predictor

In the trained models, we have used xgboost, Support vector regression, linear regression and ANN. which are giving good performance. shown in Table 1-

Table 1
Accuracy of EMR Predictor

EMR		
Model	**RSME****	**MAPE***
Linear Regression Model	2.61	24.6
Support vector regression	1.01	9.9
Artificial Neural Network	0.47	2.2
XG Boost Model	0.27	1.6

**RSME - Root Mean Squared Error
*MAPE – Mean Absolute Percentage Error

From different models created to predict Ethanol mass ratio(EMR), XG-Boost gives the best accuracy with MAPE of 1.6% followed by artificial neural network (2.2% MAPE) and Support vector regression (9.9% MAPE).

Linear regression model does not give good accuracy since it's a linear model and efficiency variation with different variables are not linear. In Support vector regression polynomial kernel is used so it is giving more better result than linear regression. Also, support vector regression differs from linear regression models due to their epsilon-insensitive tube which captures errors based on the points outside of the tube. Artificial neural network is a deep learning technique which works on neural networks, and it is better than traditional machine learning methods because multiple hidden layer are used. XGBoost is an ensembled technique which works on gradient boosting method. It is currently the most advanced technique in machine learning. It uses multiple learner trees to predict outcome.

3.2. Accuracy of Thermal Efficiency Predictor

In the trained models, we have used same algorithms, xgboost and ANN are giving the best performance. Which is shown in Table 2-

Table 2
Accuracy of Thermal Efficiency Predictor

Thermal Efficiency		
Model	**RSME***	**MAPE***
Linear Regression Model	0.6428	0.9807
Support vector regression	0.1038	0.1713
Artificial Neural Network	0.2062	0.3653
XG Boost Model	0.0625	0.0215

*RSME - Root Mean Squared Error
*MAPE – Mean Absolute Percentage Error

From different models created to predict thermal efficiency, XG Boost and Support vector regression gives the good accuracy with MAPE less than 0.2% followed by artificial neural network (0.36% MAPE) and linear regression (0.98% MAPE). All models are giving good results with more than 99% accuracy. One point to note is that accuracy of a model is totally dependent on the data. If the data is too much scattered it won't give good results.

3.3. Maximum Minimum Efficiency at different load factors

In Table 3 maximum and minimum thermal efficiency with respective EMR is given. At low loads maximum efficiency is at low EMR. At medium load maxium efficiency is at higher EMR. Whereas at high loads maximum efficiency is near to low EMR. Vice-versa is also true for minimum efficiency.

Table 3
Min and Max Efficiency at different Loads.

Sr. No.	Load Type	Load Factor	Efficiency	EMR
1.	LOW	10%	49.8%(Max) 44.5%(Min)	0% 78%
2.	MEDIUM	30%	49.77%(Max) 49.1%(Min)	62% 0%
3.	MEDIUM	50%	50.65%(Max) 49.1%(Min)	58% 0%
4.	HIGH	70%	49.75%(Max) 48.10%(Min)	22% 49%

3.4. Variation of EMR with different properties

• Variation of thermal efficiency highly depends on the applied load. At low load, thermal efficiency decreases as EMR increases, at medium load thermal efficiency first increases then decreases, similar case is found for high load also, but maximum efficiency is found at low EMR as this variation is due to the difficulty to ignite lean mixtures and higher fuel consumption in rich mixtures as shown in Figure 4 (a).

• The variation of exhaust temperature as shown in Figure 4 (b) is mostly dependent on the applied load and Its variation with respect to EMR is negligible. At low load it slightly increases, at medium load it slightly decreases and at high load it fluctuates. This is because as the load increased, torque required also increases (at same speed) which leads to increase in more fuel consumption. When more fuel is consumed by the engine, rapid combustion occurs which leads to higher exhaust temperature.

