The Aeronautical Journal (2025), **129**, pp. 1534–1549
doi:10.1017/aer.2024.164

RESEARCH ARTICLE

Aircraft takeoff speed prediction with machine learning: parameter analysis and model development

N.N. Karaburun[1], S. Arık Hatipoğlu[2] and M. Konar[2]

[1]Graduate School of Natural and Applied Science, Erciyes University, Kayseri, Türkiye
[2]Faculty of Aeronautics and Astronautics, Erciyes University, Kayseri, Türkiye
Corresponding author: S. Arık Hatipoğlu; Email: arikseda@erciyes.edu.tr

Received: 31 October 2024; **Revised:** 13 December 2024; **Accepted:** 16 December 2024

Keywords: aircraft; machine learning; takeoff speed; hyperparameter optimisation

Abstract

With developing technology, the usage areas of aircraft are constantly expanded. In aircraft designed for different missions, it is an important issue to evaluate many design possibilities and make optimum designs by taking into account many parameters that are not directly connected to each other with equal importance. In this context, issues such as safety and performance come to the fore in aircraft designs. One of the critical situations affecting flight safety is the takeoff and landing phases of the aircraft. The speed changes that occur in these stages are an important issue. In this study, takeoff speed was predicted with different machine learning algorithms using takeoff speed data of the Boeing B-737-300 type aircraft. Linear regression, support vector regression, classification and regression trees, random forest regression, Extreme Gradient Boosting algorithms were selected from machine learning algorithms for takeoff speed prediction. Base models were created with these selected algorithms and the takeoff speed was predicted by training the data. Considering the obtained results, feature engineered was applied to minimise the error values of the proposed base models. In models developed by applying feature engineered, error values were reduced and better performance was observed in takeoff speed prediction. Takeoff speed values obtained with the developed models and actual flight speed values are presented comparatively for the first time in the literature. The simulation results emphasise that the developed models can be used as an effective and alternative method for takeoff speed prediction.

Nomenclature

AI	artificial intelligence
ANN	artificial neural network
CART	classification and regression trees
GBDT	gradient boosting decision tree
IoT	Internet of Things
LR	linear regression
LSTM	long short-term memory
LSTM-RNN	long short-term memory recurrent neural network
MAE	mean absolute error
MEMS	micro electro-mechanical systems
MSE	mean square error
NODE	neural ordinary differential equations
PoF	phase of flight
QAR	quick access recorder
RF	random forest
RMSE	root mean square error
SVM	support vector machine

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal Aeronautical Society.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

SVR	support vector regression
UAV	unmanned aerial vehicle
XGBoost	Extreme Gradient Boosting

1.0 Introduction

Modern aviation has a wide range of uses, from transportation to defense industry [1–4]. In addition, with technological developments, demand and research in the field of aviation has also increased [5]. Increasing demand has led to significant developments in aircraft produced for various usage purposes, and aircraft of different sizes and types have been produced. The aircraft produced continue to be constantly improved in order to a chieve flights that are more effective by prioritising issues such as safety and performance [6]. In this context, it is important to evaluate many possibilities and make optimum designs in order for highly complex systems that involve the interaction of various interconnected parameters to perform safe and efficient missions in aircraft developed [7, 8]. However, effectiveness and safety of aircraft have always been the focus of aircraft operators and designers. The takeoff and landing phases, which are the most critical phases of every flight, are the phases where 30% of accidents occur, but it is stated that these phases are only 2% of the average flight time [9, 10].

The changing speeds of aircraft during their landing, takeoff and endurance are affected by various parameters. Parameters affecting the speed of the aircraft include various factors such as weather conditions, aerodynamic factors, aircraft weight, fuel amount and engine performance. There are many studies in the literature to accurately analyse and evaluate these parameters that affect the speed of the aircraft. One of the methods used in these studies is artificial intelligence applications. Artificial intelligence applications are preferred in many areas such as predicting possible dangers during flight, determining optimal speeds and making real-time recommendations to pilots.

Diallo presented an approach using artificial neural networks (ANNs) to estimate landing speed. They stated that the model they developed accurately predicted the landing speed for windless and high wind conditions [11].

Hu et al. proposed a model using the support vector machine (SVM) method to predict hard landing events during the landing phase of the aircraft. In their analysis, they emphasised that the SVM model they proposed was successful in predicting hard landing events [12].

Tong et al. proposed a long short-term memory (LSTM)-based architecture that they developed using artificial intelligence (AI) and Internet of Things (IoT)-based sensor data in order to estimate aircraft landing speed. They stated that the results obtained showed better performance compared to traditional methods [13].

Zhang and Zhu stated that they successfully captured the long-term dependencies of quick access recorder (QAR) data by using the LSTM model to predict hard landings, and that the model demonstrated superior ability compared to other traditional hard landing prediction methods [14].

