

Full Length Article



Advancing Fuel Spray Characterization: A Machine Learning Approach for Directly Injected Gasoline Fuel Sprays

Sadique Khan^{a,b}, Mudassir Masood^{c,d}, Mario Medina^e, Fahad Alzahrani^{a,b,*}

^a Mechanical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

^b Interdisciplinary Research Centre for Hydrogen Technologies and Carbon Management, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

^c Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

^d Interdisciplinary Research Center for Communication Systems and Sensing, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

^e Mechanical Engineering Department, California State University, Los Angeles, USA

ARTICLE INFO

Keywords:

Machine learning
Gasoline direct injection
Fuel spray
Compression ignition
Spray tip penetration
Spray angle

ABSTRACT

Compression ignition engines when operated on gasoline fuels cause significant reduction in NO_x and particulate emissions. In such advanced combustion strategy, the fuel-oxidiser mixing process, intensified by the prolonged ignition delays of gasoline fuels, directly affects the stability and efficiency of combustion. Thus, optimising fuel spray characteristics leads to optimisation of injector design, engine performance and subsequent decrease in emissions. Since spray development is a complex process that involves wide range of length and time scales, computationally expensive modelling techniques can be replaced with machine learning (ML) models. These ML models are employed to predict spray characteristics utilizing datasets generated from comprehensive spray studies. In this work, a dataset of about 5400 instances taken from non-evaporating gasoline fuel spray imaging experiments under Gasoline Compression Injection (GCI) engine conditions is used to train various ML models with data split of 70 % and 30 % for training and testing, respectively, with five-folds cross-validation performed within the training. The fuel injection pressure (60 – 150 MPa), chamber pressure (0.1 – 2 MPa), nozzle diameter, nozzle hole conicity and injection duration are used as input features to the models for predicting the spray tip penetration and spray angle. The performance of four ML models was evaluated and compared under default and tuned hyperparameters against experimental data and available physics-based correlations in the literature. The models include random forest, extreme gradient boosting, multilayer perceptron, and elastic-net. The results show that the hyperparameter-tuned extreme gradient boosting model performs best in predicting the spray parameters. The overall model performance was evaluated using the coefficient of determination (R²), mean absolute error (MAE), and root mean squared error (RMSE), resulting in values of 0.884, 0.651, and 1.571, respectively. This study presents compelling evidence demonstrating the effectiveness of ML as a powerful tool for isolating non-linear behaviors from physical processes. By effectively decoupling these behaviors, ML enhances the accuracy of predicting spray characteristics while significantly reducing computational costs. The application of ML in fuel injector design has the potential to revolutionize engine performance and contribute to substantial reductions in emissions.

1. Introduction

The global energy consumption is predicted to rise by about 36 % in 2050 [1] with around 80 % of the energy demand to be met through

combustion [1]. The transportation sector consumes one third of the global energy [2] with internal combustion engines (ICEs) being the dominant engines employed for powering transport vehicles. Forecasts show that the global ICE market is projected to grow by 27.2 billion USD

Abbreviations: ICE, Internal Combustion Engine; CI, Compression Ignition; GCI, Gasoline Compression Ignition; GDI, Gasoline Direct Injection; LDV, Light Duty Vehicle; LTC, Low Temperature Combustion; HCCI, Homogeneous Charge Compression Ignition; PM, Particulate Matter; S, Spray Tip Penetration Distance; θ , Spray Angle; RF, Random Forest; XGB, Extreme Gradient Boosting; MLP, Multilayer Perceptron; ENET, Elastic Net; MAE, Mean Absolute Error; RMSE, Root Mean Squared Error.

* Corresponding author at: Mechanical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.

E-mail address: fahadmz@kfupm.edu.sa (F. Alzahrani).

<https://doi.org/10.1016/j.fuel.2024.131980>

Received 1 February 2024; Received in revised form 18 April 2024; Accepted 19 May 2024

Available online 8 June 2024

0016-2361/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

between 2022 and 2027 [3] with a compound annual growth rate of 8.32 %. Although the engine market predicts tremendous growth, stringent policies placed with regards to vehicular emissions [4] has led researchers to develop engine technologies that can optimise ICE. Numerous advanced combustion technologies have been developed over the years to achieve improved efficiency and combat emissions. Some of these technologies include variable compression ratio engines, low-temperature combustion (LTC), boosted engines, lean combustion and many more. These technologies have the potential to enhance fuel economy up to 40 % [5–8]. LTC has been studied extensively as it provides optimum fuel consumption and consequently emissions reduction since it operates at relatively lower temperatures when compared to conventional engines. Homogeneous Charge Compression Ignition (HCCI) that falls under LTC uses the combined compression ignition technique with spark ignition methods. It can be beneficial in terms of efficiency; however, it is often difficult to control which in turn results in higher hydrocarbon (HC) and carbon monoxide (CO) emissions [9–12]. These limitations can be overcome by combining the concept of HCCI and gasoline direct injection (GDI) which is commonly known as gasoline compression ignition (GCI). Fuel is directly injected into the chamber under high pressures during GCI combustion which results in longer ignition delays that allows for better fuel–air mixing. The thermal efficiency can reach up to 40 % using GCI combustion with significantly lower levels of NO_x emissions [13,14]. An illustration of the conventional and advanced engine combustion technologies is described in Fig. 1.

Direct injection has gained popularity due to its advantages of lower emission and enhanced fuel delivery control [15]. The engine performance and emissions are significantly affected by the spray formed as a result of fuel injection (see Fig. 2) which in turn is governed by the injector nozzle design and operating conditions. Thus, it becomes of a paramount importance to characterize the internal and external nozzle flows, and how they are affected by the injector nozzle design and the engine operating conditions.

The overall spray structure can be described by a number of important parameters. The spray penetration is a characteristic of the spray that is typically defined as the maximum distance the spray traverses from the point of injection (Fig. 3). For non-evaporative conditions of the fuel, the penetration of the spray is simply known as spray penetration, while for evaporative conditions the spray penetration is divided into liquid and vapor penetration. The study of spray penetration is also crucial for understanding chamber wall impingement, a factor that contributes to emissions, in addition to comprehending the dynamics of the fuel–air mixture [17]. The spray angle is another important characteristic of the spray structure. It represents the degree to which the spray can spread allowing for quantification of the air entrained by the spray and leading to the determination of fuel

evaporation. The spray angle has been calculated in the literature by different ways [18–20]. Medina et al. [21] has calculated the spray angle for a single hole injector for gasoline direct injection as the inclusive angle at one third distance of the spray penetration, denoted by θ in Fig. 3. Furthermore, the spray breakup that occurs as the spray advances into the domain results in the formation of tiny droplets, promoting the fuel–air mixture formation. The spray breakup is categorized into primary and secondary breakup (see Fig. 4). Primary breakup occurs at the nozzle exit due to turbulence caused by sudden pressure drop across the injector and is also influenced by the implosion of cavitating bubbles that are formed near the nozzle exit. The secondary breakup takes place at a later stage downstream of the regime caused by aerodynamics instabilities on the fluid surface due to surrounding air turbulence [22,23]. According to Medina et al. [21], the spray breakup time is calculated as the point of maximum spray tip penetration rate.

Operating conditions such as fuel injection pressure and chamber pressure largely affect the spray characteristics. Literature shows that for a fixed fuel injection pressure and varying chamber pressure, the spray penetration is almost similar before the spray breakup time as the spray is driven by momentum, while at later stages the effect of chamber pressure and air entrainment dominate causing deviation in the spray penetration with penetration being faster at lower chamber pressures [24]. The same effect is observed with varying injection pressures [25]. Similarly, the spray angle is more sensitive to chamber pressures after the spray breakup time with spray angle increasing with increased chamber pressures. With regards to the effect of injector nozzle design, the injector hole count, hole diameter, hole length, conicity as well as the injector needle, which allows for opening and closing of the nozzle, affect the fuel flow rates and the resulting spray formation [24,26–30]. A schematic showing a typical injector nozzle design is shown in Fig. 5. Current studies show that converging nozzles with larger outlet hole diameters result in increased spray penetration. However, with increasing nozzle diameter, the increased spray penetration leads to fuel impingement on chamber surfaces that can affect the combustion performance [31]. Spray formation is also affected by recirculation zones formed in the flow field [32,33], cavitation [34,35] and flash boiling [36–38]. As the liquid fuel exits the nozzle, spray structure evolves through atomization which is facilitated by factors like turbulence, surface instabilities, drag and surface tension of the fluid [39]. In addition, the spray droplets undergo collision and coalescence [40,41]. Under relevant engine conditions, evaporation of the fuel occurs, and it affects the mixture formation, ignition delay period, combustion characteristics and engine out emissions [42–44].

Experimental studies have been performed to characterize injector internal flow, near nozzle flow and external flow. Internal geometry was studied by Manin et al. [45] using X-ray tomography to assess the diameter of the orifice and by Duke et al. [46] to characterize the shape of the nozzle sac and orifice. Some studies also focused on the orifice angles, which determined the extent to which the spray deflects as fluid exits the orifice [47,48]. Costa et al. [49] carried out gasoline multiple injection studies and concluded that mass flow measurements enables optimization of the fuel economy. A recent study by Medina et al. [24] carried out mass flow rate as well as momentum flux measurements for gasoline fuels directly injected using two-hole injectors. They also focused on nozzle external flow by characterising spray structures at injection pressures reaching up to 150 MPa using five different orifice geometries. Results showed that cylindrical orifices with larger outlet diameter resulted in the highest spray tip penetration, while divergent orifices with 20 % hydro-erosion rounding resulted in the largest spray angle. Numerous other studies focused on engine performance and emissions, including the study by Hoffman et al. [50]. Using Mie scattering technique, the authors obtained spray images and concluded that increasing injection pressure reduced fuel droplet size, indicating enhanced atomization. Furthermore, they found a significant reduction in particulate emissions with increased fuel injection pressure (40 MPa). Another study by Merola et al. showed that multiple injection or split

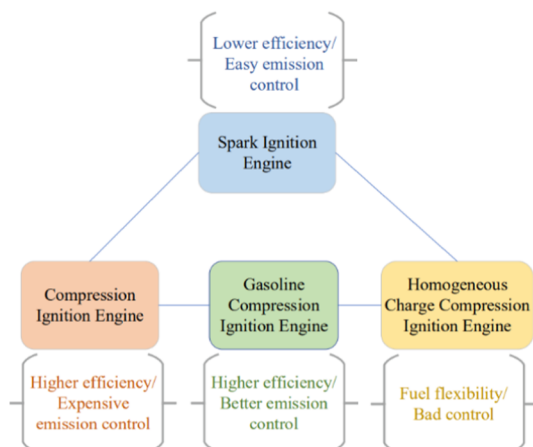


Fig. 1. Conventional and advanced engine combustion technologies.

