

Received October 18, 2020, accepted October 28, 2020, date of publication November 2, 2020, date of current version November 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035318

Prediction of Drag Force on Vehicles in a Platoon Configuration Using Machine Learning

FARWA JAFFAR¹, TAHA FARID², MUHAMMAD SAJID^{1,2},
YASAR AYAZ^{1,3}, (Senior Member, IEEE), AND MUHAMMAD JAWAD KHAN^{1,3}

¹Department of Robotics and Intelligent Machine, School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

²Department of Mechanical Engineering, School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

³National Centre for Artificial Intelligence (NCAI), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding author: Muhammad Sajid (m.sajid@smme.nust.edu.pk)

ABSTRACT Machine learning is used for extraction of valuable information from data thus helping in exploration of hidden patterns, leading to learning models that can be used for prediction. In the domain of autonomous vehicles machine learning techniques have been applied in several areas, vehicle platooning being one of them. Vehicle platooning is a vital feature of automated highways which provides the key benefits of fuel economy, road safety and environmental protection coupled with safe road transportation. However, high computational cost associated with the numerical simulation of vehicle aerodynamics makes the Computational Fluid Dynamics (CFD) study of vehicle platoon prohibitively expensive and complex. Machine learning, with its high predictive power, has emerged as a promising complement to CFD studies of external aerodynamics. This paper presents estimation error based performance comparison of five different supervised learning algorithms: Support Vector Regression, Polynomial Regression, Linear Regression and two different models of Neural Networks for prediction of aerodynamic drag coefficient corresponding to each vehicle in a two, three and four vehicle platoon configurations based on the drag coefficients provided by experimental study at different inter-vehicle distances. Predicted drag coefficients are then juxtaposed with CFD data from numerical simulations to evaluate closeness to experimental drag coefficients. Results reveal that polynomial regression model best fits the aerodynamics with 0.0223 estimation error. To the best of our knowledge no machine learning based methods have been applied before for modeling aerodynamic drag on vehicle platoon.

INDEX TERMS Aerodynamics, computational fluid dynamics, machine learning, vehicle platooning.

I. INTRODUCTION

Extracting valuable information from raw data can be used for modeling physical relationship between system parameters. Need of exploring useful information from data can help in modeling very intricate relations between physical parameters. Machine learning techniques play vital role in accurately extracting information from data. Such techniques are replacing the traditional physical modeling methods by learning from the data and letting the algorithms itself learn the model.

Transportation plays a vital role in daily life. Human safety and fuel economy have always been the goals of development in the said field. Self-driving vehicles will constitute

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangxue Li^{id}.

the future transport systems providing the benefits of human comfort, fewer accidents and fuel and time economy. Substantial research has been going on in autonomous driving area in a multitude of dimensions including vision [1], control [2], tracking [3] and navigation [4].

The advent of automated highways, as an alternative to conventional highways, is a future vision of intelligent transportation that offers road safety [5] and smooth traffic flow [6] at high speeds [7], [8]. One of the most prominent features of the automated highways is autonomous vehicle platooning [9], [10] that makes an appreciable four-facet contribution [11], [12] in this modern transportation technology: fuel economy, environmental protection [13], road safety and smooth traffic flow. In fact, studies have been conducted to elicit maximum advantage from vehicle platooning in terms of these benefits through an optimum switch

control strategy [14] as well as distributed model predictive control [15]. The phenomenon of slip-stream effect allows the leader-follower configurations of the vehicles in minimizing the aerodynamic drag on each vehicle in the road train thereby resulting in reduced fuel consumption [16], [17] against subsequently decreased aerodynamic drag. As a matter of fact, 20% of all energy losses on modern vehicles are due to aerodynamic drag, flow separation being the primary reason [18]. The recent advances in internet of things and up-gradation in autonomous vehicles has provided tremendous opportunity in establishing autonomous vehicle platooning as a viable means of road transportation [19], [20]. Fuel economy is one of the aspects of road transport which is directly related to the aerodynamic drag faced by the vehicle [21]. Experimental methods for drag measurement deploy wind tunnel using scaled down models of cars to characterize aerodynamic coefficients [22]. On-road experiments were employed by Bruneau *et al.* [23] to derive a formula to predict drag coefficient using power law approach based on wake analysis of vehicle in platoon configuration. More recently, research trends have shifted from experimental to computational approach for estimating drag on vehicle platoons. Kaluva *et al.* [24] used numerical simulations to estimate drag on tractor trailer vehicles in platoons while Tadakuma *et al.* [25] studied autonomous electric vehicles, both studies highlight the computational cost of for larger more complex aerodynamic studies of vehicle platoons.

Owing to wind tunnel constraints in experimental work and computational expense in numerical simulations, predictive power of machine learning algorithms can be used to learn from experimental data leading to an aerodynamic model which predicts drag coefficients given inter-vehicular distance and vehicle tag number in the vehicle platoon. This paper assesses the performance of five machine learning models to predict drag coefficients in comparison with those numerically computed through Computational Fluid Dynamics (CFD). As discussed later in detail, the results reveal that the machine learning models perform better than CFD in terms of both time and computational expense. This study, thus, achieves a two-fold contribution. It, first, provides a way forward for coupling machine learning with vehicle aerodynamics as a viable conjunction to advance the development of vehicle platooning in the inter-disciplinary paradigm of automated highways. Secondly, in view of the associated benefits of relative cost and time saving pertaining to the data-driven approach of machine learning, this study serves to open up avenues for exploring machine learning as a promising partial alternative to wind tunnel testing as well as CFD simulation for the aerodynamic performance study of vehicle platoons.

This paper is organized into a total of five sections. Following the introduction in SECTION I, we discuss the literature review pertaining to the CFD study of the aerodynamics of vehicle platoon in SECTION II. we also highlight the background and the related works in the domain of artificial intelligence for employing machine learning algorithms to the drag coefficient data obtained from the CFD study of

vehicle platoon aerodynamics, for data prediction. Accordingly, SECTION III elucidates the methodology of this study: outline of the machine learning regression evaluation parameters, machine learning training framework, data processing for training, explication of the CFD study for obtaining drag coefficient data, training and testing procedures along with loss functions of the machine learning algorithms, and,, a discussion of the mathematical formulations of these machine learning algorithms. In the SECTION IV, we evaluate the machine learning predictions compared to the CFD data and experimental data while assessing the performance of the machine learning algorithms in terms of prediction estimation error. In addition, we also discuss the results of the CFD study of the vehicle platoon aerodynamics. Finally, in SECTION V, we summarize our findings from this study and present future recommendations.

II. RELATED WORK

A. AERODYNAMIC STUDY OF VEHICLE PLATOON

Aerodynamic study of the vehicle platoon through CFD [26] is usually employed to study the drag characteristics of each vehicle in the platoon since experimental testing abounds considerable risk and cost. In particular, the experimental study of drag on platoons using wind tunnel testing is not feasible for platoons greater than 4 cars given the space restriction within a wind tunnel as opposed to greater length platoons for scaled-down vehicle size, let alone the original size of the vehicles. On the other hand, although there has been appreciable development in high performance computing but the CFD numerical solution are exorbitantly expensive for a multitude of practical engineering applications especially capturing flow physics in large computational domains at highly turbulent flows. The driving cause for the high computational cost of CFD is the requirement of a fine grid for a reliable simulation of the fluid flow phenomenon through a grid-independent solution. Hence, this fine grid requirement reduces the computational time step thereby translating into uneconomical usage of computational resources [27]. The associated cost of doing high fidelity CFD study of vehicle platoons that captures the very fine aerodynamics over minute details of vehicles, like flow around side-mirrors, front grill, tires and other complex geometric features, is prohibitively expensive. To quantify the budget requirement for such a study, we estimated the cost of replicating the experimental work of Calif. Pathway [28] on commercially available CFD resources like Ansys Fluent [29] via cloud computing services offered at Amazon Web Services [30] to be approximately 2 million USD. This estimate is bound to increase exponentially with the increase in the platoon size coupled with a fine grid of around 200 million elements to adequately capture the flow physics for each vehicle in the platoon.

This experimentally and computationally demanding outlook of the investigation of vehicle aerodynamics prevents an extensive study of the drag characteristics of a large vehicle

platoon covering its diverse parameters, including vehicle permutations, speeds and inter-vehicle distances. Hence, it is not economically feasible, and therefore research conducive, to conduct further experimental and computational studies on platoon aerodynamics to collect more data for data-driven predictions.

