Review *Derleme*

DOI: 10.5281/zenodo.16415720

Applications of Machine Learning Algorithms to Internal Combustion Engine Studies

Makine Öğrenmesi Algoritmalarının İçten Yanmalı Motor Çalışmalarına Uygulanması

İlhan Volkan ÖNER[1, 2, 3]

[1]Atatürk University, Department of Mechanical Engineering, Erzurum, Türkiye

[2]Volkan Engineering, AtaTeknokent, Erzurum, Türkiye

[3]Institute of Science, Nano-Science and Nano-Engineering, Atatürk University, Erzurum, Türkiye

ABSTRACT

Petroleum (diesel, gasoline) reserves are depleting as energy demand rises, and the quest for conventional fuels is growing daily. Internal combustion engine research is crucial because of this. Current research on internal combustion engines is expensive, both in terms of setting up the experiment and in terms of the fuels that are employed and consumed. Because of this, machine learning techniques have been used in recent years to estimate engine performance and exhaust emissions. As a result, less time and material are utilized, and high accuracy in estimating the engine's performance and fuel-related exhaust emissions is attained. Machine learning algorithms will be discussed in this study first, followed by an assessment of recent research and findings in the literature.

Keywords: Machine Learning, IC Engines, Alternative Fuels, Performance, Exhaust Emissions

ÖZ

Petrol (dizel, benzin) rezervleri enerji talebi arttıkça tükenmekte ve konvansiyonel yakıt arayışı her geçen gün artmaktadır. Bu nedenle, İçten yanmalı motor araştırmaları çok önemlidir. Mevcut içten yanmalı motor araştırmaları hem deneyin kurulumu hem de kullanılan ve tüketilen yakıtlar açısından pahalıdır. Bu olumsuzluklar sebebiyle, son yıllarda motor performansını ve egzoz emisyonlarını tahmin etmek için makine öğrenimi teknikleri kullanılmıştır. Bu sayede, daha az zaman ve malzeme kullanılır ve motorun performansı ile yakıtla ilgili egzoz emisyonlarının tahmininde yüksek doğruluk sağlanır. Bu çalışmada, öncelikle makine öğrenimi algoritmaları ele alınacak, ardından literatürdeki son araştırmalar ve bulgular değerlendirilecektir.

Anahtar Kelimler: Makine Öğrenmesi, İçten Yanmalı Motorlar, Alternatif Yakıtlar, Performans, Egzoz Emisyonları.

Received/Geliş Tarihi : 23.06.2025
Revision/Revizyon Tarihi: 10.07.2025
Accepted/Kabul Tarihi : 10.07.2025
Publication Date/Yayın Tarihi : 31.07.2025

Corresponding Author/Sorumlu Yazar:
E-mail: ivoner@atauni.edu.tr

Cite this article: Öner İ.V. (2025). Applications of Machine Learning Algorithms to Internal Combustion Engine Studies. *Journal of Energy Trends*, 2(1) 28-34

Introduction

Emissions resulting from the combustion of fossil fuels have been a significant global environmental issue for the past few decades. In line with increasing emission regulations and sustainability goals, the search for alternative fuels and technologies in internal combustion engines has gained significant importance. In this context, the use of clean energy sources such as hydrogen as auxiliary fuel in diesel engines has become a noteworthy method for reducing emissions. However, due to the complex structure of hybrid fuel engine systems, accurately predicting performance and emission behaviors requires big data analytics and advanced modelling techniques (Sugumaran, et al., 2023). Computational models called machine learning (ML) methods are used to extract relationships and patterns from data. Machine learning allows systems to automatically improve their performance by evaluating data and recognizing patterns without being explicitly coded for each unique task, in contrast to traditional programming, which requires explicit instructions to solve a problem. In recent years, artificial intelligence (AI) and machine learning (ML) have expanded quickly in the context of data analysis and computing, which usually enables the applications to operate intelligently (Sarker, et al., 2021; Sarker, 2021).

