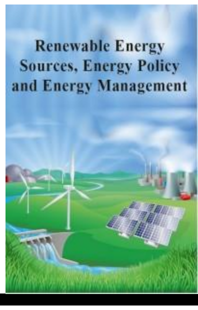
Renewable Energy Sources Energy Policy and Energy Management 2(1) [2021], pp. 1-9

Renewable Energy Sources Energy Policy and Energy Management

journal homepage: https://dergipark.org.tr/en/pub/resepem



Review Article

Support Vector Machine, Gradient Boosting and Artificial Neural Network Techniques in Internal Combustion Engine Tests: A Review

Mehmet Bilban [1*], Mehmet Selman Gökmen[1]

[1] *Necmettin Erbakan University, Seydişehir Vocational School, Seydişehir, Konya, Turkey*

ARTICLE INFO

* Corresponding author
mbilban@erbakan.edu.tr

Received 6 October 2020
Received in revised form
19 November 2020
Accepted 19 November 2020

Published by Editorial Board Members of RESEPEM

© This article is distributed by Turk Journal Park System under the CC 4.0 terms and conditions.

ABSTRACT

It is seen that researchers tend to use alternative machine learning methods in order to determine the complex relationship between engine performance data, alternative fuel blend ratios and exhaust emissions. As a result of the researches, it is observed that gradient boosting algorithm, support vector machine and artificial neural network machine learning methods are frequently used methods. Among these three methods, it is concluded that the the artificial neural network method is superior. In this study, the gradient boosting algorithm, support vector machine and artificial neural network methods are viewed. The reasons for using the artificial neural network method more than others are explained.

Keywords: artificial neural network, gradient boosting, machine learning, support vector machine, internal combustion engines

1. Introduction

It is preferred due to the use of diesel fuel in vehicles, high performance and fuel efficiency. However, many emissions are released into the atmosphere as a result of combustion. These exhaust emissions (EE) also bring environmental problems. Biodiesel is an alternative energy source for diesel engines. They can also be mixed with diesel fuels in certain proportions and filled directly into the engine without a major change in the engine [1-2].

Biodiesel is currently more expensive than diesel. A biodiesel fuel mixture ratio that can achieve less emissions while maintaining low fuel cost is desired. There is no shortcut in determining the optimum ratio of biodiesel in the mixture because it is difficult to correlate dependent parameters such as exhaust emission and fuel cost [3]. Numerous experiments are carried out on a dynamometer to determine the optimum biodiesel rate. The disadvantage of experimenting on the dynamometer is that it takes time and cost. At the same time, many sensors are needed to control the relationship between engine performance and exhaust emissions. The solution may be to create a mathematical model on biodiesel engines to eliminate all the disadvantageous problems. At the same time, the optimum biodiesel ratio can be determined by optimization methods that use the mathematical engine model. It is also known that it is very difficult to obtain a mathematical model of the engine due to complex relationships between dependent and independent parameters [3]. Data-based machine learning methods are used to reveal the complex relationships between all these parameters [4-14]. It has been observed that machine learning methods such as artificial neural network, gradient enhancement algorithm and support vector machine are frequently used by researchers to overcome all these problems. In this study, it has been a determining factor in researching the machine learning methods discussed.

An extreme gradient boost (GB) algorithm (GBA) was developed for diesel engine misfire diagnostics [15]. A support vector machine (SVM) method is proposed for online optimization problem on a marine diesel engine

[16]. Support vector machine method was used to estimate the exhaust emission and engine performance data of the diesel engine that can work with Nano Diesel blended fuels [17]. An artificial neural network (ANN) model was developed for the optimization of diesel engine operating parameters fed with a palm oil-diesel mixture [18]. ANN model was created on a diesel engine where plastic pyrolysis oil was blended with diesel and ethanol, and its usability in real-time studies was shown [19]. Canakci et al. [20] investigated the use of ANN for exhaust emission and biodiesel engine performance prediction process. Yusaf et all. [21] used ANN to estimate engine performance data with fuel generated from a mixture of palm oil and diesel. Also in the literature, various ANN methods have been developed to estimate different fuel mixtures and emission characteristics [22-24]. In this study, a research study was carried out on the engine performance of GBA, SVM and ANN machine learning methods, optimization of fuels consisting of diesel and biodiesel fuel mixtures, and also the usability of exhaust emission values in estimation processes.