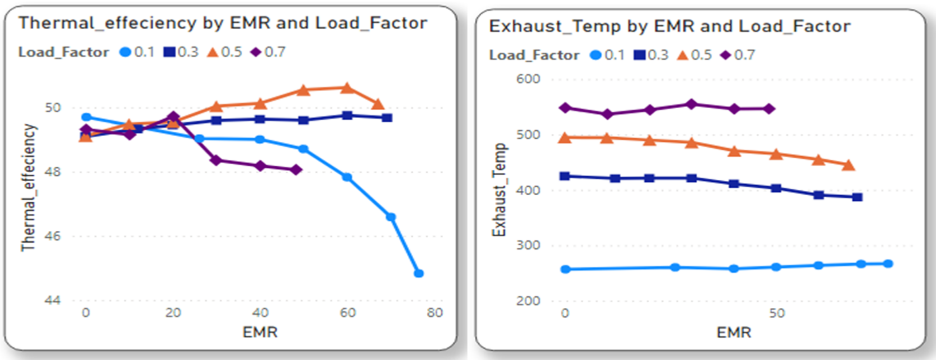


Figure 4: (a) EMR VS Thermal Efficiecnycy, (b) EMR VS Exhaust Temperature

• Variation of EMR with NOx emission is shown in Figure 5 (a). at low load, NOx emissions decrease with EMR due to the bio-chemical properties of ethanol. But at medium loads it slightly decreases then it suddenly increases with increase in EMR. At high loads NOx emissions first increase then decreases and stabilizes (slightly Increasing) after 22% of EMR.

• Variation of EMR with HC emissions is shown in Figure 5 (b). Surprisingly HC emissions increase at low loads if we increase the EMR. But as the load is increased it becomes stable and increases slightly. This is because of evaporative emission. More fuel can be trapped in the crevice volume when the EMR increases. Thus, incomplete combustion occurs due to the flame not penetrating in the crevice volume. The unburnt mixture in the crevice is then released in the exhaust stroke and is one of the major sources of unburnt HC.

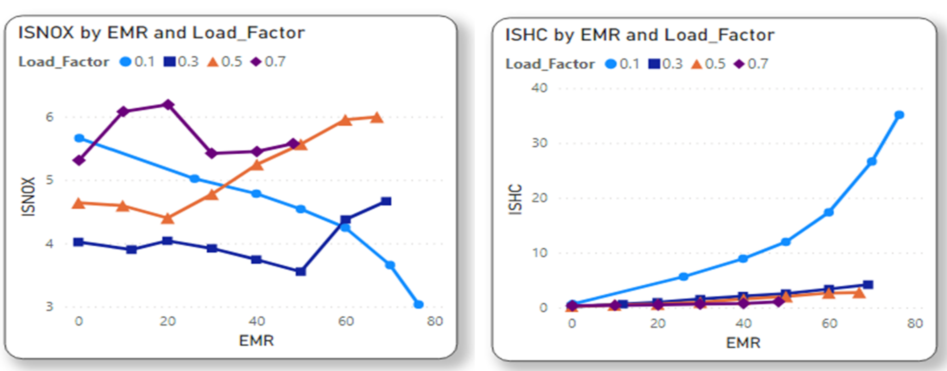


Figure 5: (a) EMR VS NOx emission, (b) EMR VS HC emission

• Variation of EMR with CO emissions is shown in Figure 6 (a). As the EMR is increased CO emissions also increases at low load. If we increase the load CO emissions decreases. At low load poor oxidation of CO to CO2 becomes a major issue. Whereas at high loads since engine temperature is high it leads to good oxidation of CO2.

• Variation of EMR with PM emissions is shown in Figure 6 (b). For all loads the initial emission of PM (at 0% EMR) is very high because it contains pure diesel. As EMR increased diesel ratio decreases. PM emissions also decreases as EMR increases because ethanol fuel contains less PM than diesel fuel [19, 20].