The effectiveness and efficiency of a machine learning solution depend on the nature and characteristics of the data as well as the capabilities of the learning algorithms. Machine-learning techniques can be applied to feature engineering and dimensionality reduction, regression, data clustering, classification analysis, association rule learning, or reinforcement learning to effectively build data-driven systems (Sarker, 2021; Han, et al., 2022; Witten, & Frank, 2002).

Traditional physics-based methods remain limited in accurately modelling these systems and in real-time control applications. Therefore, in recent years, machine learning (ML) techniques have emerged as a strong alternative to overcome these challenges. Today, the modelling, diagnosis, optimization, and control of internal combustion engines (ICE) in line with performance, emission, and efficiency targets have become increasingly complex (Aliramezani, et al., 2022).

The complex nature of the combustion process in internal combustion engines (ICE) highlights some of the limitations of traditional physics-based modelling. In recent years, machine learning (ML) algorithms, which fall under data-driven approaches, have garnered significant interest in order to overcome these limitations and predict engine behaviour with high accuracy. In this review study, the integration of machine learning into different types of internal combustion engines, the prominent methods, and the scope of the achieved successes are evaluated.

Materials and Methods

The Working Principle of Machine Learning Systems

At the core of machine learning systems lies data. Machine learning algorithms use this data to learn specific patterns and then apply them in prediction or decision-making processes. This data can be in structured forms (such as tables or numerical values) or unstructured forms (such as text, audio, or visuals).

Model selection

A machine learning model is chosen in accordance with the nature of the problem and the available data. Simple linear regression and intricate neural networks are two examples of models. The job (such as classification, regression, or clustering) and problem complexity determine the model to use.

Training

In order to learn from the patterns in the input data, the machine learning algorithm processes the data during the training phase and modifies its parameters. In supervised learning, the model's predictions are contrasted with the actual, known results. The distance between the model's predictions and the actual values is measured by a loss function, which can be the mean squared error for regression or the cross-entropy for classification. To reduce this loss, the model iteratively modifies its parameters.

Learning process

All types of machine learning are shown in Figure 1.

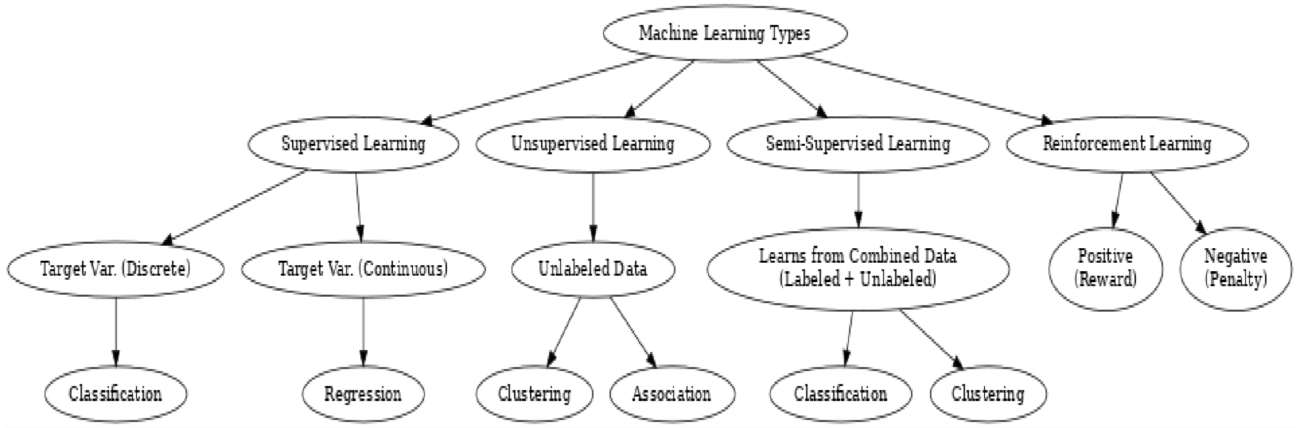
Supervised Learning: Due to the fact that the algorithm is trained on labeled data, every input has an appropriate output. The input-to-output mapping is learned by the model. When particular objectives are identified to be achieved from a specific set of inputs, supervised learning is implemented Sarker, I. H. et al. (2020). i.e., a task-driven approach. The two most popular supervised tasks are regression, which fits the data, and classification, which separates the data. Spam detection and picture classification are two examples.