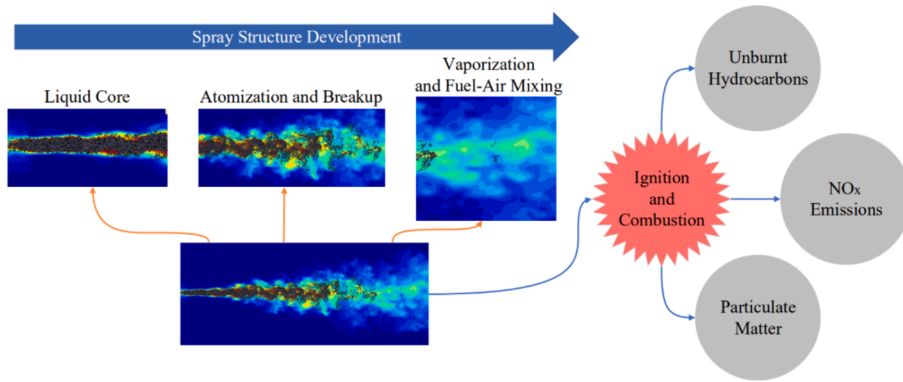


Fig. 2. Schematic representing the spray formation and its effect on combustion and emissions. Adapted from Kaario et al. [16]

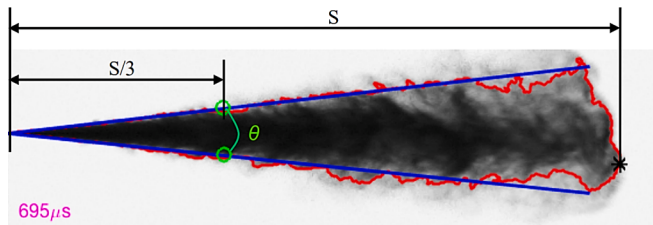


Fig. 3. A snapshot of the injection event representing the spray penetration distance and spray angle at operating condition of 60 MPa injection pressure and 2 MPa chamber pressure [21].

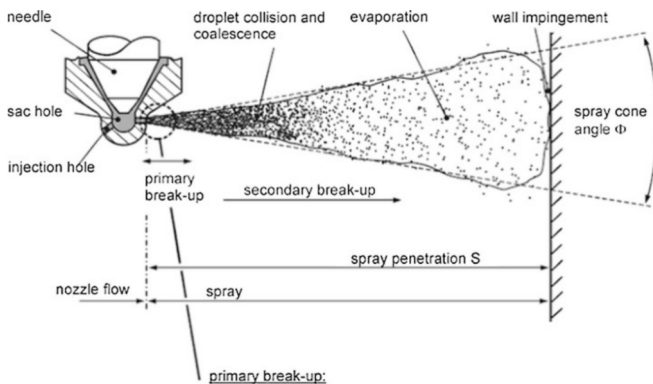


Fig. 4. Depiction of fuel spray primary and secondary breakups. Adapted from Duronio et al. [22]

injection can enhance fuel atomization that leads to better fuel–air mixing [51]. Nonetheless, direct injection can result in fuel impingement on the cylinder surface, especially at high injection pressures. The impingements are considered a major source of particulate emissions as shown in literature [52,53].

To overcome challenges associated in understanding spray behaviour using experiments [54,55], numerical studies have been carried out over the past few decades to assist in comprehending the complex spray evolution process. Internal nozzle flow studies were performed by Grief et al. [56] by employing multi-fluid Eulerian, to model mixture of gasoline and ethanol fuels while Mishra et al. [57] used a similar model coupled with volume of fluid model to track the liquid and gas interfaces under cavitating conditions. A simpler model known as the homogeneous Eulerian model was used by Moulai et al. [58] to study internal nozzle flow and they successfully validated the model with experimental results under similar conditions. Near-nozzle studies conducted by Shost et al. [59] and Befrui et al. [60] aimed to investigate the effects of orifice

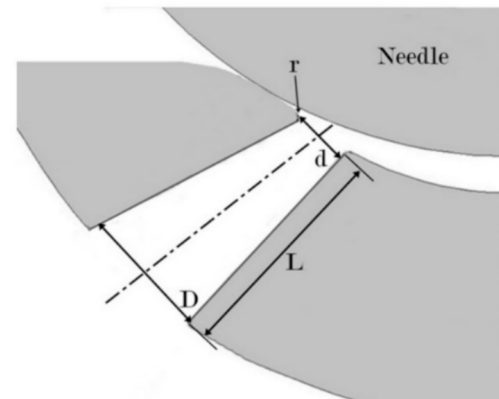


Fig. 5. Injector nozzle representing the various associated design parameters. Here, L denotes the hole length, d and D are the inlet and outlet hole diameters, and r is the rounded inlet radius.

geometry on spray pattern. In particular, they compared straight orifices with counter-bore orifices, and their results were in good agreement with experimental studies in term of spray breakup, as well as the spray cone angle and spray penetration. Another study focused on fuel deposits and how they are affected by the nozzle design, eventually affecting the combustion performance and engine out emissions [61]. Han et al. [62] using CFD code KIVA-3 attempted at mimicking the spray pattern with the experimental results using pressure swirl injector in a DISI engine. In a recent study, CFD modelling of ethanol-blend fuel sprays in a constant volume chamber was conducted using CONVERGE software [63]. The model was evaluated against results obtained from a DISI engine, where good agreement was observed.

With the recent development of injection systems that can provide high injection pressures [64], short injection pulses make the spray characterization phenomena complex and challenging due to the involvement of wide range of physical processes operating at different length and time scales. Over the years various experimental and computational studies have been carried out to characterize vaporizing and non-vaporizing sprays depending upon the study objective. The spray studies have generated datasets that can be used to train machine learning (ML) models for predicting the spray characteristics [65–68]. In a study by Chang et al. [69] involving ML prediction of spray target coordinates using GDI injectors with counter bore, it was observed that gradient tree boosting regression (GBRT) and random forest (RF) predicted the target coordinates more accurately compared to artificial neural networks (ANN), with a coefficient of determination (R^2) value of 0.99. A similar study by Chang et al. [70] aimed at predicting spray characteristics under collapse and non-collapse conditions at fixed

injection pressure of 35 MPa using different tree-based models showed that GBRT provided accurate predictions for spray tip penetration, with an R^2 value of 0.975. Koukouvinis et al. [71] used ANN with just one hidden layer and three neurons to effectively predict the spray penetration distance over time with a computational time of few seconds. In another study using GDI injector, ANN consisting of four features with dataset division of training, testing, and validation as 70 %, 20 % and 10 % respectively was used to successfully predict the rate of injection (ROI) and solenoid voltage signal with coefficient of determination (R^2) of 0.975 [72]. Hwang et al. [73] used a regression model to predict the spray topology image as well as the liquid penetration length and spray angle using ECN spray G injector, showing good agreement with the experimental observations.

The experimental and computational studies on sprays involve certain complexities. On one hand, experimental characterization of spray requires huge setup costs and limits the comprehensive understanding of the cavitation and flash boiling phenomena. On the other hand, numerical modelling of fuel sprays presents modelling challenges, primarily due to turbulence, atomization, heat transfer, and other associated phenomena. Additionally, the wide range of length and time scales adds complexity in developing the computational model. ML models have the potential to determine the dominating parameters that govern the complex spray evolution process as a cost-effective method compared with experiments and computational models [74]. While there have been some attempts to predict spray patterns and behaviour using ML models, research in this domain remains limited. Furthermore, these attempts focused on characterizing the spray dynamics at low injection pressures and there is a lack of studies that modelled the gasoline direct injection process using machine learning at high injection pressures, particularly for GCI conditions. Therefore, the objective of this study is to characterize high pressure gasoline sprays using different ML algorithms and evaluate the performance of the models in predicting the spray characteristics, in particular, the spray tip penetration and the spray angle.

2. Machine Learning Methodology

2.1. Dataset

The spray dataset was derived from experiments performed by Medina et al. [21,24] under non-evaporating operating conditions using gasoline fuel. The injection events were captured using a high-speed camera and image processing techniques were utilized to detect and measure deterministic parameters, including the spray tip penetration distance and the spray angle. Each spray imaging measurement was repeated twenty times and the spray tip penetration and the spray angle for each sample in the dataset represent the average values of the repetitions. The error bar for each experimental datapoint represents plus and minus one standard deviation from the average value. The injection pressure ranged from 60–150 MPa, and the chamber pressure ranged from 0.1–2 MPa for a total injection duration of 1 ms. The spray images were recorded at time steps of approximately 0.014 ms. While the mass flow rate of the fuel, which is an important parameter to consider, was not directly recorded during the experiments, its impact on the spray parameters is indirectly taken into consideration since it is correlated to the chamber pressure and the injection pressure through Equation (1).

$$m_f = C_d A_o \sqrt{2\rho_f \Delta P} \quad (1)$$

In the above equation, m_f represents actual mass flow rate, C_d is the discharge coefficient, A_o is the outlet cross-section area, ρ_f is the fluid density, ΔP is the difference between the injection and chamber pressure.

The summary of operating conditions is described in Table 1. Three injectors were used for the study, A (single hole, converging nozzle), B

Table 1

Summary of operating conditions.

Fuel	Reference grade gasoline
Chamber Gas	Nitrogen
Injection duration [ms]	1
Injection pressure [MPa]	60, 90, 120, 150
Chamber pressure [MPa]	0.1, 0.5, 1, 2
Chamber temperature [K]	298

(dual hole, both straight nozzle) and C (dual hole, diverging and converging nozzle), and are described in Table 2. The injection pressure (IP), chamber pressure (CP), normalized nozzle outlet diameter (OD) with respect to maximum outlet diameter, conicity (C) (defined with respect to inlet and outlet hole diameters) and instantaneous injection duration (T) were considered as input features to the ML algorithms with the total dataset comprising approximately 5400 samples. The spray tip penetration (S) and spray angle (θ) were considered as outputs features (Table 3). The dataset was shuffled and split randomly into 70 % training and 30 % testing, which is the common practice observed in the literature [75]. To avoid data leakage, only the training set was standardized. Five-fold cross validation was done within the training so that the models are evaluated on each of the folds. This is a common way to reduce overfitting and selection bias when evaluating ML models. A schematic of the ML framework used in training and creating the model is represented in Fig. 6.

2.2. Exploratory Data Analysis

The dataset containing no null values was visualized using the correlation heat map (Fig. 7) to gain insight into the linear dependencies of the output labels on the input features, as well as between the input features themselves. A value close to 1 and -1 represents strong positive and negative correlation respectively between the features while a value close to 0 represents no linear relationship. The spray angle showed a slightly positive linear dependency on the chamber pressure and the nozzle outlet diameter, while no linear relationship existed with the other input features. As the spray progresses into the domain, the spray tip penetration increases with time, as depicted in Fig. 8. This is further evident in the correlation heatmap, revealing a significant linear relationship between the spray tip penetration and injection duration.

2.3. Machine Learning Algorithms

The performance of four different ML algorithms was evaluated in predicting the characteristics of the spray. Two tree-based ensemble models, random forest (RF) and extreme gradient boosting (XGB), were selected based on their demonstrated effectiveness on tabular datasets [76]. In addition, multilayer perceptron (MLP), a feedforward algorithm that falls under the category of neural networks, was assessed and

Table 2

Injector geometry specifications used in the spray study by Medina et al. [24].

Injector	Orifice	Normalized nozzle outlet diameter ^{a)}	Conicity ^{b)}	Nozzle type
A	Hole 1	0.578	1.5	Converging
B	Hole 1	0.578	0	Straight
	Hole 2	1	0	Straight
C	Hole 1	0.789	-1.5	Diverging
	Hole 2	0.789	3.5	Converging

Note:

a) Normalized with respect to maximum outlet diameter studied ($d_{max} = 190 \mu\text{m}$).

b) conicity = $(d_i - d_o)/10$, where d_i and d_o are the inlet and outlet diameters of the nozzle respectively.