In the previous decade, plenty of research has been done to explore the potential of employing Artificial Intelligence in supplementing CFD [31]–[33]. One application of Artificial Intelligence in augmenting the Computational Fluid Dynamics is the use of a data-driven approach of aerodynamic performance evaluation where machine learning algorithms are used for the prediction of the drag coefficients for a wide range of real-world geometries and/or scenarios. He, *et al.* [34] conducted an extensive inquiry into the modeling of the drag forces that are vital to the dynamics of dense fluid-particle systems. The results of the study establish the reliability of the supervised machine learning approach especially artificial neural network (ANN) for the estimation of the drag force. Given the high computational cost of the CFD coupled with wind tunnel testing for estimating the drag coefficient of the car side silhouette design, Gunipar *et al.* [35] also resorted to the utilization of machine learning regression and neural-network methods for obtaining a mathematical model which is trained on the available drag coefficient dataset obtained from CFD simulation. This trained mathematical model, in turn, reliably predicts the coefficient of drag of a given silhouette. Similarly, Dube and Hiravennavar [36] employed data driven drag prediction for studying the aerodynamic performance of the underhood and underbody drag enablers by using linear regression, neural network, and random forest approaches to generate models for a fairly accurate prediction of the associated aerodynamic drag coefficients.

Based on these studies an alternative approach to experimental or computational studies is to employ Machine-Learning (ML) algorithms to use its high predictive power for estimating the aerodynamic drag characteristics of each vehicle in the platoon in comparison with the CFD approach that computes the numerical solution (aerodynamic drag coefficients) by solving complex Navier Stokes equations. The drag coefficients of each vehicle platoon obtained from the experimental wind-tunnel tests at various inter-vehicle distances at a speed of 23 m/s serves as the input database with known drag coefficient for machine learning models. For most aerodynamic objects the drag coefficient remains nearly constant across a broad range of speeds, we therefore focus on 23 m/s speed which is comparable to urban speed limits across the world. Given the greater engine power required to overcome air drag than tire and mechanical resistance at a speed of approximately 23 m/s [37], it is imperative to minimize the energy loss in aerodynamic resistance and, therefore, reduce fuel consumption. In this regard, it was observed that at spacings lesser than 0.25 times the car length, corresponding decrease in the aerodynamic drag coefficient is diminutive [28], [38]. Thus, in this study the minimum inter-vehicle distance was chosen to be 0.25 times

the car length. In order to address the risk of collision at such small inter-vehicle distances, Murthy and Masrur [39] present an experimentally validated approach to optimize the braking system of autonomous vehicle platoon considering the vehicle loading condition for executing a collision free braking maneuver. Moreover, the close/narrow inter-vehicle distance, used in vehicle platoon studies [28], [38] and in this study, have been substantiated by the recent advancements in autonomous vehicle ecological cooperative control [40], [41] coupled with IoT [42], [43] to ensure passenger safety. Models deployed for drag coefficient prediction include Artificial Neural Networks, Regression models and Support Vector Regression. Once the training of these models is completed, they are applied to unseen vehicle platoon data for predicting drag coefficient. This estimated data and the one obtained from CFD simulations [44] is then compared with corresponding drag coefficient data acquired from experimentation. The analogy of this general approach of reproducing the complex computational model in terms of high predictive capability, has been used to achieve remarkable curtailment in high computational demands relative to original model [45], [46].

Keeping this in perspective, this study aims at investigating the feasibility of complementing Computational Fluid Dynamics (CFD) solutions with Machine Learning in the study of aerodynamics of vehicle platoons. In particular, the aerodynamic drag coefficient of each vehicle in a two, three and four vehicle-platoon for eleven different inter-vehicle distances, ranging from 1L to 2L, will be computed through the combination of the robust yet costly numerical approach of CFD and then will be predicted by the high predictive capabilities coupled with cost-effectiveness offered by machine learning algorithms via training on experimental data from California PATH project [28].

B. MACHINE LEARNING

The study and construction of computer algorithms that can learn from the input data is the domain of machine learning. Machine learning is a branch of artificial intelligence that enables computers to create knowledge and draw inferences from data, it can learn hidden patterns inside data and represent it in the form of a model. Machine learning algorithms are being used in various applications in science and technology these days, since they are capable of detecting meaningful patterns in the data provided and apply them to new data [47]. This approach is an alternative to conventional approach for devising algorithmic solution [48]. While conventional design flow begins from data acquisition followed by mathematical modeling from fundamental physics principles, machine learning approach replaces domain knowledge with data acquisition, thereby using that data for training by using a learning algorithm to produce a trained machine. Two approaches to machine learning are supervised learning and unsupervised learning.

In supervised learning, the training dataset has pairs of inputs and ground truth labels and the algorithm learns a

TABLE 1. Prevalent evaluation metrics for regression.

Evaluation Metrics	Mathematical Formula
MAE (Mean Absolute Error)	$1/m \sum_{i=0}^m \hat{y}_i - y $
MSE (Mean Squared Error)	$1/m \sum_{i=0}^m (\hat{y}_i - y)^2$
RMSE (Root Mean Squared Error)	$1/m \sqrt{\sum_{i=0}^m (\hat{y}_i - y)^2}$

mapping from input to the output [48]. Applications of supervised learning include regression and classification. In unsupervised learning inputs have no labels. Idea of unsupervised learning is to discover the inherent properties of mechanism generating data. In this study, the problem statement of aerodynamic drag coefficient prediction falls under the category of supervised machine learning through regression. Studies pertaining to regression problems have highlighted the difference in performance of various machine learning algorithms especially epidemic regression models and artificial neural networks [49], [50]. These studies employed a selection criterion based on the performance of the machine algorithms in terms of estimation error and computation time. However, an extensive literature review reveals that there are no substantial research endeavors that aim at the application of these five machine learning algorithms to aerodynamics of vehicles and, especially, the vehicle platoons. Therefore, it can be reasoned that for a small data set, as in this study, the regression models may perform better than neural network-based approaches. Hence, we present a comparison of neural network and regression models for the prediction of drag force on vehicles in a platoon configuration. Accordingly, the procedure for conducting machine learning training and testing was carried out using common machine learning modeling and training best practices. Five supervised machine learning algorithms have been employed in this study to provide an efficient and effective alternative to both wind tunnel tests as well as CFD studies for predicting the aerodynamic performance of vehicle platoons thereby helping to advance the development of vehicle platoon through the supplementation of the domain of vehicle aerodynamics by machine learning in the interdisciplinary context of automated highways.

III. METHODOLOGY

Proposed methodology is to compare five regression models: Support Vector Regression (SVR), Linear Regression (LR), Polynomial Regression (PR) and two neural network models namely ANN-I and ANN-II to predict the drag coefficient considering the experimental data as the true values.

A. EVALUATION PARAMETER

Popular metrics used in regression problems are shown in Table 1:

Here, m is total number of examples, \hat{y} is prediction and y is true label. It is important to note that MSE is used as the evaluation metric because it punishes the larger values.

B. TRAINING PROCESS

Fig. 1 shows the training procedure. Red boxes indicate experimentally calculated drag coefficients [28] and CFD calculated drag coefficients. Fig. 2 shows the vehicle tags and configuration within the platoon of 2; 3 and 4. This is the same configuration as was employed in the pathway experiment [28] and in the simulations. Brief overview of the training procedure is presented below while details follow the current section:

- Taking the experimental data as the true labels, data set has been prepared with vehicle tag and inter-vehicular distance being two features, since these two parameters along with speed (which is constant in this case) influence drag coefficient.
- Same feature parameters have been used to calculate drag using numerical simulations (CFD).
- Data set has been used to train five models namely: Linear Regression, Polynomial Regression, Support Vector Regression and two models of Neural Networks with MSE loss and Huber Loss.
- After the training is complete, test data is used for prediction of drag coefficients.
- Predicted values are compared with experimental values and error has been calculated as shown in Table 1.
- Model showing minimum estimation error is taken as the finalized model.

C. DATASET

The dataset comprises two columns of features namely: Inter-vehicular distance and Vehicle tag. The vehicle tags numbers have been chosen as separate features because their position in platoon influences the drag faced by the vehicle. Front-most vehicle (SUV-1) will experience the greatest drag (and hence the greatest drag coefficient) which will subsequently be lessened for later vehicles, hence, making vehicle tag a salient feature. Furthermore, as the vehicles move in a platoon, they maintain a specific spacing, so this inter-vehicular distance can greatly affect the drag. Thus, it is logical to consider these two attributes and make a dataset of which Inter-vehicular distance and Vehicle tag are two columns. Vehicle tags corresponding to two, three and four vehicle platoons are SUV1, SUV2; SUV1, SUV2, SUV3; and SUV1, SUV2, SUV3 SUV4, as shown in Fig. 2. The data-points give values of drag at following inter-vehicular distances: 0.06m, 0.1m, 0.13m, 0.15m, 0.19m, 0.25m, 0.3m, 0.33m, 0.4m, 0.5m, 0.63m, 0.75m, 1m.