2. Data Estimation Process with Machine Learning Methods

Machine learning techniques make inferences from data in order to make automatic decisions, obtain information and data reduction. These techniques also include algorithms. Machine learning methods require a comprehensive data set to predict relationships between data. Machine learning techniques are divided into three subgroups as unsupervised machine learning (training with unclassified or unlabeled data), supervised machine learning (training with classified or labeled data) and reinforced machine learning (includes reward or experience-based methods) as shown figure 1 [25] .

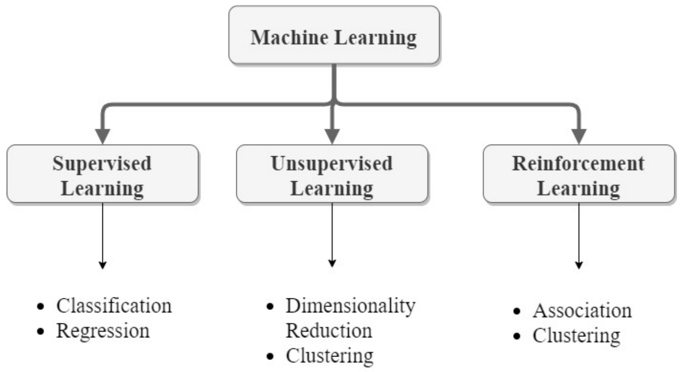


Figure 1. Classification of machine learning approaches

In a system using supervised learning, machine learning algorithms, inputs and corresponding outputs are given and system tries to establish a relationship between input-output data. The purpose of the established relationship is to predict the outputs of unused data. Supervised learning methods can be used to classify data and perform regression processes, depending on whether the data is quantitative or qualitative. Machine learning methods that apply qualitative data are used for classification processes. Machine learning methods for quantitative data perform regression operations [26].

In unsupervised learning, system tries to discover whether there is a possible relationship in this data, when a data set is given to the system without input-output information. The main purpose of unsupervised learning is to perform data analysis by dividing data into sections and clustering operations.

In reinforcement learning, system interacts with an environment and performs learning according to the outcome of the actions, i.e. it refers to the learning process based on experience In supervised learning, performing the target output estimation process, it becomes more difficult to evaluate the accuracy of the structure (segmentation and cluster) created for a specific situation in unsupervised learning [27].

Machine learning methods consist of data collection, feature engineering, model building, validation and estimation of final target outputs [28]. It is the presentation of the data obtained after data collection, experimental studies and calculation processes in the desired format after data optimization processes. Feature engineering is the process of selecting suitable features for the predicted target output. Feature engineering is very important for machine learning methods because it is very effective in determining the performance of algorithms. Model creation includes the selection of algorithms that best fit the training data set and test data sets. Verification is the process of testing the trained machine learning method on a certain amount of separated data and observing the results. In general, machine learning methods consist of training of data and testing process steps. At the same time, it refers to the process steps in which some criteria are determined for the model to be verified and the results are observed. Machine learning basic working principle structure is shown in Figure 2.

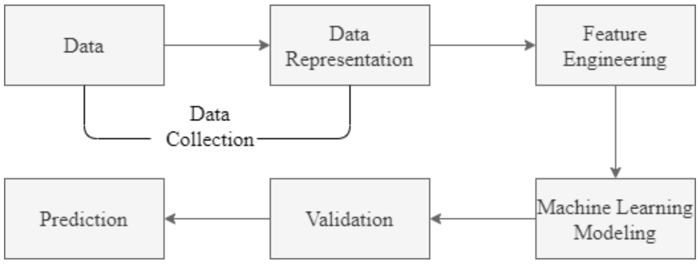


Figure 2. Basic working principle structure of machine learning

One of the important points to consider when creating a machine learning model is the complexity of the model. Learning step of a complex model can cause an overfitting problem. Overfitting means that the resulting model fits the training data very well, but has poor generalization ability over test and other data groups. Another point to consider is the process of training the model from noisy and incorrect data. Therefore, it is not desirable to create a model that fits the points corresponding to each data in the data set. The most important criterion when creating the model is to obtain a model capable of generalization. The complexity of the model is determined according to the relationship between the size of the training data and model parameters. For this reason, while creating machine learning models, simple models are created for small training data sets while larger data sets are needed to prevent overfitting problem in complex and higher dimensional data sets [27]. Hyper-parameter selection is one of the important issues when creating a machine learning model. Hyper-parameter selection varies depending to the machine learning algorithm. At the same time, parameter selection is made depending on the complexity of the data and the experience of the researcher. Therefore, the data specialist should consider a linear model or a polynomial model and other data inherent features of the data before selecting the appropriate machine learning model for the data set. It does not seem possible to talk about a single method for hyper-parameter optimization, but it has been seen that the hyper-parameter automatic selection process is among the current research topics in machine learning [29-31].