Table 3
Sample dataset.

Input Features					Output Features	
Injection pressure (IP) [MPa]	Chamber pressure (CP) [MPa]	Normalized nozzle outlet diameter (OD)	Conicity (C)	Injection time (T) [ms]	Spray tip penetration (S) [mm]	Spray angle (θ) [degree]
60	0.1	0.578	1	0.13037	24.873	7.723
120	1	1	1.136	0.11588	21.277	12.605
90	0.5	0.789	1.233	0.24625	38.583	10.281
150	2	0.578	0.900	0.44904	36.253	13.358
30	0.1	1	1	0.37622	29.801	5.205

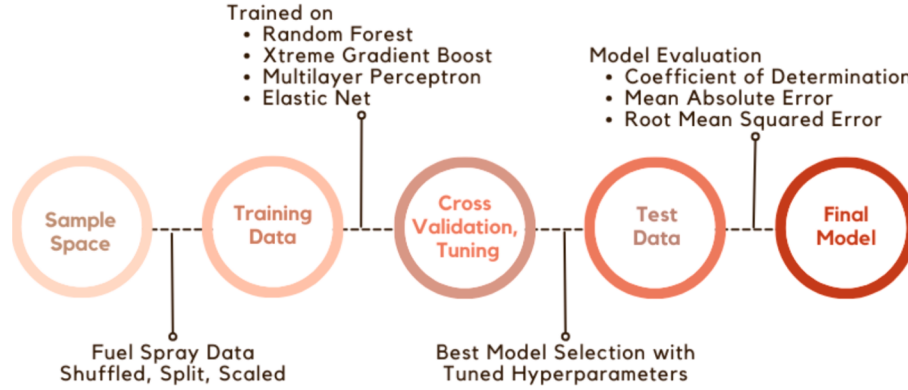


Fig. 6. Schematic of the ML model framework used in the study.

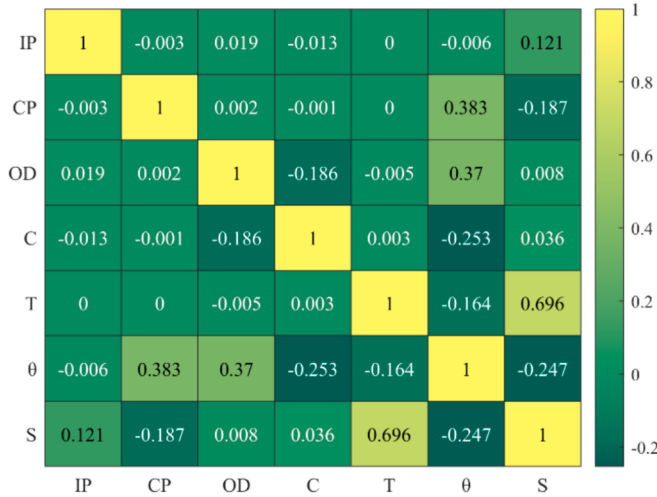


Fig. 7. Linear dependencies between input and output features.

compared with the conventional tree-based models. Furthermore, elastic-net (ENET), a regularized linear regression model, was chosen to evaluate its performance on the non-linear dataset. Scikit learn library under Python was used to test the different models along with their hyperparameters [77].

2.3.1. ML Algorithm: Random Forest (RF)

RF is an ensemble learning technique that uses multiple decision trees to make accurate predictions. It works by building a large number of decision trees, each trained on a random subset of the input and output. The predictions of the individual trees are then aggregated by taking the average [78]. Consider decisions trees given by Tree 1, Tree 2, ..., Tree n as shown in Fig. 9. The decision at each node of the tree divides the data into two subsets based on a selected feature and a threshold value. When the maximum depth or minimum number of samples is reached, the node split in the tree stops.

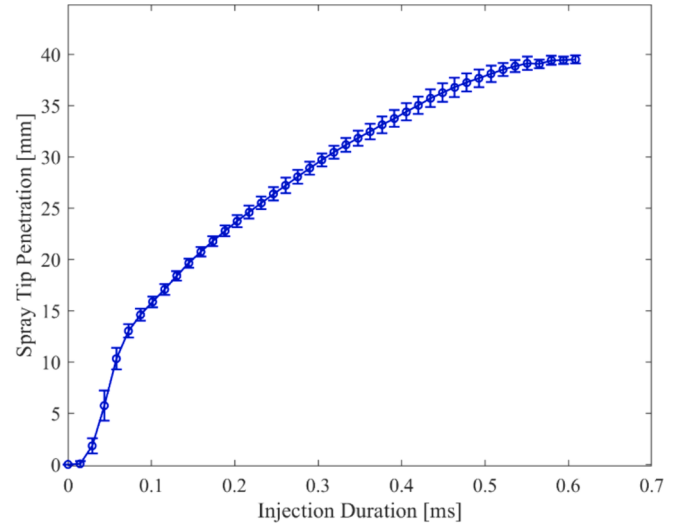


Fig. 8. Spray tip penetration distance as function on injection duration using injector A for injection pressure of 150 MPa and chamber pressure of 2 MPa [21].

For each decision tree trained on a sample of training data given in the form $[x, \theta_i]$ and using a random subset of features at each node split, prediction is made given by $h_i(x, \theta_i)$. This random selection of features results in diverse predictions by each tree leading to better generalization of the model and reduction in overfitting. The random forest averages the prediction made at each tree which is given by the following equation.

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n h_i(x, \theta_i) \quad (2)$$

The various hyperparameter settings considered while investigating the performance of Random Forest are provided in Table 5.

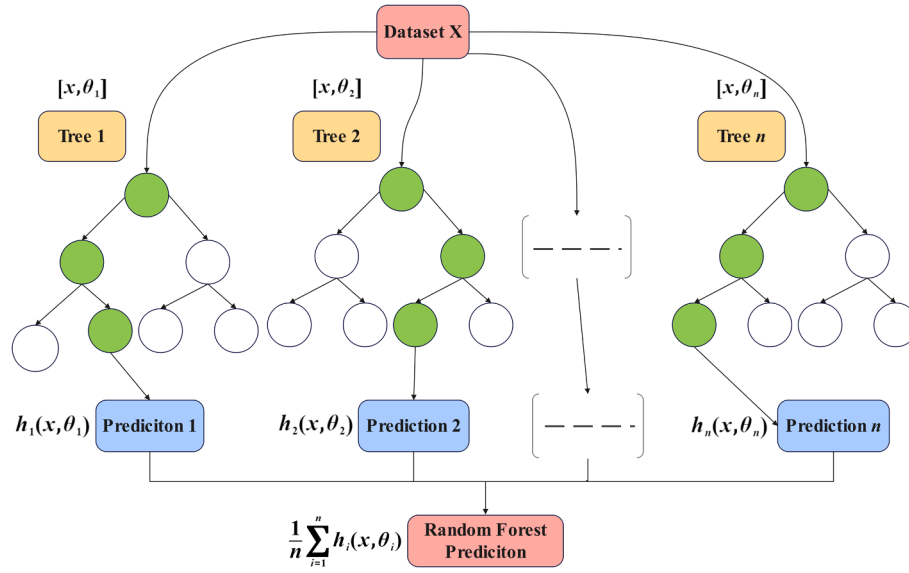


Fig. 9. Random forest flow chart.

Table 4
ML algorithms compared based on memory, speed, data and accuracy.

Model	Memory Usage	Computational Speed	Data Handling	Accuracy
RF	High- Large number of decision trees consumes memory.	Slow- Training is slower due to the large number of trees.	Proficient in analyzing and processing extensive datasets.	High- can be surpassed by boosting methods.
XGB	Moderate to High- The implementation is efficient, but it can scale with the number of trees and depth.	Moderate- Trains more rapidly than RF.	—Proficient in handling extensive datasets —adept at managing sparse data —susceptible to overfitting when working with limited datasets.	Very high- typically surpasses other models when appropriately fine-tuned.
MLP	High- Weights assigned to each neuron.	Slow- Due to backpropagation and large number of iterations.	significantly influenced by feature scaling —susceptible to overfitting if large number of features are involved.	High- Usually with complicated patterns and fine-tuning.
ENET	Low- Reduces the complexity of models by selecting only the most important features.	Faster- Significantly faster rather tree based models due to lower complexity.	—Proficient with smaller size and non-complex data.	Satisfactory- requires appropriate regularization to overcome overfitting.

2.3.2. ML Algorithm: Extreme Gradient Boosting (XGB)

XGB belongs to the class of ensemble learning techniques that can be used for classification and regression [79]. Gradient boosting, from which XGB is derived, trains sequentially weaker models obtained typically from decision trees which results in a stronger model. It is an

Table 5
Hyperparameters of the RF model.

Hyperparameters	Values	Description
Maximum depth	[5,10,20,40]	The maximum numbers of level in each decision tree
Minimum leaf samples	[1,2,3,4]	The minimal number of samples necessary for a tree's leaf node
Number of estimators	[300, 500, 700 , 900]	It refers to the number of trees in the forest
Minimum split samples	[2,4,8,16]	The minimum number of samples required to split an internal node.

iterative process which enhances the accuracy of the model in each step by rectifying the mistakes in the previous step.

In order to optimize the training process of decision trees, XGB minimizes a loss function that quantifies the discrepancy between the predicted and actual target values. The process of optimization is executed via gradient descent, in which the decision tree parameters are modified based on the gradients of the loss function with regard to the predicted values. The objective function of XGB is expressed as

$$J = \sum L(y_i, \hat{y}_i) + \sum_{i=1}^n \beta[h_i(x, \theta_i)] \quad (3)$$

where $L(y_i, \hat{y}_i)$ is the mean squared loss function between the actual value and predicted value and $\beta[h_i(x, \theta_i)]$ is the regularization term that penalizes complex models to reduce overfitting.