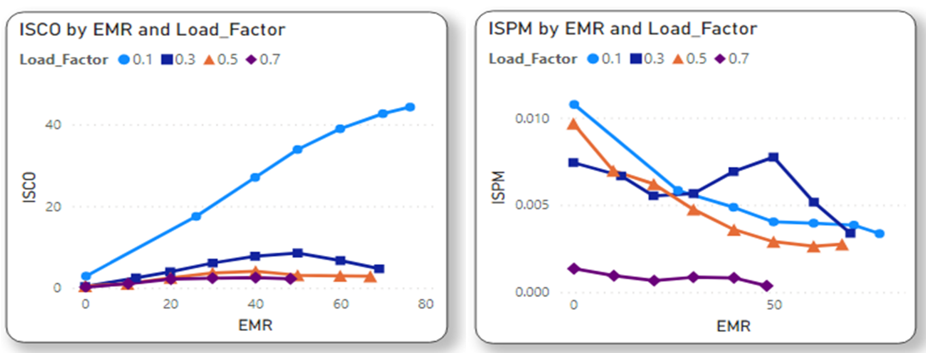


Figure 6: (a) EMR VS CO emission, (b) EMR VS PM emission

• Variation of EMR with maximum pressure rise is shown in Figure 7 (a). As the EMR increases pressure rise (PRRmax) decrease slightly at low loads. But as the load is increased the pressure rise also increases with increase in EMR.

• Variation of EMR with gross indicated mean effective pressure is shown in Figure 7 (b). The coefficient of variance of gross indicated mean effective pressure (COVofgIMEP) increases at low load as EMR is increased, but it decreases at medium and high loads if EMR is increased. At high loads the COVofgIMEP does not fluctuates resulting in less variance of gross indicated mean effective pressure.

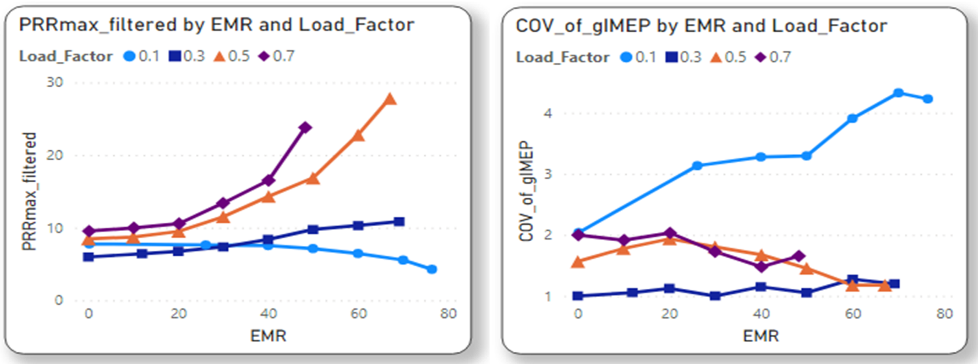


Figure 7: (a) EMR Vs PRRmax Filtered, (b) EMR Vs COV of gIMEP

• Variation of NOx emission with thermal efficiency is shown in Figure 8 (a). With increase in thermal efficiency at low load NOx emission increases due to more fuel consumption. At medium loads it is almost same. But at high loads first it decreases then increases.

• Variation of thermal efficiency emission with burn duration is shown in Figure 8 (b). As thermal efficiency increases burnup duration decreases at low load because if burn duration is less then knocking and detonation is less resulting in more fuel economy. Burnup duration is almost constant at medium loads because at medium loads fuel economy is almost constant.

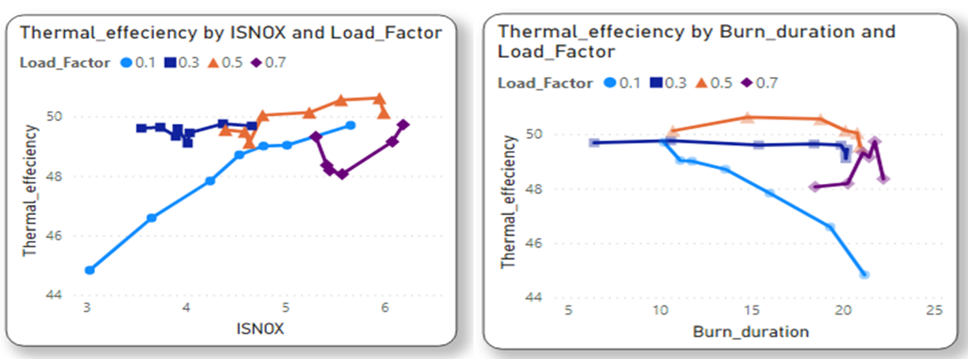


Figure 8: (a) ISNOx Vs Thermal efficiency, (b) Burn duration Vs Thermal efficiency