3. Machine Learning Methods

3.1 Gradient Boosting Algorithm

Gradient boosting algorithm is known as one of the decision tree communities [32]. GBA is used for classification and regression problems. A repetitive training process is applied on decision trees in order to minimize the lost function value. The existing training data set is used to predict the label of each data. To put more emphasis on training data with poor results, the dataset is relabeled and errors are eliminated in the prediction process. GBA is used on datasets where high predictive results are needed [33]. The GBA is shown in table 1.

Table 1. Gradient Boosting Algorithm

Gradient Boosting Algorithm

Input: $\{(x_i, y_i)\}_{i=1}^{n}$ is training data set, $L(y, F(x))$ is loss function and number of iterations is defined as M.

1. With a fixed value, the initial model is started:

$$\boldsymbol{F_0(x) = \underset{\gamma}{arg\,min} \sum_{i=1}^{N} L(y_i, \gamma_i)}$$

2. **For m=1 to M**:

$$\boldsymbol{r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} for\ i = 1, 2, 3 \ldots\ldots, n}$$

$h_m(x)$ is model prediction data and $\{(x_i, r_{im})\}_{i=1}^{n}$ is the data set for the training process:

$$\boldsymbol{\gamma = \underset{\gamma}{arg\,min} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))}$$

$$\boldsymbol{F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)}$$

3. Output $\boldsymbol{F_m(x)}$

Since GBA shows high performance in nonlinear problems, its use in multi-parameter prediction problems is increasing day by day [34-36].

3.2 Support Vector Machine

SVR was first introduced by Vapnik for quadratic optimization problems [37]. The support vector machine algorithm sets nonlinear boundaries between classes by transforming the data into a high dimensional area. It is the task of the kernel function to be able to separate the input data linearly. SVM creates a hyperplane that provides maximum separation between output classes. The training observations closest to the Hyperplane are called support vectors [38]. The kernel function is defined as consisting of $K(x_i, x_j)$pairs. It is represented by the Lagrangian factor a for each training observation. If the value of a in the training data set is not zero, these values represent support vectors. The w value represents the weight vectors and the b value represents the bias value. Consequently, the SVM classifier is defined as in equation 1 [38].

$$\mathrm{sgn}(\sum_{i}^{N} a_i y_i K(x_i, x_j) + b) \qquad \textbf{Eq.1}$$

The reasons for using the SVM method are; good generalization ability [39], transparency [40], high computational speed [41] and quadratic programming [42].

3.3 Artificial Neural Network

ANN, one of the application areas of artificial intelligence, is an effective tool used to model complex systems. ANNs provide multiple advantages such as faster statistical training with the ability to detect nonlinear relationships between independent and dependent variables. The ANN structure consists of an input layer, exit layer and one or more hidden layers.

ANN is the representation of the human nervous system consisting of neurons that communicate with each other through connections called axons. Neurons consist of one or more connections. Each link is assigned a weight value. The output connection is responsible for carrying the incoming signals to other neurons. Signals are collected according to the activation threshold level of each neuron. The output signal is determined by passing all the collected signals through an activation function. As shown equation 2, X is an input vector and y is the perceptron producing a single output [38].

$$y = \Phi(\sum_{i=1}^{n} w_i x_i + \mathrm{b}) \qquad \textbf{Eq.2}$$

w input weights vector, b bias value and Φ represents the non-linear activation function. The activation function is used to solve nonlinear problems with ANN. The activation function ensures that the output signal is not a linear function [6]. The three most commonly used activation functions in the literature are seen as hyperbolic tangent, ReLU, and sigmoid. l number of entries, m number of hidden units and let n represent the number of output units. For each hidden layer, the formulation is defined as in equation 3 [38].

$$h_j = f(\sum_{i \to j} w_{ij}^{(1)} x_i + w_{0j}^{(1)}) \qquad \textbf{Eq.3}$$

The activation function used in hidden layers is expressed as f, it is defined as $i = \{1, 2, 3, 4 \ldots\ldots l\}$ and $j = \{1, 2, 3, 4, \ldots\ldots m\}$. Output unit value is defined as in equation 4 [38].