XGB constructs decision trees such that it continuously incorporates trees into the ensemble to reduce the objective function. The flow chart for XGB algorithm is illustrated in Fig. 10. In every iteration of XGB, a new tree is trained to predict the negative gradients of the loss function with regards to the current ensemble predictions. This technique efficiently concentrates on the areas of the feature space where the ensemble is producing the most significant errors, enabling future trees to rectify these errors. The final prediction is obtained by summing all the predictions made by individual trees.

$$\hat{y} = \sum_{i=1}^n h_i(x, \theta_i) \quad (4)$$

2.3.3. ML Algorithm: Multilayer Perceptron (MLP)

MLP is a type of artificial neural network that consists of layers of

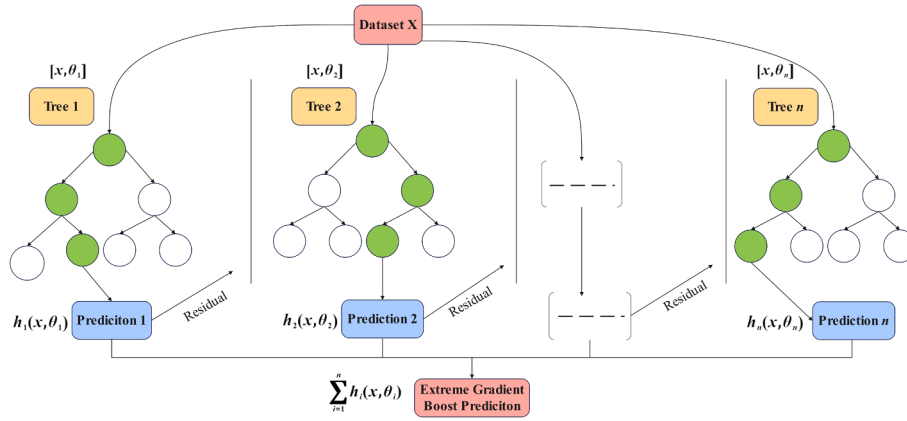


Fig. 10. Extreme Gradient Boost flow chart.

interconnected nodes. A schematic of the MLP model we employed is shown in Fig. 11 (a) where the inputs are mapped to the outputs through a single hidden layer. As shown in Fig. 11 (b), the input feature is forwarded to the intermediate layer using the subsequent propagation rule given by

$$s = \sum x_i w_i + b \quad (5)$$

where x_i and denotes the input features and w_i their corresponding weights. The bias associated with each node is represented by b . Each node in an MLP applies an activation function (z) to the weighted sum of its inputs given by Equation (6), allowing the network to model complex nonlinear relationships between inputs and outputs.

$$z = f(s) = f\left[\sum_i x_i w_i + b\right] \quad (6)$$

The current MLP model uses the following activation functions in the neural network.

$$\tanh = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x)) \quad (7)$$

In an MLP, minimizing the discrepancy between the predicted outputs and the actual targets in the training data requires modifying the weights and biases of the neurons during the learning process. The

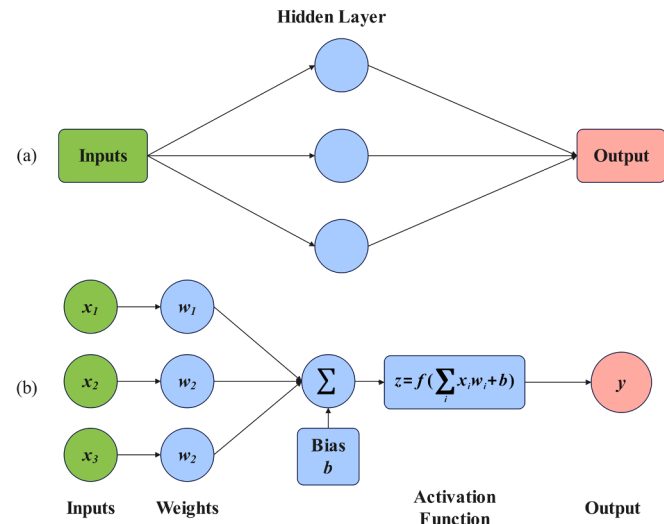


Fig. 11. Multilayer perceptron flow chart with (a) input layer, hidden layer, and output layer and (b) mapping operation at each node between input and output.

learning is carried out using backpropagation using optimization algorithms such as stochastic gradient descent (SGD) or Adaptive Moment Estimation (Adam). During backpropagation, the optimization algorithm compute these gradient of the loss function with regards to the weight and biases. Given n input-output pairs, the loss function is represented as mean squared difference of the desired output (y_i) and the actual output ($\hat{y}_i$) given by Equation (8). The various hyperparameters associated with the MLP are described in Table 7.

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

2.3.4. ML Algorithm: Elastic Net (ENET)

ENET is an extension of the simple linear regression model that combines the lasso (L1) and ridge (L2) regularization to overcome their limitations. Lasso typically yields more concise models; however, its performance is limited by the size of the dataset. In contrast, ridge regression excels at identifying groups of interrelated features but does not achieve the same level of model conciseness as Lasso. ENET is useful when dealing with complex and high dimensional datasets. The cost function contains L1 and L2 penalties alongside the mean squared error (MSE) given by Equation (9).

$$L = \text{MSE}(y, \hat{y}) + \lambda \left[\alpha \|\theta\|_1 + \frac{(1-\alpha)}{2} \|\theta\|_2^2 \right] \quad (9)$$

In the above equation, θ denotes the model parameters, λ is the regularization parameter and α is the mixing parameter between L1 and L2 penalty with value ranging from 0 to 1. Since λ and α are considered hyperparameters (see Table 8), a range of values for each can be tested during the training of the ENET algorithm.

While each of the ML algorithm stated above have their own strengths with regards to model performance and accuracy, it also has certain limitations associated with them. A brief comparison based on memory, speed, data handling and accuracy of the above-mentioned algorithms is provided in Table 4.

Table 6
Hyperparameters of the XGB model.

Hyperparameters	Values	Description
Number of estimators	[25, 50, 100 , 200, 500]	The number of boosting rounds or trees added to the model
Maximum depth	[3,5, 10 ,20]	The maximum numbers of level in each decision tree
Learning rate	[0.03, 0.05, 0.1 , 0.5]	This parameter shrinks the feature weights to make the boosting process more conservative

Table 7
Hyperparameters of the MLP model.

Hyperparameters	Values	Description
Number of hidden layers	[1,2]	Specifies the depth of the neural network
Number of neurons	[3,5,10]	The number of neurons (or nodes) in each hidden layer
Maximum iteration	[200, 500, 700 , 1000]	The maximum number of epochs or iterations over the entire training dataset
Activation function	['tanh', 'ReLU']	It is used to introduce non-linearity into the network, which allows the model to learn more complex patterns
Learning rate	[0.001, 0.01 , 0.1]	It controls the step size at each iteration while moving toward a minimum of a loss function
Optimization algorithm	['sgd', 'adam']	The method used to update weights in the network and minimize the loss function

Table 8
Hyperparameters of the ENET model.

Hyperparameters	Values	Description
Maximum iteration	[1,5,10,20,50]	The maximum number of iterations to be run by the algorithm
Mixing parameter (L1 and L2), α	[0.1, 0.3 , 0.5, 0.7, 0.9, 1]	The mixing parameter α determines the balance between L1 regularization (lasso) and L2 regularization (ridge)
Regularization parameter, λ	[0.1, 0.3, 0.5, 0.7, 0.9 , 1]	It determines the amount of shrinkage, controlling the strength of the regularization applied to the model

2.4. Hyperparameter Tuning

One of the important aspects to consider while building an ML model is hyperparameter tuning. The learning process of the ML algorithm is generally controlled by parameters referred to as hyperparameters. The hyperparameters are not learned from the training dataset, rather they must be explicitly set before training the model. These hyperparameters vary depending on the type of algorithm chosen for model training. The default hyperparameters of a model may not always be optimal since dataset complexity can vary depending on the specific problem. Therefore, tuning or adjusting these hyperparameters is necessary to achieve the best performance. Grid search is one of the methods to evaluate the model performance on all the different possible combinations of hyperparameter specified initially [80]. For a large hyperparameter search space, random search is more effective as it is less computationally expensive. Since our sample set is relatively small, we employed the grid search approach to discover the best hyperparameter settings. The range of different important hyperparameters values for the selected models RF, XGB, MLP and ENET and their description are presented in Table 5, Table 6, Table 7, and Table 8, respectively, with the best hyperparameters values, evaluated based on the performance metrics, marked in bold.

2.5. Evaluation Metrics

Three evaluation metrics, namely the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) were used to evaluate the model performance. The coefficient of determination, denoted by R^2 , evaluates the variability between the measured outputs and the predicted outputs². Its value ranges from 0 to 1, where 0 indicates that the variability in outputs cannot be correlated to the input, while 1 indicates that the variability in the output is clearly correlated to the inputs. R^2 is calculated using the following equation.

$$R^2 = 1 - \frac{SSE}{SST} \quad (10)$$

where SSE denotes the squared difference sum of the actual output (y_i) and the predicted output ($\hat{y}_i$) as represented by Equation (11), while SST denotes the squared difference sum of the actual output (y_i) and the mean of the actual output ($\bar{y}$) (see Equation (12)).

$$SSE = \sum (y_i - \hat{y}_i)^2 \quad (11)$$

$$SST = \sum (y_i - \bar{y})^2 \quad (12)$$

Similarly, MAE refers to the average of the absolute differences between the actual and predicted output values. A value close to 0 indicates minimal error. Equation (13) represents MAE, where N denotes the total number of sample points.

$$MAE = \frac{1}{N} \sum |y_i - \hat{y}_i| \quad (13)$$

Lastly, RMSE refers to the root of mean squared difference between the actual and predicted value indicated by Equation (14). A value close to 0 indicates minimal error.

$$RMSE = \sqrt{\frac{1}{N} \sum (y_i - \hat{y}_i)^2} \quad (14)$$

3. Results and Discussion

3.1. Model Performance

The performance of each of the models was examined based on the evaluation metrics R^2 , MAE, and RMSE. RF, having $R^2 = 0.830$, MAE = 0.653, and RMSE = 1.774, and XGB, having $R^2 = 0.884$, MAE = 0.651, and RMSE = 1.571, outperformed the other models as evident from their performance metrics plots (see Fig. 12). Conversely, while MLP that has $R^2 = 0.841$, MAE = 0.714, and RMSE = 1.843 performed better than the ENET model ($R^2 = 0.473$, MAE = 3.631, RMSE = 4.175), its performance was lower than RF and XGB. The R^2 value for the XGB model was the highest with a value of 0.884, indicating that the ML algorithm captured the variability in the dataset sufficiently well. The performance of the models was enhanced after tuning the model hyperparameters with only a slight increase in the accuracy of RF and XGB models. However, significant differences were observed in the R^2 , MAE, and RMSE values between the default and tuned MLP model as shown in Fig. 12.

The overall performance of MLP being lower than the tree-based models can be attributed to potential overfitting in tabular dataset with limited instances. Additionally, MLP is designed to capture highly nonlinear relationships in huge datasets. ENET being a simple linear model encompassing the regularization parameters is unable to learn the non-linearity available in the dataset. Given that MLP and ENET are sensitive to model hyperparameters, their performance is enhanced through hyperparameter tuning.

It is evident from the model performance that the tuned XGB model is able to predict the spray characteristics with good accuracy within the tested range of operating conditions. The training loss curve (see Fig. 13) demonstrates that the dataset size is sufficiently large for training, thereby providing accurate predictions of spray tip penetration and spray angle. To quantitatively compare the predicted spray characteristics with the experimental values, the spray tip penetration and spray angle predicted by different models are plotted in comparison with the experimental data for two cases. Fig. 14 represents the ML models prediction and experimental plots of spray tip penetration and spray angle using injector C, Hole 1 at injection pressure of 120 MPa and chamber pressure of 2 MPa, while Fig. 15 shows the results plotted for injector B, Hole 2 at injection pressure of 150 MPa and chamber pressure of 0.5

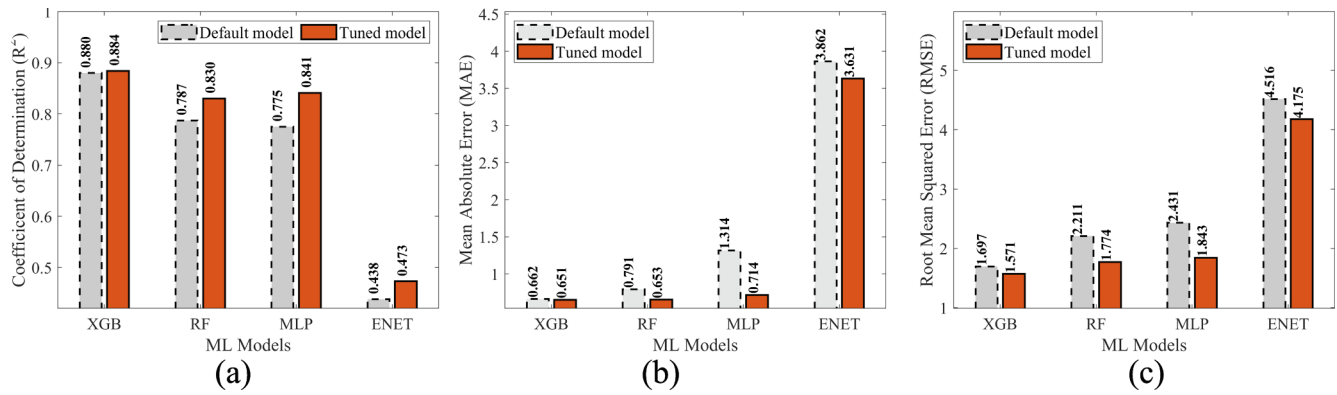


Fig. 12. Model performance (a) R^2 , (b) MAE, and (c) RMSE of tested ML models with default and tuned hyperparameters.