Fig. 3 shows dataset structure. It is important to note here that the experimental data is extracted from the results of pathway [28] experiment. There are 28, 42 and 56 data-points for platoon comprising 2, 3 and 4 vehicles.

D. COMPUTATIONAL ESTIMATION OF DRAG

The computational domain employed for the CFD study of 4-vehicle platoon is formulated on the basis of the Ahmed Body CFD results [51]. The layout of the configu-

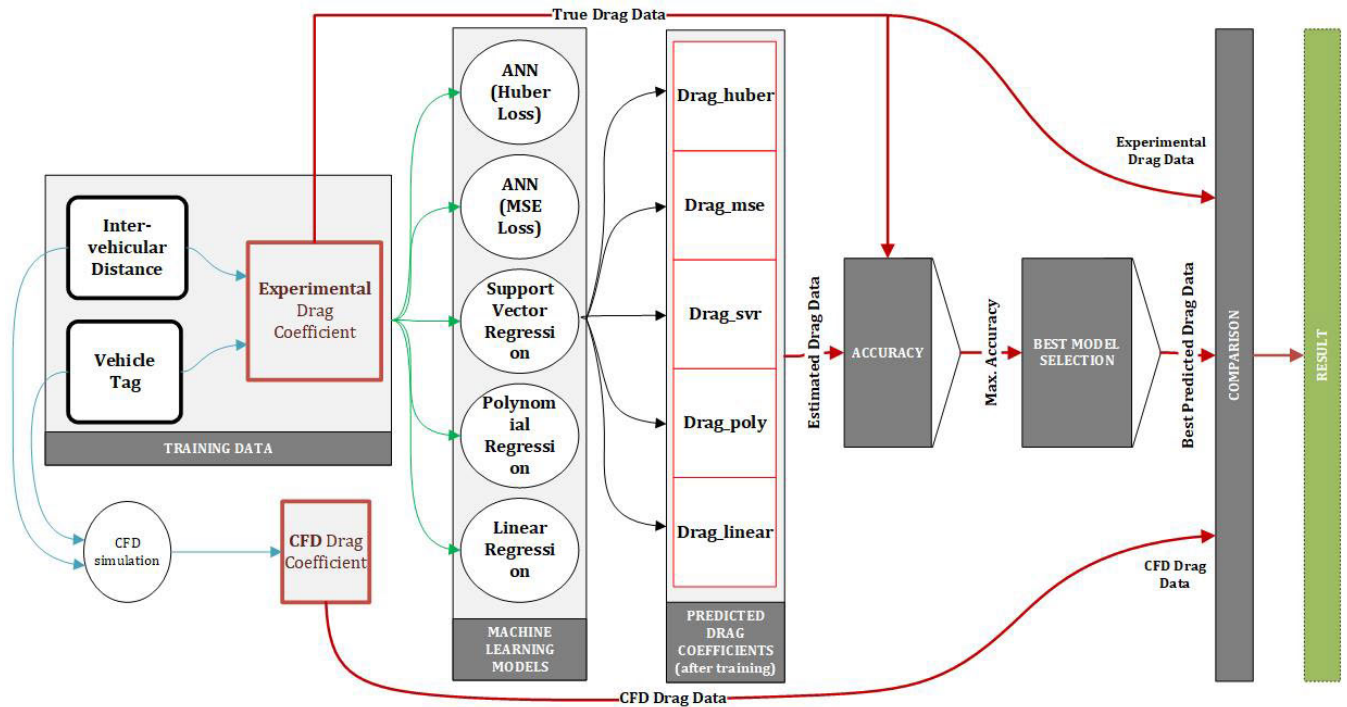


FIGURE 1. Process flow chart of the machine learning implementation.

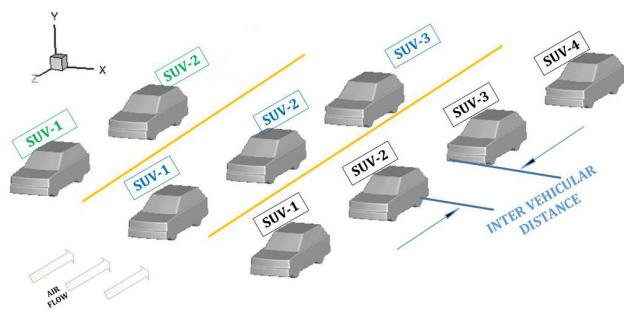


FIGURE 2. Vehicle tags corresponding to two, three and four vehicles in a platoon.

ration of 4-vehicle platoon for CFD simulation is illustrated in Fig. 2. The meshing of the computational domain for each of the inter-vehicle spacing ranging from 0.25L to 1L is carried out on consistent parameters. The values of these mesh parameters are dictated by the best practices in the computational study of external aerodynamics [52] such that the resulting mesh density ensures a numerically realizable yet computationally cost-effective CFD solution. In this regard, the specifications of the vehicle under consideration is based on the vehicles employed in Calif. Pathway [28] experiment: 1991 General Motors Lumina APV having a length of 4616 mm, width of 1890 mm, height of 1688 mm, and a ground clearance of 184 mm. On the other hand, Fig. 4 depicts the detail of mesh elements employed in the discretization of the computational domain. In particular, the computational domain of 4-vehicle platoon is discretized using, on average,

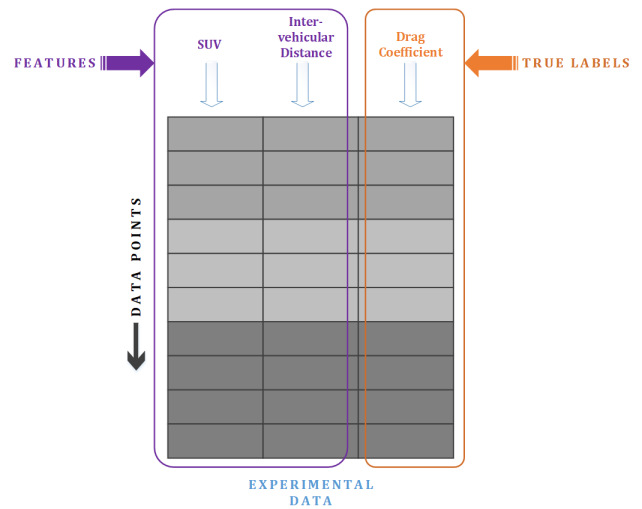


FIGURE 3. Description of Dataset.

25 million volume mesh elements with an appreciable orthogonal quality and skewness. Table 2 summarizes the average mesh statistics for the CFD study of 4-SUV platoon at inter-vehicle distances of 0.25L, 0.5L, 0.75L and 1L. The remaining aspects of the computational setup including governing equations, solver parameter and convergence are based on the work by Farid *et al.* [44].

E. TRAINING AND TESTING DATA

As evident from section III-C, there are a total of 126 data points: 10% of the data is separated as test data while the rest is used for training.

TABLE 2. Average Mesh Statistics of 4-vehicle platoon with inter-vehicle distances of 0.25L, 0.5L, 0.75L and 1L.

No. of Elements	No of Nodes	Average Aspect Ratio	Average Skewness	Average Orthogonality
27425053	14573396	3.1297	0.218405	0.779655

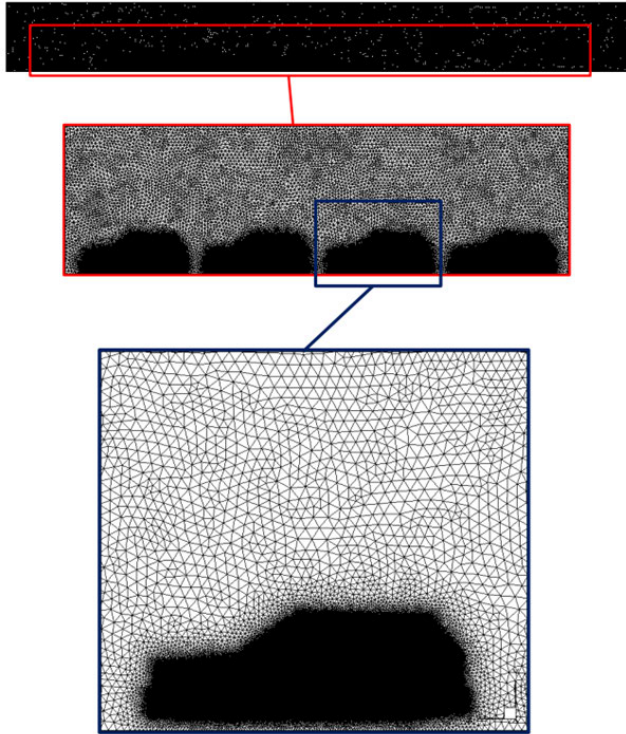


FIGURE 4. Mesh detail of 4-vehicle platoon for CFD study.