$$y_k = g(\sum_{j \to k} w_{jk}^{(2)} h_i + w_{0k}^{(2)}) \qquad \textbf{Eq.4}$$

g represents another activation function and $k = \{1, 2, 3, 4 \ldots\ldots n\}$. Output value varies according to different activation functions and weights. When the values obtained in the output layer are compared with the actual values, deviations/errors may occur. In order to reduce the error in the output layer, the weights, bias value and parameters must be adjusted. Error value is as given in equation 5 [38].

$$\Delta w = e_k * x_{jk} * \alpha \qquad \textbf{Eq.5}$$

e_k represents the error of the output unit, the x_{jk} and k values are the input values that cause the error, and the α value is defined as the learning rate that determines how much the weights need to be changed to correct the error. After the error is calculated, the change of weights in the hidden layer is calculated as in equation 6 [38].

$$e_j = w_j * e_k * d(y_k) \qquad \textbf{Eq.6}$$

d stands for derivation. In an ANN network structure, forward and reverse calculations continue until the error is reduced under a limit value [38]. The learning process of the network is carried out by repeatedly adjusting the weights. The generalization capability of the ANN model on the network is related to the selection of parameters and a correct architecture. Since there is no specific method for optimum network architecture, it varies according to the data set and the experience of the data scientist. While small networks may not perform learning process in the created architecture, complex and large networks may cause over-fitting problems with poor generalization ability. In the process of creating and testing an obtained ANN model, the existing dataset is randomly separated at certain rates.

3.4 Literature Review

Many publications in the literature related to the three mentioned methods on internal combustion engines are evaluated and given in Table 2. It is observed that machine learning methods such as ANN, SVM and GB are used to estimate the performance parameters of internal combustion engines, additive fuel mixtures and exhaust emission values. It has been observed that the ANN method is one of the most frequently used machine learning methods in problems that are difficult to model mathematically and in nonlinear regression calculations. It has been observed that the SVM method is used due to its advantage in solving regression problems with small data sets in recent studies. The GB method is preferred to combine nonparametric and decision tree-based regression processes to improve forecast performance.

Table 2. ANN, SVM and GB methods and model development parameters

Year	Research Field	Model and Data Workflow	Description	Ref.
2020	ICE&EE	**Model:** ANN **Input parameters:** Engine load and the percentage of biodiesel in the blends input parameters. **Output parameter:** performance indices BSFC, BTE and exhaust emissions like NOx, CO, and HC **Activation Function:** tansig and linear function **Validation:** R^2 and MSE	An ANN model has been developed to predict engine performance and EE values in an ICE using palm oil biodiesel.	[43]
2020	ICE&EE	**Model:** ANN **Input parameters:** It consists of Engine speed, Fuel Type, Fuel Consumption data. **Output parameter:** Engine Torque, Engine Power, Nox Emission values. **Activation Function:** tansig and linear function **Validation:** R^2 and MAE	It has been shown that in an ICE using a soybean biodiesel fuel mixture, the engine torque, engine power and NOx emission data can be estimated with ANN without the need for long-term experiments.	[44]
2019	ICE	**Model:** ANN **Input parameters:** Input parameters consist of internal engine data consisting of 109 parameters in total. **Output parameter:** indicated mean effective pressure **Activation Function:** hyperbolic tangent **Validation:** R^2 and MAE	The ability of an engine to predict indicated mean effective pressure (IMEP) using neural networks has been investigated.	[45]
2018	ICE & EE	**Model:** ANN **Input parameters:** It consists of 4 parameters in total (Engine speed, Throttle angle, exhaust temp., air flowrate) **Output parameter:** heat energy. **Activation Function:** hyperbolic tangent	It has been suggested to use ANN in order to increase the efficiency performance significantly by ensuring the recovery of heat energy from the engine.	[46]
2017	ICE	**Model:** ANN **Input parameters:** Engine map data consisting of 5 parameters: engine speed, torque, intake manifold pressure, turbine inlet pressure and intake manifold temperature were used. **Output parameter:** Volumetric efficiency **Activation Function:** Logistic function **Validation:** R^2 and MAPE	The application of the ANN method to estimate the volumetric efficiency in a diesel engine has been studied.	[47]
2020	ICE	**Model:** SVM Regression **Independent variables:** Three operating and calibration parameters were selected as independent variables, including manifold absolute pressure, fuel equivalent ratio and ignition advance angle. **Dependent variables:** torque, brake specific fuel consumption and specific NOx emission. **Kernel Function:** radial basis, kernel function denoted by q and q=0.01 **Validation:** R^2 and MAPE	The effect of HCNG mixtures with various hydrogen mixing ratios on combustion and emission properties on a 6-cylinder ICE was investigated in depth using the SVM method.	[48]
2020	ICE	**Model:** SVM **Input parameters:** ignition timing, injection timing, hydrogen volume ratio and speed variables are input parameters. **Output parameter:** Fuel conversion efficiency is expressed as power, torque and specific fuel consumption and CO. **Kernel Function:** radial basis function **(**RBF) **Validation:** R^2, MAPE, RMSE	SVM method was used to estimate engine performance data such as fuel conversion efficiency, power, torque and specific fuel consumption and CO.	[41]