• Variation of thermal efficiency emission with exhaust temperature is shown in Figure 9 (a). Exhaust temperature mostly depends on the applied load. As load factor increase exhaust temperature also increases. At low load, there is a variation of thermal efficiency at the same (slightly decreasing) exhaust temperature. For medium and high loads thermal efficiency fluctuates slightly. At medium and high loads, variation of thermal efficiency with exhaust temperature is negligible.

• Variation of thermal efficiency emission with ignition delay is shown in Figure 9 (b). For ignition delay as thermal efficiency increase ignition delay decreases, this is because as ignition delay increases the chances of detonation and knocking also increases at low loads. Whereas at medium loads ignition delay is almost constant and less as compared to low load. If load is increased further ignition delay increases and reaches the optimum level, then decreases.

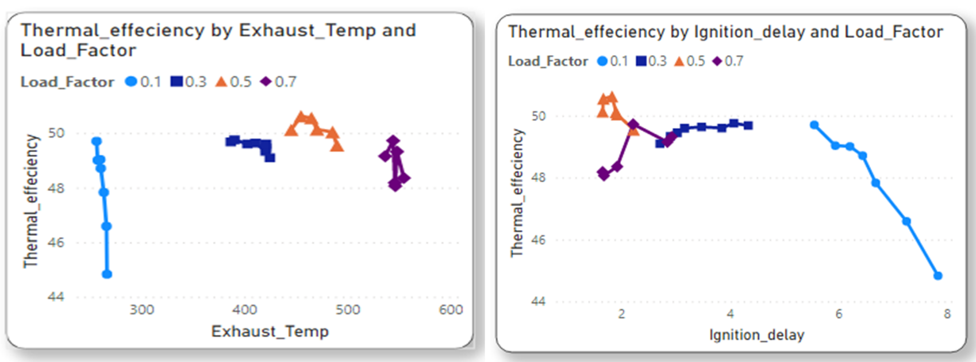


Figure 9: (a) Exhaust temp Vs Thermal efficiency, (b) Ignition delay Vs Thermal efficiency

4. Conclusion

This study is aimed at the combustion of ethanol and diesel (flex fuel) in a single cylinder of a commercially available heavy duty diesel engine. The major purpose of this work is to apply machine learning and deep learning to determine EMR and thermal efficiency. The data is prepared from a standard experiment. At where the dataset was created under four distinct loads (A10, A30, A50 and A70) ranging from low to high. Various fluctuating parameters were investigated. The results are shown below.

1. At low load as the EMR increases, thermal efficiency decreases. At low

loads, lean mixtures are difficult to ignite and flame propagation is not complete, resulting in lower thermal efficiency.

2. At medium loads, as EMR increases, the thermal efficiency increases up-to a certain limit (optimum EMR) after this point, it starts decreasing, so for 30% load - 50% load, the maximum thermal efficiency is 49.77% - 50.65% respectively. These efficiencies are achieved at 62% and 58% ethanol mass ratio respectively. This is because at medium loads, lean mixtures are difficult to ignite and flame propagation is not complete resulting in lower thermal efficiency but, rich mixtures at medium load leads to higher fuel consumption which decreases the efficiency. In between the lean and rich mixture, it gives the maximum efficiency. As the load increase maximum efficiency shifts towards lower EMR.

3. At high loads (70% load), the thermal efficiency is maximum (49.75%) at 22% EMR. Rich mixtures at high load leads to higher fuel consumption which decreases the efficiency.

4. For predicting ethanol mass ratio (EMR), linear regression model gives a mean absolute percentage error (MAPE) 24 to 25%. Whereas Support vector regression is giving a result of mean absolute percentage error (MAPE) 9 to 10%. The best models for predicting ethanol mass ratio are artificial neural networks with a mean absolute percentage error (MAPE) of 2.2 %. Whereas for the XG Boost, mean absolute percentage error (MAPE) is 1.6 %.