Unsupervised Learning: In this case, the model's job is to identify hidden structures or patterns after being trained on unlabelled data. (Sarker, 2021; Han, et al., 2022). Clustering, density estimation, feature learning, dimensionality reduction, association rule discovery, anomaly detection, and so forth are among the most popular unsupervised learning tasks (Dwight, 1998).

Reinforcement Learning: Reinforcement learning, also known as environment-driven learning, is a type of machine learning algorithm that enables computers and software agents to automatically determine the optimal behavior in a certain context or environment in order to boost productivity (Kaelbling, L. P. et al. 1996). Through interaction with the environment and feedback in the form of incentives or punishments, the algorithm learns and adapts. It is employed in robotics and gaming, among other uses.

Semi-supervised Learning: Because semi-supervised learning uses both labeled and unlabeled data, it can be thought of as a combination of the supervised and unsupervised approaches described above (Sarker, 2021; Han, et al. 2022; Sarker, et al., 2020). Combines labeled and unlabeled data; this is frequently employed when data labeling is costly or takes a long time. The ultimate goal of a semi-supervised learning model is to generate a better prediction result than what could be obtained from the model using only the labeled data. Machine translation, fraud detection, data labeling, and text classification are just a few of the fields that use semi-supervised learning (Sarker, I. H. 2021).

Figure 1.
Various types of machine learning techniques (Sarker, I. H. 2021).



Model evaluation

Following training, the model is assessed using a different set of data—referred to as the test set—that was not used during training. This aids in evaluating the model's ability to generalize to fresh, untested data. Evaluation metrics include recall, accuracy, precision, and F1 score (for classification) or RMSE (for regression).

Optimization

Many machine learning algorithms use optimization techniques like gradient descent to minimize the error and adjust model parameters. In neural networks, for example, backpropagation is used to propagate errors backward through the network, adjusting weights to reduce prediction error.

Prediction

Once trained and evaluated, the model can be deployed to make predictions or decisions on new, unseen data (inference). Depending on the type of machine learning model, these predictions can range from classifying an image to predicting house prices.

Types of Machine Learning Algorithms

Naive Bayes classifier

The Naive Bayes algorithm is widely used due to its ease of implementation and understanding, making it useful for large datasets. It is a method where classification is performed using a conditional probability approach. The functioning of the algorithm begins with the transformation of the dataset into a frequency table. Later, probabilities are calculated to create a probability table.

Linear regression and logistic regression

For tasks involving regression and classification, respectively, these are straightforward methods. While logistic regression is used for binary classification, linear regression is utilized to predict continuous values. The aim is to find the relationship between the dependent variables, which should be categorical, and the independent variables in the logistic regression algorithm, and to model it in the best possible way.

Decision trees and random forests

Decision trees separated the data into subsets based on feature criteria. Random forests build many decision trees and combine their predictions for higher accuracy. Like the tree structure found in data structures, decision trees are made up of roots, nodes, and leaves. The goal of a decision tree algorithm with this structure is to learn decision rules from the attributes in order to create a model that can forecast the value of the target variable. The goal of the random forest method is to avoid the aforementioned overfitting issue. Many decision trees are being constructed for this purpose by choosing subgroups at random from the dataset and the features.