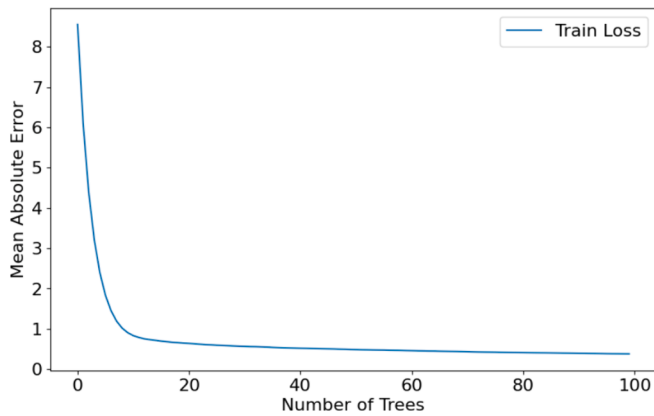


Fig. 13. Training loss curve of XGB model based on number of trees.

MPa. The figures show that the dependencies behind the spray development from the start of injection till the complete spray formation can be learned using the XGB model, having the best performance. While RF model is able to follow the spray tip penetration and spray angle curve similar to XGB model, ENET model shows significant deviation from the actual experimental values. Additionally, MLP model follows the trends of the curves but is not accurate enough in predicting the spray tip penetration and spray angle. As such, the current XGB model can be generalized to other similar datasets by fine tuning the model [81], which is also supported by the literature [79]. This allows for a pathway

to extend the ML model training to a wider range of operating conditions in order to optimize spray formation in an effort to improve engine performance and reduce engine out emissions.

Performance metrics using RMSE are provided in Table 9 for Injector A operating under various injection and chamber pressures. While the overall model performance indicates that XGB is more accurate compared to other models, the RF model shows superior performance under certain conditions. For example, at an injection pressure (IP) of 60 MPa and a chamber pressure (CP) of 2 MPa, the RMSE for spray angle with RF is 0.133, whereas it is 0.566 with XGB. Under the same conditions, the RMSE for spray tip penetration is 0.212 for XGB and 0.159 for RF. Additionally, the RMSE values for spray angle with the ENET model are significantly lower than those for spray tip penetration. This discrepancy can be attributed to the lower variation in spray angle compared to spray tip penetration as the spray develops, as observed in Fig. 14 and Fig. 15.

The spray tip penetration for chamber pressure ranging from 0.1 MPa to 2 MPa at a fixed injection pressure of 150 MPa is illustrated in Fig. 16 (a). The XGB model accurately captures the initial spray penetration distances across various chamber pressures, indicating its alignment with the physics of spray formation, which is initially driven by momentum. Similar observations are made after the spray breakup time, where the XGB model demonstrates that a strong correlation exists between spray formation and chamber pressure, attributed to mixture formation and air entrainment. Furthermore, Fig. 16 (b) shows spray penetration curves at chamber pressure of 2 MPa for injection pressures ranging from 60 MPa to 150 MPa. At early injection times, penetration curves are similar for different injection pressures and diverged later.

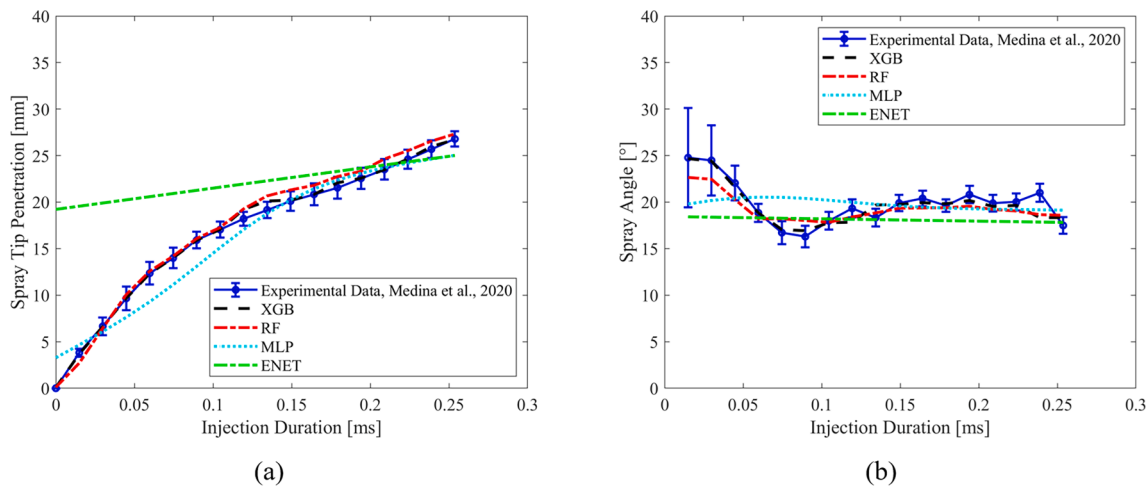


Fig. 14. (a) Spray tip penetration distance and (b) spray angle as function of time compared between experimental data [24] and hyperparameter tuned ML models using injector C, hole 1 at injection pressure-120 MPa and chamber pressure-2 MPa.

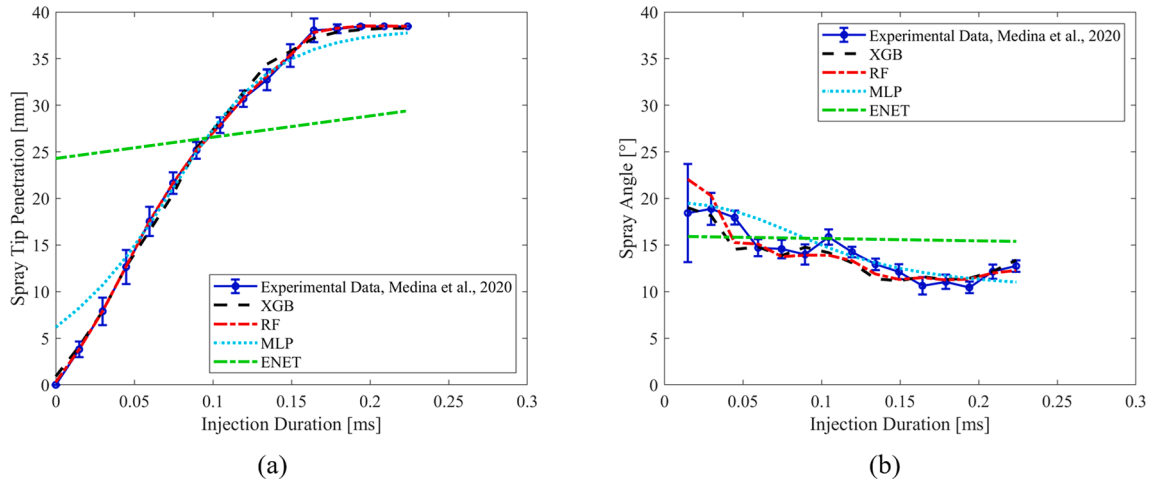


Fig. 15. (a) Spray tip penetration distance and (b) spray angle as function of time compared between experimental data [24] and hyperparameter tuned ML models using injector B, hole 2 at injection pressure-150 MPa and chamber pressure-0.5 MPa.

Table 9

Comparative Analysis of RMSE metric values for varying combination of injection pressures and chamber pressure for injector A, hole 1.

	θ [degree]	S [mm]	θ [degree]	S [mm]	θ [degree]	S [mm]	θ [degree]	S [mm]
	IP-60 MPa, CP-0.1 MPa		IP-60 MPa, CP-0.5 MPa		IP-60 MPa, CP-1 MPa		IP-60 MPa, CP-2 MPa	
XGB	0.185	0.474	0.157	0.091	0.514	0.141	0.566	0.212
RF	0.584	0.481	0.268	0.124	0.327	0.181	0.133	0.159
MLP	1.886	1.739	1.445	1.201	1.455	0.948	2.434	1.723
ENET	2.364	8.712	1.667	7.969	1.914	6.479	2.477	5.236
	IP-90 MPa, CP-0.1 MPa		IP-90 MPa, CP-0.5 MPa		IP-90 MPa, CP-1 MPa		IP-90 MPa, CP-2 MPa	
XGB	0.085	0.243	0.135	0.132	0.159	0.149	0.064	0.135
RF	0.242	0.374	0.211	0.295	0.366	0.224	0.292	0.245
MLP	1.477	1.171	1.722	1.039	1.675	1.453	1.988	1.159
ENET	1.755	8.449	1.856	7.761	2.146	6.556	2.456	5.716
	IP-120 MPa, CP-0.1 MPa		IP-120 MPa, CP-0.5 MPa		IP-120 MPa, CP-1 MPa		IP-120 MPa, CP-2 MPa	
XGB	0.183	0.163	0.192	0.251	0.836	0.209	0.133	0.118
RF	0.327	0.348	0.408	0.26	0.366	0.277	0.178	0.226
MLP	1.223	1.048	2.092	1.165	2.051	1.849	1.986	0.875
ENET	1.609	8.031	2.279	7.594	2.688	7.023	2.411	6.124
	IP-150 MPa, CP-0.1 MPa		IP-150 MPa, CP-0.5 MPa		IP-150 MPa, CP-1 MPa		IP-150 MPa, CP-2 MPa	
XGB	0.076	0.151	0.067	0.128	0.112	0.215	0.972	0.273
RF	0.206	0.618	0.261	0.322	0.248	0.282	0.425	0.372
MLP	1.249	1.218	1.512	1.581	1.775	1.815	2.018	1.337
ENET	1.421	8.051	1.788	7.485	1.921	7.443	2.573	6.136

Fig. 16 (a) and Fig. 16 (b) indicate that XGB can inherently capture the physical trends of spray development, which after breakup is predominantly controlled by chamber pressure—a similar observation to those made earlier using diesel sprays [82].