TABLE 3. Pre-processing methods for regression.

Operation	Mathematical Formulation
Standardization	$\frac{X_i - \bar{X}}{\sigma}$
Normalization	$\frac{X_i - X_{min}}{X_{max} - X_{min}}$

Pre-processing is a necessary operation which needs to be performed on data before training. Pre-processing ensures that the features are on a similar scale, enabling gradient descent to reach global minimum quickly. Table 3 shows two pre-processing methods.

Here X_i is the i th entry of column, \bar{X} is the mean of column, σ is standard deviation of column, X_{max} is maximum value in column and X_{min} is minimum value in the column.

Standardization brings the feature distribution mean at 0 and Normalization pushes the feature columns within [0, 1] range. Table 4 shows the type of pre-processing operation applied prior to training for each of the models.

F. LOSS FUNCTIONS

Loss function/Objective function characterizes how well a model performs. Depending on the objective, it can be

TABLE 4. Pre-processing method used for the models.

Pre-processing type	Model
Linear Regression	Normalization
Polynomial Regression	Normalization
Support Vector Regression	Standardization
ANN-I	Normalization
ANN-II	Normalization

TABLE 5. Loss functions used for regression.

Loss function	Mathematical Formulation
MSE ^a	$1/m \sum_{i=0}^m (\hat{y}_i - y)^2$
MAE ^b	$1/m \sum_{i=0}^m \hat{y}_i - y $
Huber	$\begin{cases} 1/2 (y - \hat{y})^2 & \text{for } y - \hat{y} \leq \delta \\ \delta y - \hat{y} - (1/2)\delta^2 & \text{otherwise} \end{cases}$

^aMean Squared Error
^bMean Absolute Error

minimized or maximized. Table 5 shows primary loss functions used for regression task.

1) MEAN SQUARED ERROR (MSE)

MSE is the most widely used loss function used for regression. It is the averaged sum of squared distances between true label value and predicted value. Key attributes of MSE include being easier to solve and continuous derivative at all points. It is also known as L2 loss.

2) MEAN ABSOLUTE ERROR (MAE)

MAE also called as L1 loss is average sum of absolute differences between our target values and predicted values. MAE derivative is continuous everywhere except at 0.

Mathematically it is evident that L1 loss is robust against outliers in data as compared to L2 loss. This is because L2 loss squares the error, which in case of outliers becomes a large number.

Apart from this, L1 loss has the same gradient value which also means it will have large gradient for small inputs. Contrary to this, MSE gives small gradient for small input values and greater for large values. There is a possibility that gradient descent might take large step and skip global minimum. These problems can be mitigated to large extent by using a loss function with the qualities of both MSE and MAE; Huber loss is one such function.

3) HUBER LOSS

Huber loss is an amalgam of MSE and MAE and so is less sensitive to outliers and also differentiable at 0. For some small value δ it behaves like L2 loss if the error is smaller than δ and becomes L1 if the error is larger than δ . It is also evident from equation given in Table 5, Huber loss approaches MAE if δ is small and becomes nearly same as MSE if δ is large. Comparison of Huber loss with different values of δ and with other loss functions is shown in Fig. 5. δ is also a hyper-parameter and is learnt the same way as other weights of the model.

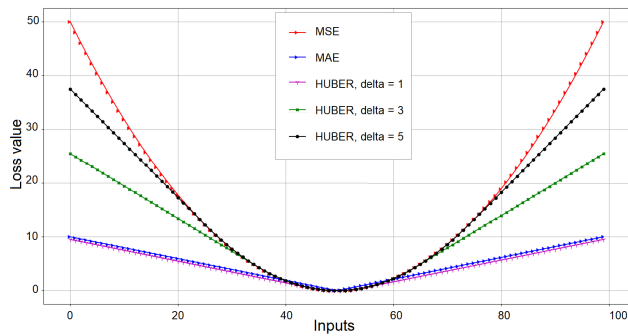


FIGURE 5. Comparison of Loss functions.

G. MACHINE LEARNING ALGORITHMS

In this research, five supervised learning algorithms have been applied. Deep learning framework employed is Keras. Same training data is used for training all the models. After training is complete, models are compared to find the best one which predicts the drag coefficient close to the original experimental data. Models used and their architecture has been briefly described below:

1) LINEAR REGRESSION

Mathematically linear regression is defined as in Equation (1):

$$h_{\theta}(x) = \hat{Y} = \theta_0 + \sum_{i=1}^N \theta_i x_i \quad (1)$$

Here, θ_0 is the bias, θ_i is the i th weight and N is the total number of features in the dataset.

As is evident from linear model equation, it tries to fit a straight line through data while minimizing the MSE loss (defined in Table 1) and the weights learnt by model behave similarly as the slope and y-intercept in equation of a straight line.

Linear regression model works well if dependent and independent variables correlate with each other to some extent [53]. In this dataset the drag faced by the front vehicle is greater than those following front vehicle, leaving the possibility of existence of a linear relationship with drag.

2) POLYNOMIAL REGRESSION

Polynomial regression is defined by Equation (2) where M is the order of polynomial, x^j denotes feature x to the power of j and w denotes coefficients vector $w_0, w_1, w_2 \dots w_M$.

$$y(x, w) = w_0 + w_1 x^1 + \dots + w_M x^M = \sum_{j=0}^M w_j x^j \quad (2)$$

It comes under the domain of linear models because although the function $y(x, w)$ is nonlinear with respect to x but is linear in terms of unknown variable w hence fitting the data in $R^{(M+1)}$ space. Values of the coefficients are determined by fitting this polynomial to the training data by using MSE loss function as indicated in Table 5. Polynomial regression

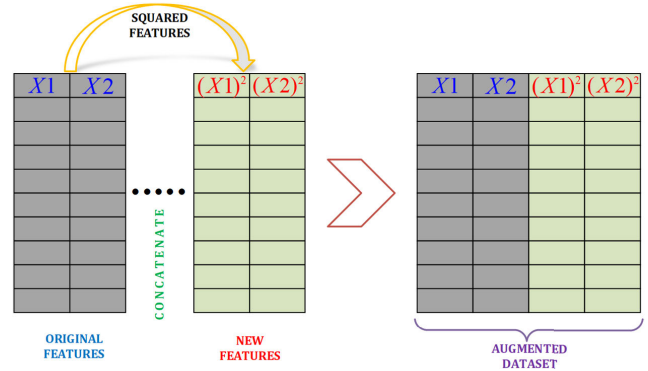


FIGURE 6. New features set for polynomial regression.

model in this research uses a second degree polynomial and Fig. 6 shows new features made in the dataset. One of the problems faced by polynomial regression is that as the degree of polynomial grows the magnitude of the learnt coefficients typically gets larger and so the model becomes tuned to the random noise on target values [54]. Less magnitude of the weights also means that noise in the input will not affect the model performance much whereas if the magnitude of learnt coefficients is large, any noisy data will render the model prone to wrong predictions. This problem can be avoided using regularization. Polynomials are flexible and useful where a model must be developed empirically and can fit a wide range of curvatures giving a good approximation of the relationship.

3) SUPPORT VECTOR REGRESSION

Support Vector algorithm is a non-linear generalization of generalized portrait algorithm, firmly grounded in statistical learning theory framework. Such algorithms are designed to fit well on the unseen data. In support vector based algorithms, data is transformed into higher dimensional space(kernel trick) and is then classified or regressed based on the type of algorithm.

Support vector regression tries to fit the best line within a predefined or threshold error value instead of minimizing the error between the predicted and the actual value [55]. Also, Support Vector Regression is independent of the dimensionality of input data [56].

Input data is transformed into high dimension by applying kernel, followed by formulation of correlation matrix using which weights are learnt. These weights are used to estimate the test data. Regression takes place within high dimensional vector space. The linear regression within the (transformed) vector space is somewhat different than least squares method [46].