2019	ICE & HCNG	**Model:** SVM Regression **Independent variables:** hydrogen blend ratio, the fuel–air equivalence ratio and the RMS voltage of the electric field applied **Dependent variables:** mean flame propagation speed and the peak combustion pressure **Kernel Function:** radial basis function **(**RBF) **Validation:** R^2 and MAPE	SVM method was investigated for the prediction of flame propagation and combustion properties under Hydrogen enriched compressed natural gas (HCNG) electric fields.	[49]
2018	ICE & EE	**Model:** SVM Regression **Independent variables:** Engine speed, manifold absolute pressure, fuel equivalent ratio, ignition advance angle **Dependent variables:** NO_x emission **Kernel Function:** radial basis function **(**RBF) **Validation:** R^2 and MAPE	A study was conducted on the factors affecting the prediction accuracy of the SVM model for the NOx emission of a hydrogen enriched compressed natural gas engine.	[50]
2020	EE	**Model:** GB Regression **Input variables:** It consists of speed, acceleration, road grade, passenger load, speed and acceleration data within 10 seconds. **Output variables:** CO_2 emission **Function:** loss (a negative binomial log-likelihood) **Validation:** MAPE, MAE, RMSE, and R2	It has been concluded that the GB regression method is the most suitable method to estimate the results in order to estimate the CO_2 emission rates in buses using different fuel types.	[51]
2019	EE	**Model:** GB Regression **Input variables:** CO, CO_2, HC, NO_x **Output variables:** liquefied natural gas (LNG) emission **Function:** loss (a negative binomial log-likelihood) **Validation:** MAPE, MAE, RMSE	The GB regression method was used to analyze and estimate the emission rates for the bus using LNG.	[52]

4. Conclusions

As a result of the literature researches, it is seen that machine learning methods are used that can realize the relationships between the engine performance values and exhaust emission values of diesel and biodiesel-added fuel blends in real time. As can be seen from the studies, machine learning methods have been used as an alternative to traditional methods in order to cope with test processes and cost problems. It has been observed that machine learning methods such as GBA, SVM and ANN are frequently used. Many researchers tend to use black box (generated by algorithm with data) modeling techniques such as ANN and SVM more in scientific studies. These methods find relationships between variables on experimental data with acceptable precision and without the need for statistical assumptions. In a study, it was stated that SVM method is a more generalized and acceptable model than ANN method for marine diesel engine performance data [53]. It has been observed that the ANN is used more than other machine learning methods. Because it has ability to better model the complex relationships in ICE data. In order to explain the relationship between many engine, fuel and exhaust emission parameters to be discussed in the future, it has been observed that the interest in machine learning methods discussed in this study and other machine learning methods will increase.

5. References

[1] M. S. Grabosk and R. L. McCormick, "Combustion Of Fat And Vegetable Oil Derived Fuels in Diesel Engines," *Science (80-.).*, vol. 24, no. 97, pp. 125–164, 1998.

[2] A. S. Ramadhas, S. Jayaraj, and C. Muraleedharan, "Use of vegetable oils as I.C. engine fuels - A review," *Renew. Energy*, vol. 29, no. 5, pp. 727–742, 2004, doi: 10.1016/j.renene.2003.09.008.

[3] K. I. Wong, P. K. Wong, C. S. Cheung, and C. M. Vong, "Modeling and optimization of biodiesel engine performance using advanced machine learning methods," *Energy*, vol. 55, no. x, pp. 519–528, 2013, doi: 10.1016/j.energy.2013.03.057.

[4] J. Grahovac, A. Jokić, J. Dodić, D. Vučurović, and S. Dodić, "Modelling and prediction of bioethanol production from intermediates and byproduct of sugar beet processing using neural networks," *Renew. Energy*, vol. 85, pp. 953–958, 2016, doi: 10.1016/j.renene.2015.07.054.