Since computational cost is an important matter to consider in modelling, its evaluation becomes important. The ML modelling time in this study was significantly low. For example, the hyperparameter-tuned XGB model was compiled and evaluated in 422 s on a CPU with 12 processors, in contrast to the several hours of computational time typically required for numerical spray studies using high-performance computing. Similar observations were made by Hwang et al. [74] in their study. The computational time associated with each machine learning model following the application of hyperparameter tuning is presented at the end in Fig. 17.

3.2. XGB Model Interpretability

To interpret the XGB model, the importance of each feature contributing to the predictions in spray tip penetration and spray angle is presented in Fig. 18. The feature importance values are calculated based on the tree node split across all trees by each feature. As the spray progresses and evolves over time, the injection duration demonstrates a substantial impact on predicting both the spray tip penetration and spray angle. Furthermore, in line with the findings from experimental spray studies by Medina et al. [21] and previous studies on diesel sprays [82], ML spray modelling observations show that the chamber pressure is more dominant than injection pressure in characterizing the spray in terms of spray angle and spray tip penetration. Finally, the nozzle outlet diameter and conicity have relatively lower contribution in predicting

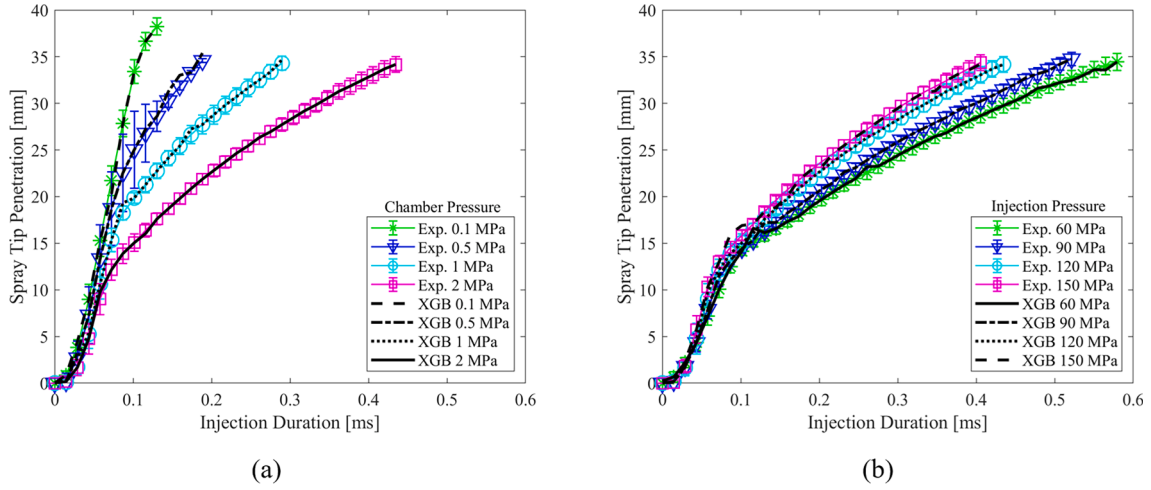


Fig. 16. Spray tip penetration for (a) fixed injection pressure:150 MPa and chamber pressures:0.1–2 MPa (b) fixed chamber pressure:2MPa and injection pressures-600:150 MPa for injector A, hole 1.

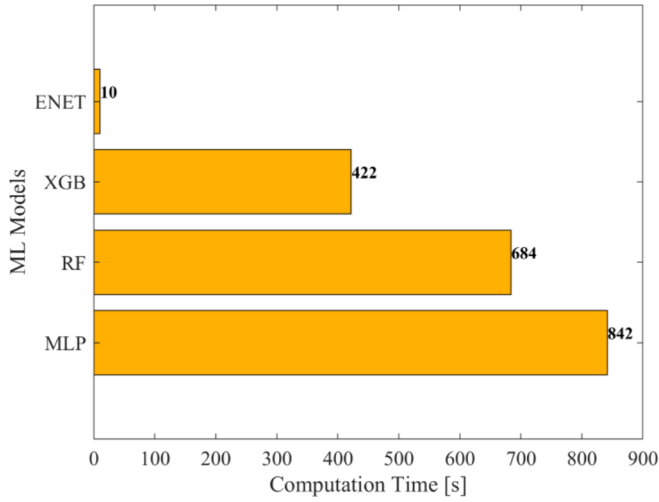


Fig. 17. Computation time corresponding to each of the hyperparameter tuned ML models.

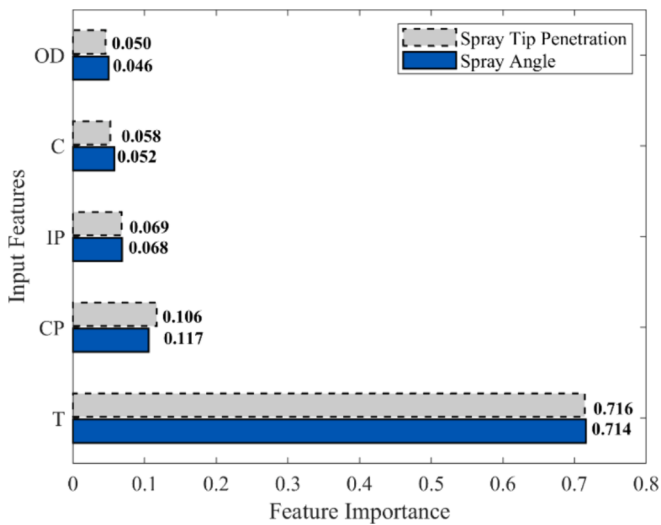


Fig. 18. Feature importance plot illustrating the contribution of each input feature in predicting the spray angle and spray tip penetration.

the spray characteristics.

3.3. XGB Model Comparison with Empirical Correlations

Several studies have been carried out to characterize sprays using empirical correlations for spray tip penetration based on the fluid properties and operating conditions. Some of the parameters on which these correlations depend include injector diameter (d), difference of injection pressure and chamber pressure (ΔP), density of fluid (ρ_f), and density of chamber fluid (ρ_c). Hiroyasu and Arai [83] derived a correlation for diesel fuel which is given by Equations (15–17).

$$S(t) = 0.39 \sqrt{\frac{2\Delta P}{\rho_f}} t \quad 0 < t < t_b \quad (15)$$

$$S(t) = 2.95 (\Delta P / \rho_c)^{1/4} \sqrt{dt} \quad t > t_b \quad (16)$$

$$t_b = 28.65 \sqrt{\frac{\rho_f d}{\rho_c \Delta P}} \quad (17)$$

In the above equations, t_b denotes the spray break up time. The correlations above are in SI units.

Tian et al. [64] defined spray tip penetration correlation for high pressure gasoline fuels with no breakup time under non-evaporating (Equation (18)) and evaporating (Equation (19)) conditions. The correlation contained parameters similar to the ones used for diesel fuel correlation development except for the non-evaporating condition in which ambient chamber temperature was also included.

$$S_{non-evap}(t) = 0.0635 \rho_c^{-0.25} d^{0.5} t^{1.033} \Delta P^{0.71} \quad (18)$$

$$S_{evap}(t) = 8.019 \rho_c^{-0.319} d^{0.561} t^{0.462} \Delta P^{0.212} (294/T_c)^{0.693} \quad (19)$$

In the above expression, ΔP is given in kPa, d in mm, t in ms and ρ_c in g/cm³. Here T_c represents the chamber temperature in K.

A comparison of the best ML model of the current study, i.e., the tuned XGB model, with the experimental data [21] and the spray correlations for spray tip penetration is depicted in Fig. 19. As seen in Fig. 19, the XGB model, which has an R^2 value of 0.884, MAE value of 0.651, and RMSE value of 1.571, agrees well with the experimental data. The spray penetration correlation developed by Tian et al., lacking consideration for spray breakup, results in significant deviations of penetration distance values from the actual ones. On the other hand, the penetration distance equations derived by Hiroyasu and Arai align

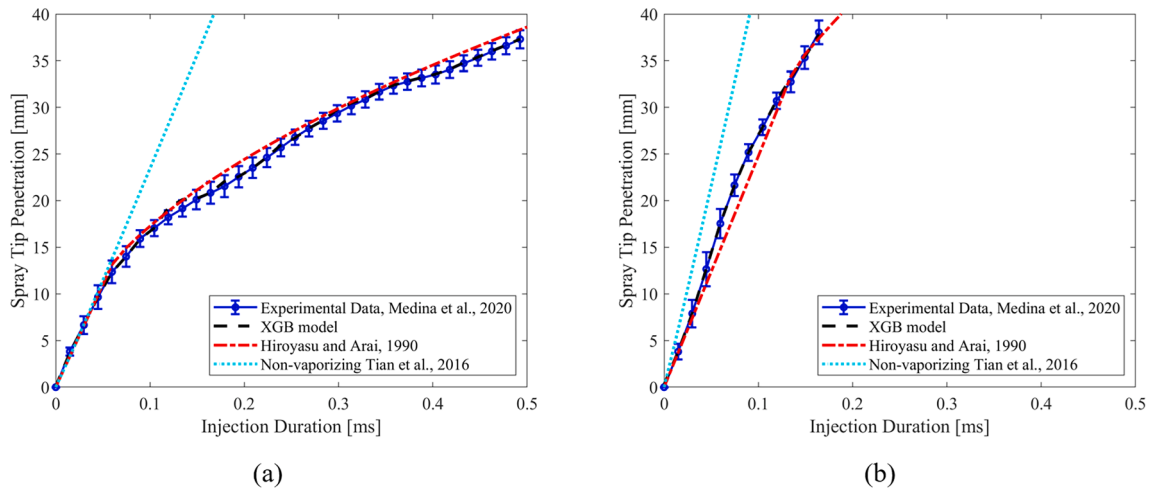


Fig. 19. Spray tip penetration distance as function of time compared between spray correlations, experimental data [24], and the XGB model (a) using injector C, hole 1 at injection pressure-120 MPa and chamber pressure-2 MPa (b) using injector B, hole 2 at injection pressure-150 MPa and chamber pressure-0.5 MPa.