Support vector machines (SVMs):

SVMs are mostly used for classification; they identify the hyperplane that divides data points into distinct classes the best. Support vector machines aim to find a plane or hyperplane that can best separate the classes by using features, while minimizing classification error by selecting the line with the highest margin.

k-Nearest neighbors (k-NN)

A non-parametric technique that divides data points into groups according to the feature space's majority class of their closest neighbors. In the implementation of the algorithm, all training data is stored and the model is not created until a new sample needs to be classified. When it comes to classifying and predicting a new sample, unlike other algorithms, the classification process is completed each time using the training dataset instead of using parameters determined by preprocessing. Thus, no probability assumption is made regarding the distribution of the training dataset (Sarker, 2021).

Neural networks and deep learning

Layers of interconnected nodes, or neurons, make up neural networks, which are capable of modeling intricate, non-linear interactions (McCulloch, & Pitts, 1943). Large neural networks with numerous hidden layers that can extract hierarchical characteristics from data are used in deep learning. While long short-term memory (LSTM) networks and recurrent neural networks (RNNs) are utilized for sequential data like time series or language processing, convolutional neural networks (CNNs) are specialized for picture data.

Clustering algorithms (e.g., k-means, DBSCAN)

Without labeled data, these algorithms combine comparable data points into groups. Applications such as anomaly detection and market segmentation benefit from clustering.

Dimensionality reduction (e.g., PCA, t-SNE)

Methods for cutting down on a dataset's features without sacrificing its most crucial elements. It is helpful for enhancing model performance and data visualization.

Key Concepts in Machine Learning

Overfitting and underfitting

Poor generalization results from overfitting, which happens when a model learns the noise in the training data in addition to the underlying pattern. When a model is too basic to identify the underlying pattern in the data, underfitting takes place.

Cross-validation

A method for making sure the model works properly with various data subsets. The process entails creating folds out of the data and training the model on various combinations of these folds several times.

Regularization

By increasing the model's complexity, regularization techniques such as L1 (Lasso) and L2 (Ridge) regularization work to prevent overfitting.

Feature engineering

The procedure for choosing, altering, or adding characteristics to increase the predictive capacity of the model. The performance of a model can be significantly impacted by its features.

Hyperparameter tuning

The hyperparameters of machine learning models, such as the regularization strength and learning rate, are predetermined before training. Finding the ideal hyperparameter combination to optimize performance is known as hyperparameter tuning.

Machine learning algorithms basically use data to learn and then use that knowledge to forecast or make judgments. Data collection, model selection, training, evaluation, and optimization are the phases that the system goes through in order to deploy a model that can make decisions based on fresh data. Recommendation systems, autonomous cars, computer vision, natural language processing, and many other industries employ machine learning.

Results

Banta et al. (2024) examine the capabilities of machine learning models in predicting turbulent combustion rates in spark-ignition engines operating on hydrogen-natural gas mixtures. The speed of the vehicle is a critical parameter that has a direct impact on engine performance and emissions. Calculating the turbulent combustion rate using traditional methods requires complex mathematical models and long computation times, while machine learning techniques aim to significantly accelerate this process. In this context, various machine learning algorithms, such as linear regression, support vector machines, random forests, and artificial neural networks, have been applied, and their performances compared. The results obtained indicate that machine learning models can provide faster and more accurate results in predicting combustion rates compared to traditional methods.

Bukkarapu et al. (2024) investigate the feasibility of applying machine learning algorithms for the classification of different combustion regimes in biodiesel-fueled homogeneous charge compression ignition (HCCI) engines. Biodiesel has the potential to reduce dependence on fossil fuels and provide a more sustainable transportation solution. However, the combustion regimes of HCCI engines in biodiesel usage can vary significantly depending on engine load, compression ratio, and the reactivity of the biodiesel. HCCI engines operate through the auto-ignition of a homogeneous air-fuel mixture, making it crucial to accurately classify the combustion regime to ensure the engine runs efficiently and stably. However, due to the compositional diversity of different types of biodiesel and the operating conditions of the engines, it is quite challenging to determine these combustion regimes through experimental methods. This study has developed a model to classify four fundamental combustion regimes for biodiesel-fueled HCCI engines (engine misfire, diluted stable combustion, diluted stable combustion, and knocking) using machine learning methods.