Borup et al. preferred ANNs and linear regression (LR) algorithm to estimate angle-of-attack, sideslip angle and airspeed using low-cost micro electro-mechanical systems (MEMS) based pressure sensors integrated on unmanned aerial vehicle (UAV). In their simulations, they stated that they generally obtained lower error values with the ANN algorithm compared to LR [15].

Kovarik et al. performed simulations with long short-term memory recurrent neural network (LSTM-RNN), SVM and neural ordinary differential equations (NODE) using many flight data to improve phase of flight (PoF) estimation. In their results, they emphasised that LSTM-RNN models are more suitable for PoF prediction [16].

In the work presenting by Puranik et al., they proposed a model based on the random forest (RF) algorithm to estimate the landing true airspeed and ground speed. They emphasised that their proposed model performed better than existing techniques in the literature in estimating the true air speed and ground speed during landing [10].

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

Using flight data records in their study, Jarry et al. examined target parameters such as fuel flow rate, flap and landing gear setting with a model based on LSTM [17].

Kang et al. proposed a model based on LSTM architecture using QAR data to predict aircraft landing speed. Because of their simulations, they emphasised that the prediction accuracy of their proposed method is significantly higher than traditional methods [18].

In their study, Ozkat et al. used vibration data collected from a multi-rotor UAV to estimate the remaining useful life of the UAV with the LSTM method [19].

Ozkat proposed a wavelet scattering LSTM autoencoder network to detect propeller failures at an early stage by analysing vibration signals. He made predictions with the proposed method using the vibration data generated by making an intentional change in one of the propeller blades of a multi-rotor UAV and emphasised that he obtained satisfactory results [20].

Karaburun et al. estimated the battery state of charge (SOC) using SVR, RF and LSTM methods in order to estimate the UAV's flight time. They stated that the estimated values they obtained were consistent with the true values and emphasised that the methods they proposed could be used effectively in battery SOC estimation [21].

When the studies in the literature are examined, there are no models in which different machine learning algorithms are used and compared in takeoff speed prediction. In this study, takeoff speed prediction was performed using different machine learning algorithms. Many parameters affect aircraft takeoff speed, but in this study, depending on the data set shared by Boeing, pressure altitude, temperature, weight, flap angle data were selected as independent variables, and takeoff speed was selected as dependent variable [22]. The selected dependent-independent variables were trained with machine learning algorithms such as LR, (SVR), classification and regression trees (CART), RF and Extreme Gradient Boosting (XGBoost), and, the results obtained were presented comparatively for the first time in the literature. The presented results emphasise that machine learning algorithms can be used as an alternative and innovative method for takeoff speed prediction.

2.0 Machine learning

Machine learning, a sub-branch of artificial intelligence, is the technology of developing computer algorithms that can imitate human intelligence and perform self-learning. It is the modeling of systems, which make predictions by making inferences on data through mathematical and statistical processes, by computers [23]. Machine learning algorithms are divided into three categories: supervised, unsupervised and reinforcement learning.

i. Supervised learning examines how the relationships between the dependent and the independent variables are learned based on the label information of the data set. Training data consists of both inputs and outputs. If the outputs are categorical, classification algorithms are used, if they are numerical, regression algorithms are used [24].
ii. Unsupervised learning aims to understand and discover the relationships between unlabeled data. In this type of learning, there is no dependent variable and there is data without a cause and effect relationship. Clustering operations are performed according to the similarities of the observations to each other [25].
iii. Reinforcement learning is based on an agent using the trial-and-error method to interact with the environment in a certain situation and complete the given task in the most successful way. Its main purpose is to develop the optimal strategy to obtain the most reward in order to achieve the best result [26].

The machine learning process, whose flow diagram is given in Fig. 1, is done by performing certain steps. First, data is collected, subjected to preprocessing steps, and received as input. In the preprocessing stage, data cleaning, checking of missing data and taking appropriate actions according to the problem situation are carried out [27]. Next, model selection and feature engineered steps are applied.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