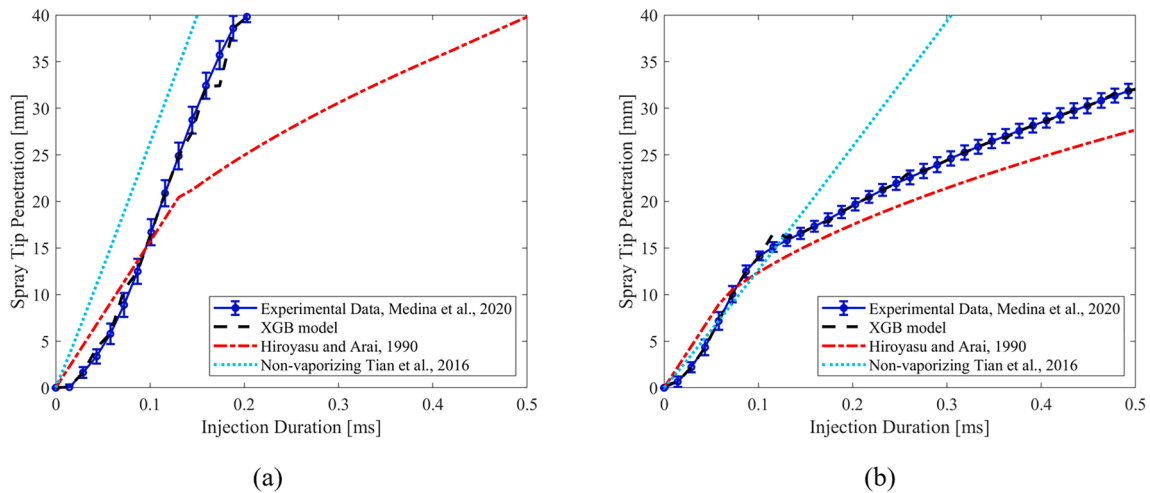


Fig. 20. Spray tip penetration distance as function of time compared between spray correlations, experimental data [24], and the XGB model using injector A, hole 1 at injection pressure-60 MPa and (a) chamber pressure-0.1 MPa (b) chamber pressure-2 MPa.

relatively well with the experimental results. However, it is observed that these correlations for spray penetration underperform at chamber pressure of 0.1 MPa, especially after spray breakup, as depicted in

Fig. 20 (a) and Fig. 21 (a). Furthermore, it should be emphasized that while these correlations were originally derived from fundamental principles, they were validated exclusively with datasets from diesel

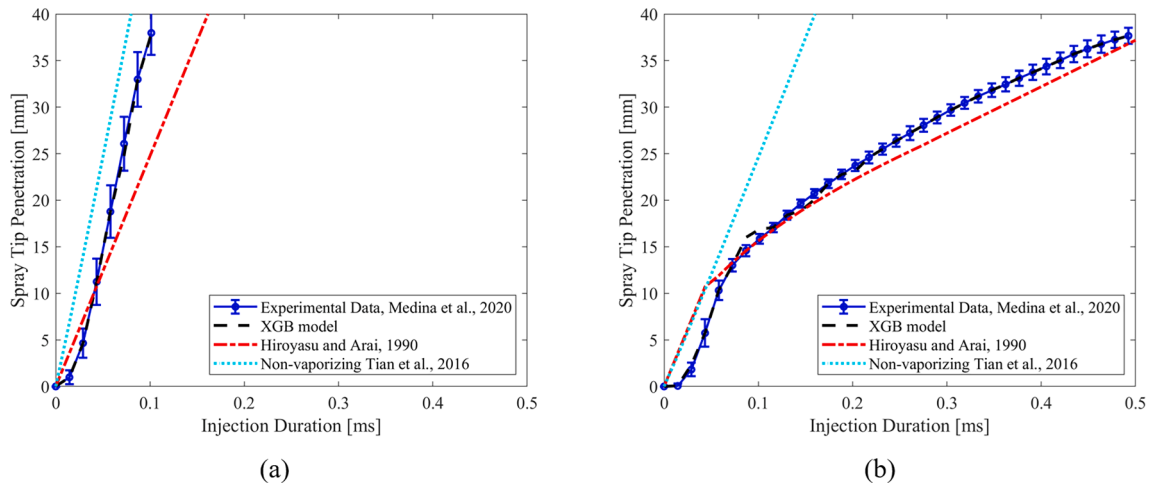


Fig. 21. Spray tip penetration distance as function of time compared between spray correlations, experimental data [24], and the XGB model using injector A, hole 1 at injection pressure-150 MPa and (a) chamber pressure-0.1 MPa (b) chamber pressure-2 MPa.

fuels. This specificity limits their broader applicability to various engine operational conditions. The enhanced accuracy by the ML model indicates that ML algorithms are a useful tool that can be used to decouple non-linear behaviours from the physical processes to improve prediction.

4. Conclusion

In this study, machine learning algorithms were leveraged to predict the characteristics of high-pressure gasoline sprays. The ML algorithms were trained using experimental imaging data from previous studies. Initial analyses revealed weak linear correlations between operating conditions, injector geometry, and spray features. As such, three non-linear and one linear ML models were used: random forest, extreme gradient boosting (XGB), multilayer perceptron (MLP), and elastic net, respectively. Based on the coefficient of determination, mean absolute error, and root mean squared error, the tree-based models (random forest and XGB) demonstrated superior performance over elastic net and MLP. This indicates the strength of traditional machine learning algorithms for nonlinear datasets over deep neural networks. Furthermore, the XGB model's ability to decouple nonlinear behaviours underscores its utility to identifying physical trends in spray formation, enhancing predictive accuracy and reducing computational demands.

While data-driven ML methods can accurately predict spray characteristics, their performance may be hindered by the lack of sufficiently high-quality data. Additionally, generalizing the model to similar datasets may require fine-tuning to ensure accurate predictions. To overcome this limitation and improve the models, training data can include instantaneous spray images and their tabular dataset. Another avenue of future work can include physics-driven ML approaches to derived spray correlations such as penetration distance and spray breakup time. The application of ML in fuel injector design has the potential to revolutionize engine performance and contribute to substantial reductions in emissions. This work underscores the transformative impact of ML in advancing the optimization of fuel injection systems, leading to improved engine efficiency and environmental sustainability.

CRedit authorship contribution statement

Sadique Khan: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Mudassir Masood:** Writing – review & editing, Supervision, Methodology, Investigation. **Mario Medina:** Writing – review & editing, Supervision, Investigation, Data curation. **Fahad Alzahrani:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

The data used in this study are available from the corresponding authors upon reasonable request.

Acknowledgements

This work was supported by the Interdisciplinary Research Center for Hydrogen Technologies and Carbon Management at King Fahd University of Petroleum and Minerals [Project No. INHE2209], and it has been (partially) supported by the National Science Foundation, USA [Award No. HRD-2112554].

References

- [1] Energy Information Administration, "International Energy Outlook," Washington, DC, ISBN 2025869592, 2019.
- [2] IEA (2022), Transport, IEA, Paris, License: CC BY 4.0, ([cit. 2023-05-19]).
- [3] Internal Combustion Engine (ICE) Market Growth by End-user, Type, and Geography - Forecast and Analysis 2023-2027, Technavio.
- [4] Regulation (EU) 2023/851 of the European Parliament and of the Council of 19 April 2023 amending Regulation (EU) 2019/631 as regards strengthening the CO₂ emission performance standards for new passenger cars and new light commercial vehicles in line with the Union's increased climate ambition.
- [5] Lavoie GA, Ortiz-Soto E, Babajimopoulos A, Martz JB, Assanis DN. Thermodynamic sweet spot for high efficiency, dilute, boosted gasoline engines. *Int J Engine Res* 2013;14(3).
- [6] Kalghatgi GT, Risberg P, Ångström HE. Advantages of fuels with high resistance to auto-ignition in late-injection, low-temperature, compression ignition combustion. In: SAE Technical Papers 2006.
- [7] Mofijur M, Hasan MM, Mahlia TMI, Rahman SMA, Silitonga AS, Ong HC. Performance and emission parameters of homogeneous charge compression ignition (HCCI) engine: a review. *Energies* 2019;12.
- [8] Shaik A, Moorthi NSV, Rudramoorthy R. Variable compression ratio engine: A future power plant for automobiles - An overview. *Proc Inst Mech Eng Part D J Automobile Eng* 2007;221(9).
- [9] Dec JE. Advanced compression-ignition engines - Understanding the in-cylinder processes. *Proceedings of the Combustion Institute*. 2009;32 II.
- [10] Chang K, Lavoie GA, Babajimopoulos A, Filipi ZS, Assanis DN. Control of a multi-cylinder HCCI engine during transient operation by modulating residual gas fraction to compensate for wall temperature effects. In: SAE Technical Papers 2007.
- [11] Dec JE, Sjöberg M. Isolating the effects of fuel chemistry on combustion phasing in an HCCI engine and the potential of fuel stratification for ignition control. In: SAE Technical Papers 2004.
- [12] Stanglmaier RH, Roberts CE. Homogeneous charge compression ignition (HCCI): Benefits, compromises, and future engine applications. In: SAE Technical Papers 1999.
- [13] Dempsey AB, Curran S, Wagner R, Cannella W. Effect of premixed fuel preparation for partially premixed combustion with a low octane gasoline on a light-duty multicylinder compression ignition engine. *J Eng Gas Turbine Power* 2015;137(11).
- [14] Kalghatgi GT, Risberg P, Ångström HE. Partially pre-mixed auto-ignition of gasoline to attain low smoke and low NO_x at high load in a compression ignition engine and comparison with a diesel fuel. In: SAE Technical Papers 2007.
- [15] Chincholkar SP, Suryawanshi JG. Gasoline direct injection: an efficient technology. *Energy Procedia* 2016 Dec;90:666–72.
- [16] Kaario OT, Vuorinen V, Kahila H, Im HG, Larmi M. The effect of fuel on high velocity evaporating fuel sprays: Large-Eddy simulation of Spray A with various fuels. *Int J Engine Res* 2020 Jan 19;21(1):26–42.
- [17] Montanaro A, Malaguti S, Alfuso S. Wall impingement process of a multi-hole GDI spray: experimental and numerical investigation. In 2012.
- [18] Gao J, Jiang D, Huang Z. Spray properties of alternative fuels: a comparative analysis of ethanol-gasoline blends and gasoline. *Fuel* 2007 Jul;86(10–11):1645–50.
- [19] Manin J, Jung Y, Skeen SA, Pickett LM, Parrish SE, Markle L. Experimental characterization of DI gasoline injection processes. In 2015.
- [20] Song H, Xiao J, Chen Y, Huang Z. The effects of deposits on spray behaviors of a gasoline direct injector. *Fuel* 2016 Sep;180:506–13.
- [21] Medina M, Fatouraie M, Wooldridge M. High-Speed Imaging Studies of Gasoline Fuel Sprays at Fuel Injection Pressures from 300 to 1500 bar. In: SAE Technical Papers 2018.
- [22] Duronio F, De Vita A, Allocca L, Anatone M. Gasoline direct injection engines – A review of latest technologies and trends. Part 1: Spray breakup process. *Fuel* 2020 Apr;265:116948.
- [23] Lefebvre AH, McDonell VG. Atomization and sprays. CRC Press; 2017.
- [24] Medina M, Bautista A, Wooldridge M, Payri R. The effects of injector geometry and operating conditions on spray mass, momentum and development using high-pressure gasoline. *Fuel* 2021 Jun;294:120468.
- [25] Shervani-Tabar MT, Sheykhvazayefi M, Ghorbani M. Numerical study on the effect of the injection pressure on spray penetration length. *Appl Math Model* 2013 Aug;37(14–15):7778–88.
- [26] Saha K, Som S, Battistoni M, Li Y, Pomraning E, Senecal PK. Numerical investigation of two-phase flow evolution of in- and near-nozzle regions of a gasoline direct injection engine during needle transients. *SAE Int J Engines* 2016 Apr 5;9(2). 2016–01–0870.
- [27] Mouvanal S, Burkhardt A, Bakshi S, Chatterjee D. Numerical study of purging of a gasoline direct injection nozzle at the end of injection. *Int J Engine Res* 2021 May 4;22(5):1670–84.
- [28] Tu PW, Xu H, Srivastava DK, Dean K, Jing D, Cao L, et al. Numerical investigation of GDI injector nozzle geometry on spray characteristics. In 2015.
- [29] Jiang C, Parker MC, Helie J, Spencer A, Garner CP, Wigley G. Impact of gasoline direct injection fuel injector hole geometry on spray characteristics under flash boiling and ambient conditions. *Fuel* 2019 Apr;241:71–82.
- [30] Li X, Cheng Y, Ji S, Yang X, Wang L. Sensitivity Analysis of fuel injection characteristics of GDI injector to injector nozzle diameter. *Energies (Basel)* 2019 Jan 30;12(3):434.