Batool et al. (2024) have investigated the use of machine learning algorithms to describe and classify different heat emission patterns in low-temperature combustion (LTC) engines. Low-temperature combustion engines provide high thermal efficiency while producing low nitrogen oxide (NO_x) and soot emissions. However, the high rate of increase in cylinder pressure limits the operating range of these engines. For this reason, controlling combustion is of great importance to ensure the safe operation of the engine. This study aims to classify heat release rate (HRR) patterns under over six hundred motor operating conditions using supervised machine learning algorithms such as decision trees, K-nearest neighbors (KNN), and support vector machines (SVM), as well as unsupervised machine learning algorithms like K-means clustering. SVM has emerged as the most successful method, with an accuracy rate of 92.4%. The main contribution of the study is the development of a model predictive control (MPC) framework for cycle-to-cycle combustion control of a multimodal LTC engine using these classification algorithms.

The high cost and time-consuming nature of engine calibration, as well as critical issues such as combustion instability, emission prediction, knocking, and combustion mode transitions, have been highlighted in the context of the advantages provided by ML. Aliramezani et al. (2022), in their review study, thoroughly examined the current and potential applications of machine learning on internal combustion engines. In their study, ML methods are classified into three main categories: supervised learning, unsupervised learning, and reinforcement learning. Thanks to ML algorithms, it is possible to learn the relationships between complex inputs and outputs with high accuracy without the physical knowledge of motor systems. Additionally, the grey-box approach, which combines physical modeling with ML-based models, is recommended for real-time control and high-accuracy modeling. The successes of methods such as artificial neural networks (ANN), support vector machines (SVM), extreme learning machines (ELM), and Gaussian processes (GP) in engine modeling, diagnosis, and control have been compared. This also paves the way for advanced applications such as information sharing and fault diagnosis among similar engines. The integration of machine learning with technologies such as cloud computing and vehicle-to-infrastructure (V2I) communication also enables the development of highly accurate, customized control strategies using large datasets collected from engines. As a result of the study, machine learning has been found to offer innovative solutions to the multidimensional problems faced by internal combustion engine technologies and is considered to form the basis for smarter, more flexible, and efficient engine control systems in the future [6].

Al-jabiri et al. (2024) experimentally investigated the effects of different ratios (B10–B40) of sunflower biodiesel blends on a single-cylinder, four-stroke compression-ignition diesel engine in their study. Experimental data were evaluated based on parameters such as brake thermal efficiency (BTE), specific fuel consumption (BSFC), and exhaust emissions (CO, NOx, HC). To improve the model outputs, the Ameliorative Whale Optimization Algorithm (AWOA), developed as a multi-objective optimization technique, was applied, maximizing BTE while minimizing harmful emissions such as CO and NOx, as well as BSFC. Then, these experimental data were modeled using the Alternating Model Tree (AMT) algorithm, achieving high prediction accuracy ($R^2 > 0.98$). AWOA has been compared with the classical Whale Optimization Algorithm (WOA) and Particle Swarm Optimization (PSO), and it has been observed that it provides faster convergence and more accurate results. In this study, it is significant in terms of demonstrating both the potential of biodiesel fuels to reduce environmental impacts and the effectiveness of artificial intelligence-based modeling and optimization techniques in this field. Additionally, it has been demonstrated that the negative effects of biodiesel on efficiency can be significantly improved through appropriate optimization methods.