Fig. 7 demonstrates the support vector machine parameters. Only the points outside the boundary region contribute to the cost insofar, as the deviations are penalized in a linear fashion. The error function for support vector regression can be stated as Equation (3) while Equation (4), Equation (5), Equation (6) and Equation (7) demonstrate the constraints of

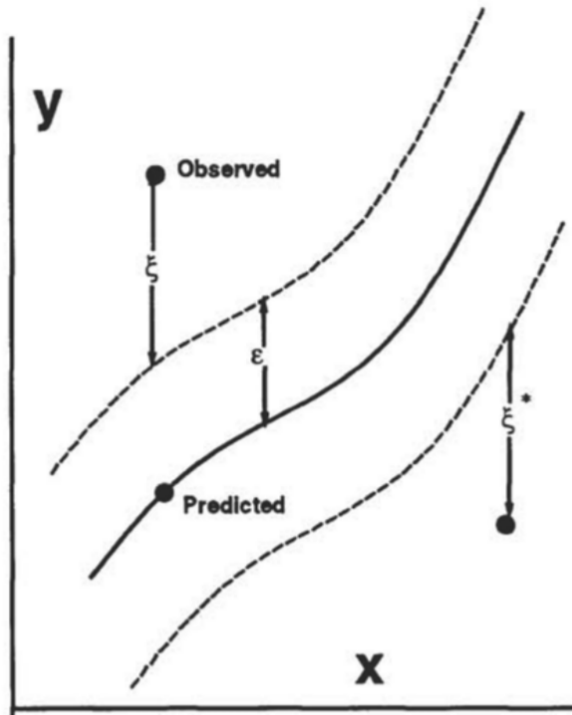


FIGURE 7. Parameters of SVR, (taken from [56]), ϵ = margin, ξ is the distance from nearest points.

optimization problem.

$$E = C \sum_{n=1}^N (\epsilon_n + \hat{\epsilon}_n) + \frac{1}{2} \|w^2\| \quad (3)$$

$$\epsilon_n \geq 0 \quad (4)$$

$$\hat{\epsilon}_n \geq 0 \quad (5)$$

$$y(X_n) + \epsilon + \epsilon_n \geq t_n \quad (6)$$

$$y(X_n) - \epsilon - \epsilon_n \leq t_n \quad (7)$$

In these equations, t_n is the target and $\epsilon_n, \hat{\epsilon}_n$ are the slack variables. $\epsilon_n > 0$ refers to point where $t_n > y(X_n) + \epsilon$ and $\hat{\epsilon}_n > 0$ refers to a point where $t_n < y(X_n) - \epsilon$.

4) NEURAL NETWORKS

Artificial Neural Networks are an imitation of the human brain. The main idea is to create a network of simple processing units called neurons, which perform computations. A transfer function is applied on the weighted sum of the inputs to each neuron and the result is forwarded as the output value of that particular neuron [57]. They are also called as universal function approximators because given enough hidden layers and neurons, they are able to approximate any function. Neural networks are used in supervised and unsupervised machine learning algorithms.

Fig. 8 shows generic model of a neural network. The index of the neuron is i , it receives inputs from N other neurons. The strength of the connection from neuron j to neuron i is denoted by w_{ij} . The function $\theta_H(b)$ is the activation function. The threshold value for neuron i is denoted by μ_i .

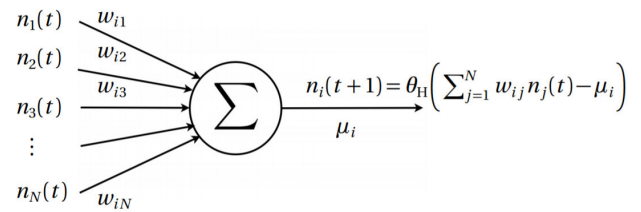


FIGURE 8. Neural Network Model.

TABLE 6. Preprocessing method used for models.

Activation Function	Mathematical Formulation
Sigmoid	$1/(1 + exp^{-z})$
Relu	$max(0, z)$
Softmax	$\frac{exp^{z_i}}{\sum_j exp^{z_j}}$

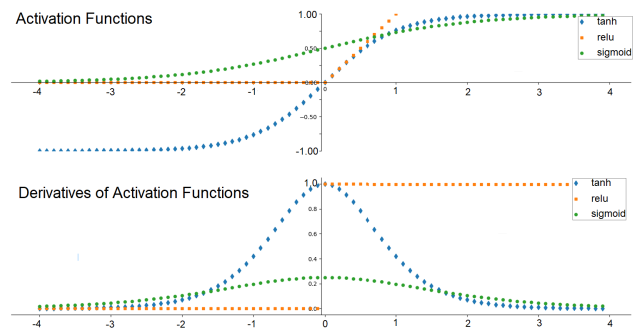


FIGURE 9. Activation Functions Comparisons.

The index $t = 0, 1, 2, 3$ labels the discrete time sequence of computation steps.

Neural Network applies subsequent transformations on the input data as it passes through hidden layers, including linear (affine transformation) (Equation (8)) followed by a nonlinear transformation (activation function denoted by σ (Equation (9)) for each hidden layer.

$$Z = W \cdot X + B \quad (8)$$

$$Activation_{output} = \sigma(Z) \quad (9)$$

For problems related to regression, no activation function is applied in the output layer to output raw score value. Table 6 shows mathematical formulation of different activation functions. \tanh function is also used as activation but is more prevalent in recurrent neural networks. Sigmoid activation function squashes the output between $[0, 1]$ range. Relu activation makes input values less than z to zero.

Nonlinear transformations (activation functions) have their own inherent properties of derivatives as shown in Fig. 9, suitable for the problem type [58], [59]. This allows neural networks to fit on any type of data.

Fig. 9 shows a comparison of activation function plots along with their derivatives. Derivatives are important in the back-propagation. ReLU activation function solves the problem of vanishing gradient faced by sigmoid and tanh

TABLE 7. Comparison of ANN-I and ANN-II.

Model Attribute	ANN-I	ANN-II
Input neurons	2	2
Hidden Layers	3	2
Loss function used	MSE	Huber
Output neurons	1	1

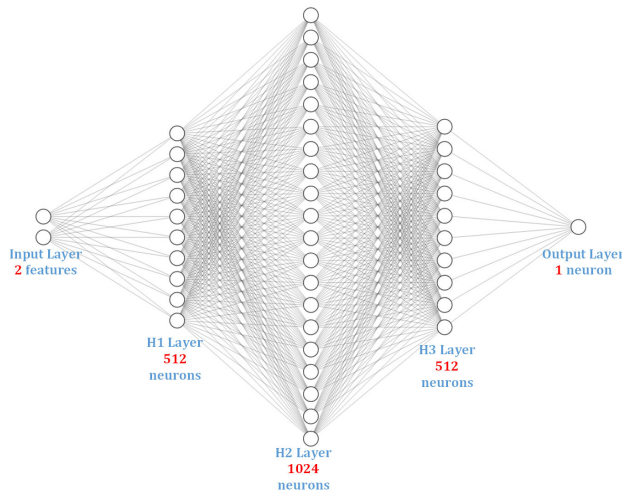


FIGURE 10. ANN-I model.

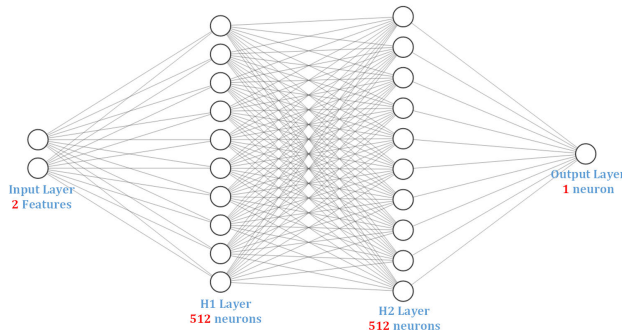


FIGURE 11. ANN-II model.

activation functions. It is clear that at smaller or larger values of input, *sigmoid* and *tan h* derivatives tend to be closer to zero ($Derivative_{\sigma(x)} \approx 0 | \sigma = (sigmoid, tan h) \wedge (x \gg 0 \vee x \ll 0)$), this declines learning of the corresponding weights during back-propagation. As compared to this ReLU derivative is a large constant value for $x > 0$ where x is input. This constant value of ReLUs results in faster learning [60]. An added benefit is that $ReLU(x) = 0 | x \leq 0$ which results in sparsity and sparse representations are beneficial as compared to dense representations because forward and backward propagation consist of a series of matrix operations.

Two different models of neural networks namely ANN-I and ANN-II are employed in this research. Table 7 shows a comparison of both models whereas architectures of both models are shown in Fig. 10 and Fig. 11.