5. Four different models were also trained to predict the thermal efficiency of flex-fuel engine. The MAPE of the linear regression model was less than 1%, whereas for the Support vector regression, it was 0.2%. For artificial neural networks, there is a MAPE of 0.4%, and the XG Boost is also giving MAPE of 0.02%. All these models are giving good accuracy for a range of observations. The simplest model is the linear regression model and the most complicated is the artificial neural network and XG Boost.

References

[1] R. Frey, New ethanol cars in brazil in february 2022, Fuel (2020).

[2] IANS, Government monitoring scope of flexible fuel vehicles to amplify use of bio-fuels in india, Fuel (2021).

[3] times of indian, India has achieved 10% ethanol blending ahead of deadline, says pm modi; no more pollution tax from october 1, Fuel (2022).

[4] A. K. Della-Bianca, B E Gombert, Us federal sample records for fuel ethanol industry, Fuel (2013).

[5] afdc, Flex fuel vehciles model year 2022: Alternative fuel and advanced technolgy vehicles, Fuel (2022).

[6] J. Han, L. Somers, R. Cracknell, A. Joedicke, R. Wardle, V. R. R. Mohan, Experimental investigation of ethanol/diesel dual-fuel combustion in a heavy-duty diesel engine, Fuel 275 (2020) 117867.

[7] I. Antonopoulos, V. Robu, B. Couraud, D. Kirli, S. Norbu, A. Kiprakis, D. Flynn, S. Elizondo-Gonzalez, S. Wattam, Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review, Renewable and Sustainable Energy Reviews 130 (2020) 109899.

[8] C. Z. Chen, Q. Wu, Z. Y. Li, L. Xiao, Z. Y. Hu, Diagnosis of alzheimer's disease based on deeply-fused nets, Combinatorial Chemistry & High Throughput Screening 24 (6) (2021) 781–789.

[9] J. Johnson, Artificial intelligence, drone swarming and escalation risks in future warfare, The RUSI Journal 165 (2) (2020) 26–36.

[10] S. Reitmann, M. Schultz, An adaptive framework for optimization and prediction of air traffic management (sub-) systems with machine learning, Aerospace 9 (2) (2022) 77.

[11] M. Sadeghifar, M. Javidikia, V. Songmene, M. Jahazi, Finite element simulation-based predictive regression modeling and optimum solution for grain size in machining of ti6al4v alloy: Influence of tool geometry and cutting conditions, Simulation Modelling Practice and Theory 104 (2020) 102141.

[12] H. Zhang, B. He, G. Lu, Y. Zhu, A simulation and machine learning based optimization method for integrated pedestrian facilities planning and staff assignment problem in the multi-mode rail transit transfer station, Simulation Modelling Practice and Theory 115 (2022) 102449.

[13] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, A. J. Aljaaf, A systematic review on supervised and unsupervised machine learning algorithms for data science, Supervised and unsupervised learning for data science (2020) 3–21.

[14] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, K. C. Chatzisavvas, A comparison of machine learning techniques for customer churn prediction, Simulation Modelling Practice and Theory 55 (2015) 1–9.

[15] Automeris - a tool to convert graphs into excel/csv data.

[16] D. C. Montgomery, E. A. Peck, G. G. Vining, Introduction to linear regression analysis, John Wiley & Sons, 2021.

[17] M. Awad, R. Khanna, Support vector regression, in: Efficient learning machines, Springer, 2015, pp. 67–80.

[18] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, et al., Xgboost: extreme gradient boosting, R package version 0.4-2 1 (4) (2015) 1–4.

[19] L. Zhu, C. S. Cheung, W. Zhang, J. Fang, Z. Huang, Effects of ethanol–biodiesel blends and diesel oxidation catalyst (doc) on particulate and unregulated emissions, Fuel 113 (2013) 690–696.

[20] S. Sakai, D. Rothamer, Impact of ethanol blending on particulate emissions from a spark-ignition direct-injection engine, Fuel 236 (2019) 1548–1558.