- [31] Lee Z, Kim T, Park S, Park S. Review on spray, combustion, and emission characteristics of recent developed direct-injection spark ignition (DISI) engine system with multi-hole type injector. *Fuel* 2020 Jan;259:116209.
- [32] Martí-Aldaraví, P. "Development of a computational model for a simultaneous simulation of internal flow and spray break-up of the Diesel injection process". PhD thesis. Valencia: Universitat Politècnica de València, 2014. doi: 10.4995/Thesis/10251/43719.
- [33] Salvador FJ, Martínez-López J, Caballer M, De Alfonso C. Study of the influence of the needle lift on the internal flow and cavitation phenomenon in diesel injector nozzles by CFD using RANS methods. *Energy Convers Manag* 2013 Feb;66:246–56.
- [34] De La Morena Borja J. ESTUDIO DE LA INFLUENCIA DE LAS CARACTERÍSTICAS DEL FLUJO INTERNO EN TOBERAS SOBRE EL PROCESO DE INYECCIÓN DIÉSEL EN CAMPO PRÓXIMO. [Valencia (Spain)]: Universitat Politècnica de València; 2011.
- [35] Suh HK, Lee CS. Effect of cavitation in nozzle orifice on the diesel fuel atomization characteristics. *Int J Heat Fluid Flow* 2008 Aug;29(4):1001–9.
- [36] Kawano D, Ishii H, Suzuki H, Goto Y, Odaka M, Senda J. Numerical study on flash-boiling spray of multicomponent fuel. *Heat Transfer—Asian. Research* 2006 Jul;35(5):369–85.
- [37] Senda J, Wada Y, Kawano D, Fujimoto H. Improvement of combustion and emissions in diesel engines by means of enhanced mixture formation based on flash boiling of mixed fuel. *Int J Engine Res* 2008 Feb 1;9(1):15–27.
- [38] He X, Li Y, Liu C, Sjöberg M, Vuilleumier D, Liu F, et al. Characteristics of spray and wall wetting under flash-boiling and non-flashing conditions at varying ambient pressures. *Fuel* 2020 Mar;264:116683.
- [39] Lefebvre AH, et al. *Atomization and Sprays*. 2nd Ed. Boca Raton, FL: CRC Press; 2017.
- [40] Ashgriz N, Poo JY. Coalescence and separation in binary collisions of liquid drops. *J Fluid Mech* 1990 Dec;26(221):183–204.
- [41] Reitz RD, et al. Structure of high-pressure fuel sprays. In: *SAE Paper 1987;870598*.
- [42] *Mixture Formation in Internal Combustion Engine*. Berlin/Heidelberg: Springer-Verlag; 2006.
- [43] Jin, J. D. et al. "A Model for Multicomponent Droplet Vaporization at High Ambient Pressures". In: *SAE Technical Papers* (1985).
- [44] Abramzon B, Sirignano W. Droplet vaporization model for spray combustion calculations. In: *26th Aerospace Sciences Meeting*. Reston, Virginia: American Institute of Aeronautics and Astronautics; 1988.
- [45] Manin J, Jung Y, Skeen SA, Pickett LM, Parrish SE, Markle L. Experimental Characterization of DI Gasoline Injection Processes. In 2015.
- [46] Duke DJ, Kastengren AL, Matusik KE, Swantek AB, Powell CF, Payri R, et al. Internal and near nozzle measurements of Engine Combustion Network "Spray G" gasoline direct injectors. *Exp Therm Fluid Sci* 2017 Nov;88:608–21.
- [47] Shahangian, N. et al. "Spray orientation assessment and correction method for gdi momentum flux measurements". In: October (2019), pp. 231–241.
- [48] Payri R, Gimeno J, Martí-Aldaraví P, Vaquerizo D. Momentum Flux Measurements on an ECN GDI injector. In 2015.
- [49] Costa M, Allocca L, Montanaro A, Sorge U, Iorio B. Multiple Injection in a Mixed Mode GDI Boosted Engine. In 2010.
- [50] Hoffmann G, Befrui B, Berndorfer A, Piock WF, Varble DL. Fuel System Pressure Increase for Enhanced Performance of GDI Multi-Hole Injection Systems. *SAE Int J Engines*. 2014 Apr 1;7(1):2014-01–1209.
- [51] Merola SS, Irimescu A, Tornatore C, Marchitto L, Valentino G. Split injection in a DISI engine fuelled with butanol and gasoline analyzed through integrated methodologies. *SAE Int J Engines* 2015;8(2).
- [52] Gawlica T, Samenfink W, Schünemann E, Koch T. Model-based optimization of multi-hole injector spray targeting for gasoline direct injection. In 2019.
- [53] Singh R, Burch T, Lavoie G, Wooldridge M, Fatouraie M. Effects of Fuel Injection Events of Ethanol and Gasoline Blends on Boosted Direct-Injection Engine Performance. In: *SAE Technical Papers*. 2017.
- [54] Bornschlegel S, Conrad C, Durst A, Wang J, Wensing M. Multi-hole gasoline direct injection: In-nozzle flow and primary breakup investigated in transparent nozzles and with X-ray. *Int J Engine Res* 2018 Jan 12;19(1):67–77.
- [55] Chaves H, Donath S. Binary cavitation in a transparent three hole GDI nozzle. In: *Proceedings ILASS-Europe 2017 28th Conference on Liquid Atomization and Spray Systems*. Valencia: Universitat Politècnica València; 2017.
- [56] Greif D, Struel J. Numerical Study of Transient Multi Component Fuel Injection. In 2013.
- [57] Mishra R, Jiwani D, Jarrahbashi D. Three-component multi-fluid modeling of pseudo-cavitation phenomenon in diesel injector nozzles. *Int J Engine Res* 2022 Apr 19;23(4):591–613.
- [58] Moulai M, Grover R, Parrish S, Schmidt D. Internal and Near-Nozzle Flow in a Multi-Hole Gasoline Injector Under Flashing and Non-Flashing Conditions. In 2015.
- [59] Shost MA, Lai MC, Befrui B, Spiekermann P, Varble DL. GDI Nozzle Parameter Studies Using LES and Spray Imaging Methods. In 2014.
- [60] Befrui B, Aye A, Spiekermann P, Varble DL, Shost MA, Lai MC, et al. GDI Skew-Angled Nozzle Flow and Near-Field Spray Analysis using Optical and X-Ray Imaging and VOF-LES Computational Fluid Dynamics. In 2013.
- [61] Wang B, Badawy T, Hutchins P, Tu P, Xu H, Zhang X. Numerical Investigation of the Deposit Effect on GDI Injector Nozzle Flow. *Energy Procedia* 2017 May;105:1671–6.
- [62] Han Z, Fan L, Reitz RD. Multidimensional modeling of spray atomization and air-fuel mixing in a direct-injection spark-ignition engine. In: *SAE Technical Papers* 1997.
- [63] Cheng AS "Ed," Ratcliff MA, McCormick RL. Modeling ethanol-blend fuel sprays under direct-injection spark-ignition engine conditions. *Energy Fuels*. 2023 Feb 16; 37(4):3031–47.
- [64] Tian J, Zhao M, Long W, Nishida K, Fujikawa T, Zhang W. Experimental study on spray characteristics under ultra-high injection pressure for DISI engines. *Fuel* 2016;186.
- [65] Nowruzi H, Ghassemi H, Amini E, Sohrabi-asl I. Prediction of impinging spray penetration and cone angle under different injection and ambient conditions by means of CFD and ANNs. *J Braz Soc Mech Sci Eng* 2017;39(10).
- [66] Nobari AH, Khorasani-Gerdekouhi F, Gulam N, Ashgriz N. assigned by the OpenConf System) Application of Deep Learning Convolutional Neural Network for Spray Characterization. 2020.
- [67] Liu Y, Tian J, Song Z, Li F, Zhou W, Lin Q. Spray characteristics of diesel, biodiesel, polyoxymethylene dimethyl ethers blends and prediction of spray tip penetration using artificial neural network. *Phys Fluids* 2022;34(1). Jan 1.
- [68] Jeyaseelan T, Son M, Sander T, Zigan L. Experimental and modeling analysis of the transient spray characteristics of cyclopentane at sub- and transcritical conditions using a machine learning approach. *Phys Fluids* 2023 Aug 1;35(8).
- [69] Chang M, Jeong M, Park S, Kim HI, Park JH, Park S. Study on predictions of spray target position of gasoline direct injection injectors with multi-hole using physical model and machine learning. *Fuel Process Technol* 2023 Aug;247:107774.
- [70] Chang M, Park S. Predictions and analysis of flash boiling spray characteristics of gasoline direct injection injectors based on optimized machine learning algorithm. *Energy* 2023 Jan;262:125304.
- [71] Koukouvinis P, Rodriguez C, Hwang J, Karathanassis I, Gavaises M, Pickett L. Machine Learning and transcritical sprays: a demonstration study of their potential in ECN Spray-A. *Int J Engine Res* 2022 Sep 1;23(9):1556–72.
- [72] Oh H, Hwang J, Pickett LM, Han D. Machine-learning based prediction of injection rate and solenoid voltage characteristics in GDI injectors. *Fuel* 2022;311.
- [73] Hwang J, Lee P, Mun S, Karathanassis IK, Koukouvinis P, Pickett LM, et al. Machine-learning enabled prediction of 3D spray under engine combustion network spray G conditions. *Fuel* 2021 Jun;1:293.
- [74] Hwang J, Lee P, Mun S, Karathanassis IK, Koukouvinis F, Tagliante F, et al. A new pathway for prediction of gasoline sprays using machine-learning algorithms. *SAE Int J Adv Curr Pract Mobil* 2022 Mar 29;5(1). 2022–01–0492.
- [75] Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Van LH, Tran VQ, et al. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Math Probl Eng* 2021 Feb;5(2021):1–15.
- [76] Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? 2022 Jul 18; Available from: <http://arxiv.org/abs/2207.08815>.
- [77] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. 2012 Jan 2.
- [78] Breiman L. Random Forests. *Mach Learn [Internet]*. 2001;45(1):5–32. Available from: DOI: 10.1023/A:1010933404324.
- [79] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Mar 8; Available from: <http://arxiv.org/abs/1603.02754>.
- [80] Feurer M, Hutter F. Hyperparameter Optimization. In: Hutter F, Kotthoff L, Vanschoren J, editors. *Automated Machine Learning: Methods, Systems, Challenges [Internet]*. Cham: Springer International Publishing; 2019. p. 3–33. Available from: DOI: 10.1007/978-3-030-05318-5_1.
- [81] Sun X, Fu J, Zhou F, Luo B, Liu J. Performance prediction and multi-objective optimization for the Atkinson cycle engine using eXtreme Gradient Boosting. *Therm Sci Eng Prog* 2024 Feb;48:102402.
- [82] Naber J, Siebers DL. Effects of gas density and vaporization on penetration and dispersion of diesel sprays. In 1996.
- [83] Hiroyasu H, Arai M. Structures of fuel sprays in diesel engines. Available from SAE Transactions [Internet] 1990;99:1050–61. <http://www.jstor.org/stable/44548562>.