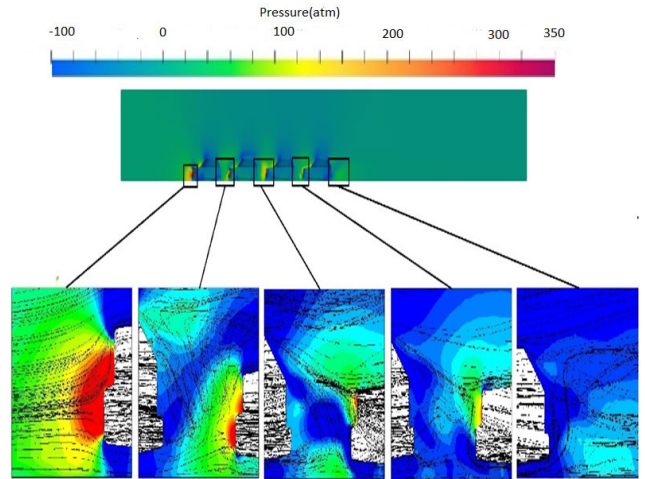


FIGURE 12. CFD visualization of the aerodynamic characteristics of 4-vehicle platoon using pressure contours and air-velocity streamlines.

IV. RESULTS AND DISCUSSION

The CFD results obtained after the simulation of computational domain in OpenFOAM® are illustrated in Fig. 12 through colored contour plots and air flow streamlines that help visualize the pressure distribution and air velocity, respectively. The flow physics in the regions of interest near the front and back bumper of the vehicles has been specifically highlighted to study the general aerodynamic behavior of 4-vehicle platoon. The head-on interaction of the first vehicle with the incoming air results in the development of stagnation region at the front bumper of the vehicle, that encapsulates a high pressure thereby resulting in maximum aerodynamic drag coefficient of the lead vehicle in the platoon. This is followed by the development of boundary layer owing to the air flow over vehicle body and road surfaces, which is characterized by the velocity gradient. The interaction of the air with the vehicle bodies at the gaps between the first, second and third vehicles in the platoon results in the formation of vortices as indicated by the recirculation of the air flow. These vortices are of almost similar density, so this translates into approximately similar drag coefficients for the second and third vehicles. Although, the vortex formation at the front bumper of last vehicle is marked by a relatively less concentration but the development of trailing vortices in the wake of the last vehicle reciprocates into a high drag coefficient for the last vehicle in the platoon.

Table 8 shows the estimation error results. It seems that polynomial regression is the best model for this data, giving 0.0223 estimation error. Apart from estimation error, the generalization capability of the model is a major concern when dealing with sparse data sets. The machine learning predictions based on small dataset, like the one employed in this work, are expected to suffer from over-fitting: increase in error computed for validation data occurring in concurrence to the decrease in error computed for training data. To address the issue of over-fitting, various techniques have been sug-

TABLE 8. Estimation error comparison.

	Linear Regression	Polynomial Regression	Support Vector Regression	ANN-I	ANN-II
Estimation Error	0.2427	0.0223	0.085	0.0491	0.0498
Parameters	3	5	rbf-kernel	1,052,161	264,705
Update Method	Gradient Descent	Gradient Descent	Gradient Descent	Gradient Descent	Gradient Descent

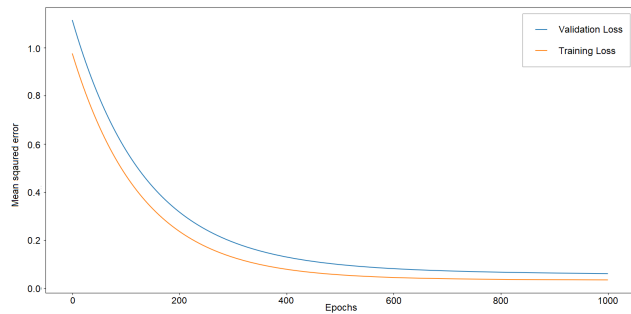


FIGURE 13. Polynomial Regression training and validation loss.

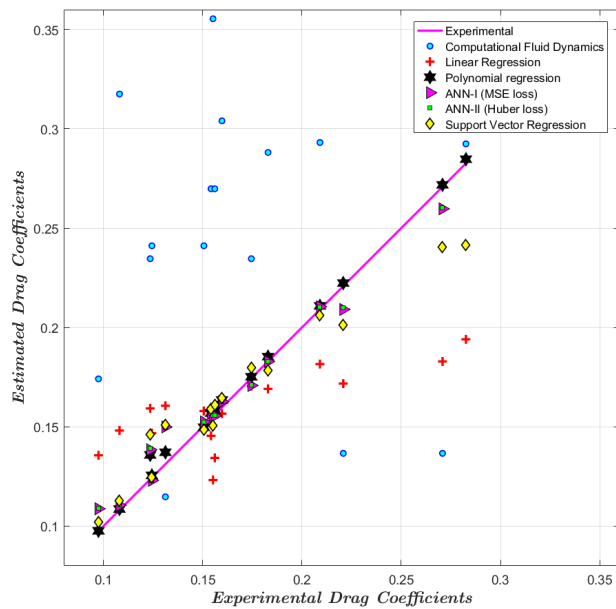


FIGURE 14. Comparison of models.

gested in the literature to improve ANN generalization capabilities. In this regard, cross-validation [61] is a widely used and accepted approach for scarce dataset. To address this concern, we employed cross-validation approach which compares training and validation errors at each iteration of the model to provide an optimal criteria. The mean square error from training and validation data are shown to converge towards their respective minimum values in Fig. 13, indicating that the model does not suffer from over-fitting and is general in nature.

Fig. 14 shows predicted drag coefficient from all five models compared with CFD data and experimentally calculated

drag coefficients. As is evident from the figure, linear model performs the worst as compared with other prediction algorithms. The least accurate model among prediction algorithms (linear regression) is able to provide better approximations to experimentally calculated drag coefficient values as compared to numerical simulations. While in terms of processing time, none of the models take as much time for training as CFD takes to arrive at a solution. ANN-I and ANN-II both show approximately same results whereas SVR performs better than ANN but worse than linear model. Overall, the polynomial regression has the relatively lowest estimation error, as illustrated in Fig. 14, while the performance of linear regression is worst in terms of prediction estimation error. This poor performance of linear regression is a direct consequence of under-fitting which results in a high RMSE value, and, therefore, high prediction estimation error. Accordingly, to reduce the prediction estimation error, the complexity of the model has to be increased by increasing the degree of the polynomial fitting on the data. As a result, the second order polynomial regression model of relatively higher complexity, used in this study, produced the best performance in data prediction by virtue of the subsequent increase in the magnitude of the learnt coefficients as well as the increased capability of the higher-degree i.e. second-order polynomial to become tuned to the random noise on the target values. Moreover, since the size of the data set is smaller for the proper training of a deep neural network as well as support vector regression, so this also explains the better performance of polynomial regression compared to other machine learning algorithms.

V. CONCLUSION

In this study, we have compared different machine learning approaches for the estimation of the aerodynamic drag coefficients for 2 to 4 vehicle platoon configurations. Inter-vehicle distances and vehicle tags were used as features for training and testing of the data. The data was regressed using support vector regression, polynomial regression, linear regression and neural networks for (comparison and) prediction of coefficients. Different models for neural networks consisting of two and three layers were used. Mean square error was used as loss function for the two layered network whereas, Huber model was used for the three layered network to compute the loss. Gradient decent algorithm was used as back propagation criterion for all machine learning models. Polynomial regression computed the lowest estimation error (0.0223) whereas the neural network models, ANN-I and ANN-II, computed 0.0491 and 0.0498 estimation errors, respectively. The neural network estimation errors were higher than the estimation error of polynomial regression due to limited training data. Upon further availability of data, a neural network may perform better. Therefore, it is proposed that in comparison to a numerical simulation, a neural network model can approximate drag coefficients for multi vehicular platoon configurations effectively with comparable performance and less coefficient estimation time.

ACKNOWLEDGMENT

The authors would like to thank SMME CAD-CAM lab and the help and support of Ahmed Husnain Johar and Talha Yousuf for results compilation and the valuable and constructive suggestions of Adeel and Tahir during the planning and development of this research work.