[5] V. Cocco Mariani, S. Hennings Och, L. dos Santos Coelho, and E. Domingues, "Pressure prediction of a spark ignition single cylinder engine using optimized extreme learning machine models," *Appl. Energy*, vol. 249, no. February, pp. 204–221, 2019, doi: 10.1016/j.apenergy.2019.04.126.

[6] A. Domínguez-Sáez, G. A. Rattá, and C. C. Barrios, "Prediction of exhaust emission in transient

conditions of a diesel engine fueled with animal fat using Artificial Neural Network and Symbolic Regression," *Energy*, vol. 149, pp. 675–683, 2018, doi: 10.1016/j.energy.2018.02.080.

[7] B. Liu, J. Hu, F. Yan, R. F. Turkson, and F. Lin, "A novel optimal support vector machine ensemble model for NO X emissions prediction of a diesel engine," *Meas. J. Int. Meas. Confed.*, vol. 92, no. X, pp. 183–192, 2016, doi: 10.1016/j.measurement.2016.06.015.

[8] S. Roy, R. Banerjee, A. K. Das, and P. K. Bose, "Development of an ANN based system identification tool to estimate the performance-emission characteristics of a CRDI assisted CNG dual fuel diesel engine," *J. Nat. Gas Sci. Eng.*, vol. 21, no. x, pp. 147–158, 2014, doi: 10.1016/j.jngse.2014.08.002.

[9] H. Almér, "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles," *Master Sci. Thesis*, p. 75, 2015, doi: 10.1016/j.apcata.2012.08.042.

[10] Y. Pan, S. Chen, F. Qiao, S. V. Ukkusuri, and K. Tang, "Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees," *Sci. Total Environ.*, vol. 660, pp. 741–750, 2019, doi: 10.1016/j.scitotenv.2019.01.054.

[11] H. Taghavifar, S. Khalilarya, and S. Jafarmadar, "Diesel engine spray characteristics prediction with hybridized artificial neural network optimized by genetic algorithm," *Energy*, vol. 71, pp. 656–664, 2014, doi: 10.1016/j.energy.2014.05.006.

[12] J. Rezaei, M. Shahbakhti, B. Bahri, and A. A. Aziz, "Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks," *Appl. Energy*, vol. 138, pp. 460–473, 2015, doi: 10.1016/j.apenergy.2014.10.088.

[13] S. O. Giwa, S. O. Adekomaya, K. O. Adama, and M. O. Mukaila, "Prediction of selected biodiesel fuel properties using artificial neural network," *Front. Energy*, vol. 9, no. 4, pp. 433–445, 2015, doi: 10.1007/s11708-015-0383-5.

[14] A. Banerjee, D. Varshney, S. Kumar, P. Chaudhary, and V. K. Gupta, "Biodiesel production from castor oil: ANN modeling and kinetic parameter estimation," *Int. J. Ind. Chem.*, vol. 8, no. 3, pp. 253–262, 2017, doi: 10.1007/s40090-017-0122-3.

[15] J. Tao, C. Qin, W. Li, and C. Liu, "Intelligent fault diagnosis of diesel engines via extreme gradient boosting and high-accuracy time–frequency information of vibration signals," *Sensors (Switzerland)*, vol. 19, no. 15, 2019, doi: 10.3390/s19153280.

[16] X. Niu, H. Wang, S. Hu, C. Yang, and Y. Wang, "Multi-objective online optimization of a marine diesel engine using NSGA-II coupled with enhancing trained support vector machine," *Appl. Therm. Eng.*, vol. 137, no. April 2017, pp. 218–227, 2018, doi: 10.1016/j.applthermaleng.2018.03.080.

[17] M. Ghanbari, G. Najafi, B. Ghobadian, R. Mamat, M. M. Noor, and A. Moosavian, "Support Vector Machine to predict diesel engine performance and emission parameters fueled with nano-particles additive to diesel fuel," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 100, no. 1, 2015, doi: 10.1088/1757-899X/100/1/012069.

[18] S. Uslu, "Optimization of diesel engine operating parameters fueled with palm oil-diesel blend: Comparative evaluation between response surface methodology (RSM) and artificial neural network (ANN)," *Fuel*, vol. 276, no. January, p. 117990, 2020, doi: 10.1016/j.fuel.2020.117990.

[19] S. V. Khandal, S. B. Gadwal, V. A. Raikar, T. M. Yunus Khan, and I. A. Badruddin, "An experimental-based artificial neural network performance study of common rail direct injection engine run on plastic pyrolysis oil," *Int. J. Sustain. Eng.*, vol. 00, no. 00, pp. 1–10, 2020, doi: 10.1080/19397038.2020.1773568.