The experimentally obtained data were analyzed using 29 different regression algorithms. In the study by Venkatesk et al. (2023), the effectiveness of machine learning (ML) algorithms in predicting the exhaust emissions of a dual-fuel compression ignition (CI) engine operating on hydrogen and diesel was examined. As input parameters, hydrogen concentration, engine load, diesel fuel input, engine speed, and equivalence ratio were used; as output, emissions such as NOx, CO_2, hydrocarbons (HC), and smoke were predicted. Among the algorithms used in their study, it was determined that the Pace Regression, Radial Basis Function Regressor (RBFreg), Multilayer Perceptron Regressor (MLPreg), and Alternating Model Tree (AMT) models achieved high correlation coefficients for each type of emission. Especially, the MLP regressor model has stood out in providing the most balanced and high-accuracy predictions for all emission parameters (R in the range of 0.85–0.99). Additionally, using the J48 decision tree algorithm, the effects of physical engine parameters on emissions have been interpreted, and the variables that most significantly impact emission formation have been identified (Sugumaran, V. et al. 2023).

In the study by Bai et al. (2023), a dual-fuel engine configuration was implemented using hydrogen gas in addition to WGO. Hydrogen, due to its high flame speed and calorific value, increases combustion efficiency and reduces harmful emissions. Among the algorithms used are Multiple Linear Regression (MLR), Decision Trees (DT), Random Forest (RF), and Support Vector Regression (SVR). In addition to these experimental data, Machine Learning (ML) algorithms were used to predict engine behaviors. Within the scope of the study, the engine was tested with 5%, 10%, and 15% hydrogen energy shares; performance and emission parameters such as NOx, CO, HC, smoke, and brake thermal efficiency (BTE) were analyzed. These models were trained with independent variables such as brake power and brake-specific fuel consumption (BSFC) and compared in predicting emission and performance parameters. As evaluation criteria, R^2, MAE, MSE, and RMSE values were used, and the results showed that the MLR model had the highest compatibility with the experimental data ($R^2 \approx 0.999$).

Ma et al. (2024), in their study, systematically examined the effects of fuel properties (RON—Research Octane Number, S—Fuel Sensitivity) and environmental conditions (pressure, temperature, oxygen concentration) on spray combustion and NOx emissions in GCI engines using the Machine Learning (ML)-supported Global Sensitivity Analysis (GSA) method. In their studies, computational fluid dynamics (CFD) simulations were conducted for 200 different operating points, and ML models were trained using the obtained data. In the GSA analysis, the Sobol method was used to evaluate the primary and interactive effects of the parameters on the target outputs. The study also compares the prediction performance of different ML algorithms (Polynomial Regression, Gradient Boosting Regressor, Extra Trees, etc.) and determines the most suitable models based on target outputs ($R^2 > 0.90$).

In their study, Khac et al. (2023) developed artificial neural network (ANN)-based models that enable the prediction of NO_x and CO_2 emissions using in-cylinder pressure data. The data used in the study were obtained from a four-cylinder marine diesel engine. Input data consists of the main components obtained from in-cylinder pressure signals and engine parameters such as engine speed, load, injection timing, and exhaust temperature. During the model development process, multilayer perceptron (MLP) and radial basis function (RBF) networks were compared. The results have shown that MLP networks have lower error rates and higher accuracy in predicting both NO_x and CO_2 emissions.

In the study by Williams et al. (2022), they investigated the prediction of the ROI profile, which is difficult to obtain through traditional methods, using Machine Learning (ML)-based models and the use of these predictions in CFD simulations. For this purpose, ROI (Rate of Injection) data were collected at different pressures and durations from a heavy-duty diesel injector using the Bosch tube method, and then both Random Forest and Neural Network models were trained using this data. The developed ML models were able to generate an ROI profile with high accuracy for the given rail pressure and injection time. Especially the Neural Network model produced consistent predictions even under low- and high-pressure conditions; it was observed that the Random Forest model was successful in interpolation but weak in extrapolation. The produced ROI profiles were later used in CFD simulations, and the impact of different ROI shapes on combustion characteristics and emissions was analyzed. It has been shown that ML-based ROI profiles produce more accurate results compared to traditional trapezoidal or stretched ROI profiles, providing significant improvements in critical outputs such as ignition timing and NO_x formation.