REFERENCES

- [1] G. A. P. de Moraes, L. B. Marcos, J. N. A. D. Bueno, N. F. de Resende, M. H. Terra, and V. Grassi, "Vision-based robust control framework based on deep reinforcement learning applied to autonomous ground vehicles," *Control Eng. Pract.*, vol. 104, Nov. 2020, Art. no. 104630, doi: [10.1016/j.conengprac.2020.104630](https://doi.org/10.1016/j.conengprac.2020.104630).
- [2] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, "A survey of deep learning applications to autonomous vehicle control," *IEEE Trans. Intell. Transp. Syst.*, early access, Jan. 7, 2020, doi: [10.1109/TITS.2019.2962338](https://doi.org/10.1109/TITS.2019.2962338).
- [3] M. Park and Y. Kang, "Experimental verification of a drift controller for autonomous vehicle tracking: A circular trajectory using LQR method," *Int. J. Control, Autom. Syst.*, vol. 18, pp. 1–13, Sep. 2020, doi: [10.1007/s12555-019-0757-2](https://doi.org/10.1007/s12555-019-0757-2).
- [4] H. A. Ibrahim, A. T. Azar, Z. F. Ibrahim, and H. H. Ammar, "A hybrid deep learning based autonomous vehicle navigation and obstacles avoidance," in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, A.-E. Hassanien, A. T. Azar, T. Gaber, D. Oliva, and F. M. Tolba, Eds. Cham, Switzerland: Springer, 2020, pp. 296–307.
- [5] J. Carbaugh, D. N. Godbole, and R. Sengupta, "Safety and capacity analysis of automated and manual highway systems," *Transp. Res. C, Emerg. Technol.*, vol. 6, nos. 1–2, pp. 69–99, 1998, doi: [10.1016/S0968-090X\(98\)00009-6](https://doi.org/10.1016/S0968-090X(98)00009-6).
- [6] P. Ioannou, Ed., *Automated Highway Systems*. New York, NY, USA: Springer, 1997, doi: [10.1007/978-1-4757-4573-3](https://doi.org/10.1007/978-1-4757-4573-3).
- [7] R. E. Fenton and R. J. Mayhan, "Automated highway studies at the Ohio state University—An overview," *IEEE Trans. Veh. Technol.*, vol. 40, no. 1, pp. 100–113, Feb. 1991, doi: [10.1109/25.69978](https://doi.org/10.1109/25.69978).
- [8] R. E. Fenton, "IVHS/AHS: Driving into the future," *IEEE Control Syst. Mag.*, vol. 14, no. 6, pp. 13–20, Dec. 1994, doi: [10.1109/37.334411](https://doi.org/10.1109/37.334411).
- [9] S. Maiti, S. Winter, and L. Kulik, "A conceptualization of vehicle platoons and platoon operations," *Transp. Res. Part C, Emerg. Technol.*, vol. 80, pp. 1–19, Jul. 2017, doi: [10.1016/j.trc.2017.04.005](https://doi.org/10.1016/j.trc.2017.04.005).
- [10] M. Muratori, J. Holden, M. Lammert, A. Duran, S. Young, and J. Gonder, "Potentials for platooning in US highway freight transport," Nat. Renew. Energy Lab (NREL), Golden, CO, USA, Tech. Rep. NREL/CP-5400-67618, 2017, doi: [10.4271/15472450.2017.1291351](https://doi.org/10.4271/15472450.2017.1291351).
- [11] D. Milakis, B. van Arem, and B. van Wee, "Policy and society related implications of automated driving: A review of literature and directions for future research," *J. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 324–348, Jul. 2017, doi: [10.1080/15472450.2017.1291351](https://doi.org/10.1080/15472450.2017.1291351).
- [12] M. J. Barth, "The effects of AHS on the environment," in *Automated Highway Systems*. Boston, MA, USA: Springer, 1997, doi: [10.1007/978-1-4757-4573-3_13](https://doi.org/10.1007/978-1-4757-4573-3_13).
- [13] A. Davila and M. Nombela, "Platooning-safe and eco-friendly mobility," SAE Tech. Paper 2012-01-0488, 2012, p. 5, doi: [10.4271/2012-01-0488](https://doi.org/10.4271/2012-01-0488).
- [14] C. Zhai, Y. Liu, and F. Luo, "A switched control strategy of heterogeneous vehicle platoon for multiple objectives with state constraints," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1883–1896, May 2019.
- [15] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 3, pp. 899–910, May 2017.
- [16] K.-Y. Liang, J. Martensson, and K. H. Johansson, "Fuel-saving potentials of platooning evaluated through sparse heavy-duty vehicle position data," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 1061–1068, doi: [10.1109/IVS.2014.6856540](https://doi.org/10.1109/IVS.2014.6856540).
- [17] C. Bonnet and H. Fritz, "Fuel consumption reduction in a platoon: Experimental results with two electronically coupled trucks at close spacing," SAE Tech. Paper 2000-01-3056, 2000, p. 9, doi: [10.4271/2000-01-3056](https://doi.org/10.4271/2000-01-3056).
- [18] J. A. Vernet, R. Örlü, and P. H. Alfredsson, "Flow separation control by dielectric barrier discharge plasma actuation via pulsed momentum injection," *AIP Adv.*, vol. 8, no. 7, Jul. 2018, Art. no. 075229, doi: [10.1063/1.5037770](https://doi.org/10.1063/1.5037770).
- [19] C. Thorpe, T. Jochem, and D. Pomerleau, "Automated highways and the free agent demonstration," in *Robotics Research*. London, U.K.: Springer, 1998, pp. 246–254.
- [20] S. Ostria, M. Lawrence, and D. Pickrell, "Capacity-induced increases in the quantity of travel with special reference to IVHS," in *Proc. Nat. Policy Conf. Intell. Transp. Syst. Environ.*, Arlington, VA, USA, 1994, p. 206.
- [21] A. A. Hussein and H. A. Rakha, "Vehicle platooning impact on drag coefficients and energy/fuel saving implications," 2020, *arXiv:2001.00560*. [Online]. Available: <http://arxiv.org/abs/2001.00560>
- [22] R. Thirumalainambi and J. Bardina, "Training data requirement for a neural network to predict aerodynamic coefficients," *Proc. SPIE*, vol. 5102, pp. 92–103, Apr. 2003.
- [23] C.-H. Bruneau, K. Khadra, and I. Mortazavi, "Flow analysis of square-back simplified vehicles in platoon," *Int. J. Heat Fluid Flow*, vol. 66, pp. 43–59, Aug. 2017, doi: [10.1016/j.ijheatfluidflow.2017.05.008](https://doi.org/10.1016/j.ijheatfluidflow.2017.05.008).
- [24] S. T. Kaluva, A. Pathak, and A. Ongel, "Aerodynamic drag analysis of autonomous electric vehicle platoons," *Energies*, vol. 13, no. 15, p. 4028, Aug. 2020.
- [25] K. Tadakuma, T. Doi, M. Shida, and K. Maeda, "Prediction formula of aerodynamic drag reduction in multiple-vehicle platooning based on wake analysis and on-road experiments," *SAE Int. J. Passenger Cars-Mech. Syst.*, vol. 9, no. 2, pp. 645–656, Apr. 2016, doi: [10.4271/2016-01-1596](https://doi.org/10.4271/2016-01-1596).
- [26] R. Lietz, L. Larson, P. Bachant, J. Goldstein, R. Silveira, M. Shademan, P. Ireland, and K. Mooney, "An extensive validation of an open source based solution for automobile external aerodynamics," SAE Tech. Paper 2017-01-1524, 2017, p. 16.
- [27] B. N. Hanna, N. T. Dinh, R. W. Youngblood, and I. A. Bolotnov, "Coarse-grid computational fluid dynamic (CG-CFD) error prediction using machine learning," *J. Fluids Eng.*, p. 68, Oct. 2017. [Online]. Available: <https://arxiv.org/abs/1710.09105v1>
- [28] M. Zabat, N. Stabile, S. Farascari, and F. Browand, "The aerodynamic performance of platoons: A final report," California Partners Adv. Transp. Technol., UC Berkeley, Berkeley, CA, USA, Tech. Rep. UCB-ITS-PRR-95-35, 1995.
- [29] Fluent Ansys. (2015). *Academic Research*. Release 14, p. 3. [Online]. Available: <http://community.hartree.stfc.ac.uk/wiki/site/admin/fluent.pdf>
- [30] *Amazon Web Services*. Accessed: May 15, 2020. [Online]. Available: <http://aws.amazon.com/es/ec2/>
- [31] P. Kutler, U. B. Mehta, and A. Andrews, "Progress and challenges in the application of artificial intelligence to computational fluid dynamics," *AAI J.*, vol. 26, no. 1, pp. 40–46, 1988, doi: [10.2514/3.9848](https://doi.org/10.2514/3.9848).
- [32] P. Kutler, U. B. Mehta, and A. Andrews, "Potential application of artificial intelligence concepts to numerical aerodynamic simulation," in *Proc. 9th Int. Conf. Numer. Methods Fluid Dyn.* Berlin, Germany: Springer, 1985, doi: [10.1007/3-540-13917-6_160](https://doi.org/10.1007/3-540-13917-6_160).
- [33] P. Kutler and U. Mehta, *Computational Aerodynamics and Artificial Intelligence*. Snowmass, CO, USA: American Institute of Aeronautics and Astronautics, 2012, doi: [10.2514/6.1984-1531](https://doi.org/10.2514/6.1984-1531).
- [34] L. He and D. K. Tafti, "A supervised machine learning approach for predicting variable drag forces on spherical particles in suspension," *Powder Technol.*, vol. 345, pp. 379–389, Mar. 2019, doi: [10.1016/j.powtec.2019.01.013](https://doi.org/10.1016/j.powtec.2019.01.013).
- [35] E. Gunpinar, U. C. Coskun, M. Ozsipahi, and S. Gunpinar, "A generative design and drag coefficient prediction system for sedan car side silhouettes based on computational fluid dynamics," *Comput.-Aided Des.*, vol. 111, pp. 65–79, Jun. 2019, doi: [10.1016/j.cad.2019.02.003](https://doi.org/10.1016/j.cad.2019.02.003).
- [36] P. Dube and S. Hiravennavar, "Machine learning approach to predict aerodynamic performance of underhood and underbody drag enablers," General Motors Tech. Center, Bengaluru, India, Tech. Rep. 2020-01-0684, 2020.
- [37] K. B. Kelly and H. J. Holcombe, "Aerodynamics for body engineers," *SAE Trans.*, vol. 72, pp. 570–578, 1964.
- [38] M. Michaelian and F. Browand, "Field experiments demonstrate fuel savings for close-following," California Partners Adv. Transp. Technol., UC Berkeley, Berkeley, CA, USA, Tech. Rep. UCB-ITS-PRR-2000-14, 2000.
- [39] D. K. Murthy and A. Masrur, "Braking in close following platoons: The law of the weakest," in *Proc. Euromicro Conf. Digit. Syst. Design (DSD)*, Aug./Sep. 2016, pp. 613–620.