[20] M. Canakci, A. Erdil, and E. Arcaklioğlu, "Performance and exhaust emissions of a biodiesel engine," *Appl. Energy*, vol. 83, no. 6, pp. 594–605, 2006, doi: 10.1016/j.apenergy.2005.05.003.

[21] T. F. Yusaf, B. F. Yousif, and M. M. Elawad, "Crude palm oil fuel for diesel-engines: Experimental and ANN simulation approaches," *Energy*, vol. 36, no. 8, pp. 4871–4878, 2011, doi: 10.1016/j.energy.2011.05.032.

[22] T. Ozgur, G. Tuccar, M. Ozcanli, and K. Aydin, "Prediction of emissions of a diesel engine fueled with soybean biodiesel using artificial neural networks," *Energy Educ. Sci. Technol. Part A Energy Sci. Res.*, vol. 27, no. 2, pp. 301–312, 2011.

[23] Shivakumar, P. Srinivasa Pai, and B. R. Shrinivasa Rao, "Artificial Neural Network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings," *Appl. Energy*, vol. 88, no. 7, pp. 2344–2354, 2011, doi: 10.1016/j.apenergy.2010.12.030.

[24] H. Mohamed Ismail, H. K. Ng, C. W. Queck, and S. Gan, "Artificial neural networks modelling of engine-out responses for a light-duty diesel engine fuelled with biodiesel blends," *Appl. Energy*, vol. 92, pp. 769–777, 2012, doi: 10.1016/j.apenergy.2011.08.027.

[25] A. U. Osarogiagbon, F. Khan, R. Venkatesan, and P. Gillard, "Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations," *Process Saf. Environ. Prot.*, 2020, doi:

10.1016/j.colsurfa.2020.124658.

[26] R. James, G., Witten, D., Hastie, T., Tibshirani, *An Introduction to Statistical Learning -with Applications in R*. 2013.

[27] C. Gambella, B. Ghaddar, and J. Naoum-Sawaya, "Optimization Models for Machine Learning: A Survey," no. xxxx, 2019, doi: 10.1016/j.ejor.2020.08.045.

[28] J. Wei, X. Chu, X. Sun, K. Xu, H. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," *InfoMat*, vol. 1, no. 3, pp. 338–358, 2019, doi: 10.1002/inf2.12028.

[29] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[30] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, pp. 1–21, 2019.

[31] M. Wistuba, A. Rawat, and T. Pedapati, "A Survey on Neural Architecture Search," 2019, [Online]. Available: http://arxiv.org/abs/1905.01392.

[32] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," *Encycl. Syst. Biol.*, pp. 358–360, 2008, doi: 10.1007/978-1-4419-9863-7_941.

[33] G. P. Gupta and M. Kulariya, "A Framework for Fast and Efficient Cyber Security Network Intrusion Detection Using Apache Spark," *Procedia Comput. Sci.*, vol. 93, no. September, pp. 824–831, 2016, doi: 10.1016/j.procs.2016.07.238.

[34] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *J. Anim. Ecol.*, vol. 77, no. 4, pp. 802–813, 2008, doi: 10.1111/j.1365-2656.2008.01390.x.

[35] R. M. Yang, G. L. Zhang, F. Liu, Y. Y. Lu, F. Yang, F. Yang, M. Yang, Y. G. Zhao, and D. C. Li, "Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem," *Ecol. Indic.*, vol. 60, pp. 870–878, 2016, doi: 10.1016/j.ecolind.2015.08.036.

[36] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Sol. Energy*, vol. 150, pp. 423–436, 2017, doi: 10.1016/j.solener.2017.04.066.

[37] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Adv. Neural Inf. Process. Syst.*, pp. 281–287, 1997.

[38] P. Golbayani, I. Florescu, and R. Chatterjee, "A comparative study of forecasting Corporate Credit Ratings using Neural Networks, Support Vector Machines, and Decision Trees," *North Am. J. Econ. Financ.*, vol. 54, no. August 2019, p. 101251, 2020, doi: 10.1016/j.najef.2020.101251.

[39] W. Li and Z. Liu, "A method of SVM with normalization in intrusion detection," *Procedia Environ. Sci.*, vol. 11, no. PART A, pp. 256–262, 2011, doi: 10.1016/j.proenv.2011.12.040.