Conclusion

This study presents a summary of research on internal combustion engines that utilizes various machine learning algorithms to predict different motor performance parameters and exhaust emission values. Machine learning methods enable high-accuracy predictions. Additionally, studies have shown that computation times have significantly decreased compared to traditional methods.

The common findings of these studies reveal that machine learning can be effectively applied in both classical engine types (SI, CI) and advanced combustion modes (HCCI, RCCI, LTC). The strengths and weaknesses of different algorithms for prediction (regression), classification, and control applications have been identified. Especially in terms of real-time applications:

1. Algorithms such as Random Forest and SVM offer high accuracy and adaptability.
2. Decision trees are advantageous in terms of interpretability.
3. Deep learning has potential in more complex systems but requires more data and computational power.

In the future, hybrid combinations of physics-based models with data-driven approaches and the transition to digital twin systems will open new horizons in engine modeling and control.

Peer-review: *Externally peer-reviewed*
Financial disclosure: *This research received no external funding.*
Conflict of Interest: *The author has no conflicts of interest to declare.*

References

Aliramezani, M., Koch, C. R., & Shahbakhti, M. (2022). Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions. *Progress in Energy and Combustion Science, 88, 100967.*

Al-jabiri, A. A., Balla, H. H., Al-zuhairy, M. S., Alahmer, H., Al-Manea, A., Al-Rbaihat, R., & Alahmer, A. (2024). Applied AMT machine learning and multi-objective optimization for enhanced performance and reduced environmental impact of sunflower oil biodiesel in compression ignition engine. *International Journal of Thermofluids, 24, 100838.*

Bai, F. J. J. S., Shanmugaiah, K., Sonthalia, A., Devarajan, Y., & Varuvel, E. G. (2023). Application of machine learning algorithms for predicting the engine characteristics of a wheat germ oil–Hydrogen fuelled dual fuel engine. *International journal of hydrogen energy, 48(60), 23308-23322.*

Banta, N. J. I., Patrick, N., Offole, F., & Mouangue, R. (2024). Machine learning models for the prediction of turbulent combustion speed for hydrogen-natural gas spark ignition engines. *Heliyon, 10(9).*

Batool, S., Naber, J. D., & Shahbakhti, M. (2024). Machine learning approaches for identification of heat release shapes in a low temperature combustion engine for control applications. *Control Engineering Practice, 144, 105838.*

Bukkarapu, K. R., & Krishnasamy, A. (2024). Evaluating the feasibility of machine learning algorithms for combustion regime classification in biodiesel-fueled homogeneous charge compression ignition engines. *Fuel, 374, 132406.*

Dwight, J. (1998). Aluminium design and construction. CRC Press.

Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research, 4, 237-285.*

Khac, H. N., Modabberian, A., Zenger, K., Niskanen, K., West, A., Zhang, Y., ... & Mikulski, M. (2023). Machine learning methods for emissions prediction in combustion engines with multiple cylinders. IFAC-PapersOnLine, 56(2), 3072-3078.

Ma, Y., Yang, D., & Xie, D. (2024). Investigating the effect of fuel properties and environmental parameters on low-octane gasoline-like fuel spray combustion and emissions using machine learning-global sensitivity analysis method. *Energy, 306, 132551.*

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5(4), 115-133.*

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science, 2(3), 173.*

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science, 2(3), 160.*

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data, 7(1), 41.*

Sugumaran, V., Thangavel, V., Vijayaragavan, M., Subramanian, B., JS, F. J., & Varuvel, E. G. (2023). Efficacy of machine learning algorithms in estimating emissions in a dual fuel compression ignition engine operating on hydrogen and diesel. International *Journal of Hydrogen Energy, 48(99), 39599-39611.*

Williams, Z., Moiz, A., Cung, K., Smith, M., Briggs, T., Bitsis, C., & Miwa, J. (2022). Generation of Rate-of-Injection (ROI) profile for Computational Fluid Dynamics (CFD) model of Internal Combustion Engine (ICE) using machine learning. *Energy and AI, 8, 100148.*

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record, 31(1), 76-77.*