- [40] C. Zhai, F. Luo, Y. Liu, and Z. Chen, "Ecological cooperative look-ahead control for automated vehicles travelling on freeways with varying slopes," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1208–1221, Feb. 2019, doi: [10.1109/TVT.2018.2886221](https://doi.org/10.1109/TVT.2018.2886221).
- [41] B. Sakhdari and N. L. Azad, "A distributed reference governor approach to ecological cooperative adaptive cruise control," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1496–1507, May 2018.
- [42] Z. Zhou, Z. Akhtar, K. L. Man, and K. Siddique, "A deep learning platooning-based video information-sharing Internet of Things framework for autonomous driving systems," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 11, pp. 1–11, 2019, doi: [10.1177/1550147719883133](https://doi.org/10.1177/1550147719883133).
- [43] J. Back, S. Choi, and Y. Shin, "A study on the election of suitable leader vehicle in vehicle platooning using the raft algorithm," in *Proc. 11th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2019, pp. 756–761.
- [44] T. Farid, A. Shakeel, and M. Sajid, "An evaluation of open source CFD for study of aerodynamics of vehicle platooning," in *Proc. ASME-JSME-KSME 8th Joint Fluids Eng. Conf.*, vol. 2, pp. V002T02A068-1–V002T02A068-7, 2019, doi: [10.1115/AJKFluids2019-5130](https://doi.org/10.1115/AJKFluids2019-5130).
- [45] T. Mansi, B. Georgescu, J. Hussan, P. J. Hunter, A. Kamen, and D. Comaniciu, "Data-driven reduction of a cardiac myofilament model," in *Functional Imaging and Modeling of the Heart*, S. Ourselin, D. Rueckert, and N. Smith, Eds. Berlin, Germany: Springer, 2013, pp. 232–240.
- [46] K. Tøndel, U. G. Indahl, A. B. Gjuvslund, J. Vik, P. Hunter, S. W. Omholt, and H. Martens, "Hierarchical cluster-based partial least squares regression (HC-PLSR) is an efficient tool for metamodeling of nonlinear dynamic models," *BMC Syst. Biol.*, vol. 5, no. 1, p. 90, 2011, doi: [10.1186/1752-0509-5-90](https://doi.org/10.1186/1752-0509-5-90).
- [47] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, and A. G. Doyle, "Predicting reaction performance in C–N cross-coupling using machine learning," *Science*, vol. 360, no. 6385, pp. 186–190, Apr. 2018.
- [48] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, Dec. 2018, doi: [10.1109/TCCN.2018.2881442](https://doi.org/10.1109/TCCN.2018.2881442).
- [49] A. F. Sheta, S. E. M. Ahmed, and H. Faris, "A comparison between regression, artificial neural networks and support vector machines for predicting stock market index," *Soft Comput.*, vol. 4, no. 7, pp. 1–9, 2015.
- [50] Z. Liu, D. Wu, Y. Liu, Z. Han, L. Lun, J. Gao, G. Jin, and G. Cao, "Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction," *Energy Explor. Exploitation*, vol. 37, no. 4, pp. 1426–1451, Jul. 2019, doi: [10.1177/0144598718822400](https://doi.org/10.1177/0144598718822400).
- [51] S. R. Ahmed, G. Ramm, and G. Faltn, "Some salient features of the time-averaged ground vehicle wake," *SAE Trans.*, vol. 93, pp. 473–503, 1984.
- [52] M. Lanfrit, "Best practice guidelines for handling automotive external aerodynamics with FLUENT," Fluent Deutschland GmbH, Darmstadt, Germany, Tech. Rep. Version 1.2 (Feb 9th 2005), 2005.
- [53] K. A. Marill, "Advanced statistics: Linear regression, Part I: Simple linear regression," *Academic Emergency Med.*, vol. 11, no. 1, pp. 87–93, Jan. 2004, doi: [10.1197/j.aem.2003.09.005](https://doi.org/10.1197/j.aem.2003.09.005).
- [54] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [55] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- [56] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 155–161.
- [57] J. Rabault, M. Kuchta, A. Jensen, U. Reglade, and N. Cerardi, "Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control," 2018, *arXiv:1808.07664*. [Online]. Available: <http://arxiv.org/abs/1808.07664>
- [58] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: <http://arxiv.org/abs/1803.08375>
- [59] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018, *arXiv:1811.03378*. [Online]. Available: <http://arxiv.org/abs/1811.03378>
- [60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [61] A. W. Moore, "Cross-validation for detecting and preventing overfitting," *School Comput. Sci.*, Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., 2001.



FARWA JAFFAR received the bachelor's degree in computer science from the Islamia University of Bahawalpur, and the M.Eng. degree in robotics and intelligent machine engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2016 and 2020, respectively. Her research interest includes interdisciplinary applications of machine learning.



TAHA FARID received the degree in mechanical engineering from the School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), Pakistan. He has research experience in the field of computational fluid dynamics (CFD) of autonomous vehicle platooning using open source code. His research interests include aimed at addressing the problems of design, optimization, uncertainty quantification in computational engineering using artificial intelligence, cloud computing, and the Internet of Things (IoT).



MUHAMMAD SAJID received the B.Eng. degree from the National University of Sciences and Technology (NUST), the master's degree from the École Nationale Supérieure d'Arts et Métiers (ENSAM) ParisTech, France, and the Ph.D. degree from the University of Cergy Pontoise, France, in 2010. He completed a Postdoctoral Research from Texas A&M University, Qatar. He is currently an Assistant Professor with the National University of Sciences and Technology (NUST), Pakistan. His research interests include experimental and computational fluid dynamics, and flow visualization, as well as multiphase fluid flow and heat transfer phenomena.



YASAR AYAZ (Senior Member, IEEE) received the Ph.D. degree in robotics and machine intelligence from Tohoku University, Sendai, Japan, in 2009. He is currently a specially appointed Associate Professor with the Graduate School of Engineering, Tohoku University. He is also the Chairman and the Central Project Director of the Pakistan's National Center of Artificial Intelligence (NCAI) and also the Head of the Department of Robotics and Intelligent Machine Engineering, NUST, Pakistan. His research interests include human-machine interaction, humanoids, and motion planning.



MUHAMMAD JAWAD KHAN received the B.E. and M.S. degrees in mechatronics engineering from Air University, Pakistan, in 2007 and 2010, respectively, and the Ph.D. degree in mechanical engineering from Pusan National University, South Korea, in 2018. He was involved in hybrid EEG-fNIRS-based BCI for rehabilitation during his Ph.D. He is currently an Assistant Professor with the School of Mechanical and Manufacturing Engineering, NUST, Pakistan. His research interests include hybrid brain-computer interfacing, rehabilitation, and machine learning.

...