[40] N. K. Natt, H. Kaur, and G. P. S. Raghava, "Prediction of transmembrane regions of β-barrel proteins using ANN- and SVM-based methods," *Proteins Struct. Funct. Genet.*, vol. 56, no. 1, pp. 11–18, 2004, doi: 10.1002/prot.20092.

[41] J. Zareei and A. Rohani, "Optimization and study of performance parameters in an engine fueled with hydrogen," *Int. J. Hydrogen Energy*, vol. 45, no. 1, pp. 322–336, 2020, doi: 10.1016/j.ijhydene.2019.10.250.

[42] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support Vector Machine-Based Expert System for Reliable Heartbeat Recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, 2004, doi: 10.1109/TBME.2004.824138.

[43] S. Raghuvaran, B. Ashok, B. Veluchamy, and N. Ganesh, "Evaluation of performance and exhaust emission of C.I diesel engine fuel with palm oil biodiesel using an artificial neural network," *Mater. Today Proc.*, no. xxxx, 2020, doi: 10.1016/j.matpr.2020.06.344.

[44] R. Kenanoğlu, M. K. Baltacıoğlu, M. H. Demir, and M. Erkınay Özdemir, "Performance & emission analysis of HHO enriched dual-fuelled diesel engine with artificial neural network prediction approaches," *Int. J. Hydrogen Energy*, vol. 45, no. 49, pp. 26357–26369, 2020, doi: 10.1016/j.ijhydene.2020.02.108.

[45] A. Di Mauro, H. Chen, and V. Sick, "Neural network prediction of cycle-to-cycle power variability in a spark-ignited internal combustion engine," *Proc. Combust. Inst.*, vol. 37, no. 4, pp. 4937–4944, 2019, doi: 10.1016/j.proci.2018.08.058.

[46] S. G. Herawan, K. Talib, and A. Putra, "Prediction of heat energy from the naturally aspirated internal combustion engine exhaust gas using artificial neural network," *Procedia Comput. Sci.*, vol. 135, pp. 267–274, 2018, doi: 10.1016/j.procs.2018.08.174.

[47] J. M. Luján, H. Climent, L. M. García-Cuevas, and A. Moratal, "Volumetric efficiency modelling of internal combustion engines based on a novel adaptive learning algorithm of artificial neural networks," *Appl. Therm. Eng.*, vol. 123, pp. 625–634, 2017, doi: 10.1016/j.applthermaleng.2017.05.087.

[48] D. Hao, R. K. Mehra, S. Luo, Z. Nie, X. Ren, and M. Fanhua, "Experimental study of hydrogen enriched

compressed natural gas (HCNG) engine and application of support vector machine (SVM) on prediction of engine performance at specific condition," *Int. J. Hydrogen Energy*, vol. 45, no. 8, pp. 5309–5325, 2020, doi: 10.1016/j.ijhydene.2019.04.039.

[49] H. Duan, Z. Li, B. Wang, R. K. Mehra, S. Luo, C. Xu, B. Sun, X. Wang, and F. Ma, "Experimental study of premixed hydrogen enriched natural gas under an alternating-current (AC) electric field and application of support vector machine (SVM) on electric field assisted combustion," *Fuel*, vol. 258, no. June, p. 115934, 2019, doi: 10.1016/j.fuel.2019.115934.

[50] H. Duan, Y. Huang, R. K. Mehra, P. Song, and F. Ma, "Study on influencing factors of prediction accuracy of support vector machine (SVM) model for NOx emission of a hydrogen enriched compressed natural gas engine," *Fuel*, vol. 234, no. June, pp. 954–964, 2018, doi: 10.1016/j.fuel.2018.07.009.

[51] Y. Pan, F. Qiao, K. Tang, S. Chen, and S. V. Ukkusuri, "Understanding and estimating the carbon dioxide emissions for urban buses at different road locations: A comparison between new-energy buses and conventional diesel buses," *Sci. Total Environ.*, vol. 703, p. 135533, 2020, doi: 10.1016/j.scitotenv.2019.135533.

[52] Y. Pan, S. Chen, F. Qiao, S. V. Ukkusuri, and K. Tang, "Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees," *Sci. Total Environ.*, vol. 660, pp. 741–750, 2019, doi: 10.1016/j.scitotenv.2019.01.054.

[53] X. Niu, C. Yang, H. Wang, and Y. Wang, "Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine," *Appl. Therm. Eng.*, vol. 111, pp. 1353–1364, 2017, doi: 10.1016/j.applthermaleng.2016.10.042.