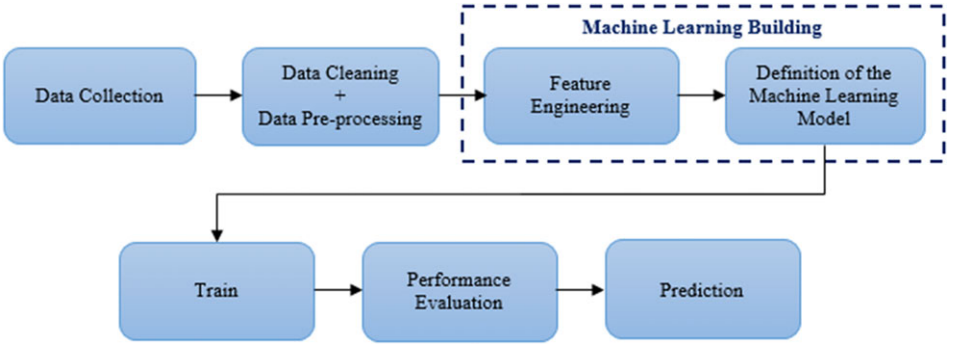


Figure 1. *Flow diagram of the machine learning process.*

It is very important that the model selection is determined with appropriate features depending on the type of problem. Feature engineered is a critical preliminary step for machine learning algorithms. Feature extraction is the process of identifying and removing unnecessary features in order to reduce the problem size and make the algorithms work faster and more effectively. After feature engineered is implemented, the selected model is properly installed. Necessary parameter adjustments of the established model are made to obtain better results. As a result of model training, certain performance evaluation metrics are used to evaluate the performance of the model and model development steps are applied when necessary [28–30]. Finally, the predictive values of the model are obtained and the results are analysed. This process can be repeated and continuously improved in machine learning projects [31].

The types of machine learning algorithm preferred in this study are briefly explained below.

2.1 Linear regression

Linear regression (LR) is a supervised learning algorithm that determines how the dependent variable constantly changes according to a set of independent variables and the relationship between them [32]. The widely used linear regression is generally preferred in small data sets thanks to its simple structure. The parameters of the model can be easily interpreted. This makes it easier for analysts to understand the relationships between variables and explain the outputs of the model [33]. The mathematical model of simple LR is given in Equation (1) [34].

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{1}$$

Here, n is the number of individuals ($i = 1, \ldots, n$), y_i is the dependent variable, x_i is the independent variable, e_i is a random error term representing uncertainties, and β_0 and β_1 are the regression coefficients [34].

2.2 Support vector regression

Support vector machines (SVM) consist of two basic categories as classification and solving regression problems. The basic working principle is to perform classification or regression operations on data points by creating a hyperplane [35, 36].

SVM used in regression problems are called as support vector regression (SVR). SVR is a supervised learning model that aims to achieve good performance on both linear and nonlinear data sets. SVR tries to learn an optimal hyperplane with the most points within a margin using support vectors representing the closest data points on the hyperplane. This margin provides an organisation that is increasing the generalisation ability of the model [37].

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

The relationship between input and output in SVR is formulated with a linear function as given in Equation (2) [38].

$$f(x) = w^T \phi(x) + b \tag{2}$$

Here, $f(x)$ is the predicted value, w is the weight vector, and $\phi(x)$ is the kernel function. Various kernel functions such as linear, polynomial, radial basis function and sigmoid, which form the basis of SVR, can be selected depending on the problem and data set and are used to move the data to a high-dimensional feature space [39]. b is the bias value and provides the effect that determines how the classes will be separated.

2.3 Classification and regression trees

The classification and regression trees (CART) algorithm is a machine learning algorithm used in many applications such as data mining, classification, regression and feature selection. CART, which has the main purpose of transforming complex structures into simple decision structures, is an algorithm built on a single tree where a set of decision rules are applied [40]. CART starts with the root node at the top and each node represents a feature. The root node represents the most important variable. The features are divided into sub-branches [41]. When a structured variable is determined, CART creates a classification tree. When a continuous variable is determined, CART creates a regression tree. The tree structure makes decisions by evaluating the features in the data set and classifies the data or makes regression predictions for data set with these decisions [42]. The clarity and interpretability of the tree structure is one of the advantages that is effective in choosing the algorithm. However, being based on a single tree can lead to limitations due to instability. For this reason, ensemble methods such as random forest, where many trees are evaluated simultaneously, have been developed [43].

2.4 Random forest regression

Random forest (RF) is an ensemble learning method that works by combining many decision trees for classification and regression operations. It aims to create a general and effective model by combining different learning algorithms. RF regression combines the predictions of decision trees and uses the voting method for classification problems and the mean value for regression problems [44, 45]. The basic working principle of the RF algorithm is given in Equation (3).

$$\hat{y} = \frac{1}{n} \sum_{1=i}^{1} h_i(x) \tag{3}$$

Here, $\hat{y}$ is the general estimate, n is the total number of trees, h_i is the estimated value of the decision tree, and x is the mean of the independent variables [46].

In RF regression, decision trees are combined into a hierarchical structure. This arrangement, which prevents the over-learning problem of a single tree structure, ensures that each tree has a hierarchical structure. Each tree represents the applied constraints, starting from a root node. The diversity between the trees reduces the tendency of the model to over-learning, allowing it to make more general and reliable predictions [42, 47, 48].

The advantages of RF include working effectively on complex data sets, being resistant to over-learning and outliers. It is used to obtain successful results especially on large and complex data sets [24]. These features make RF stand out as a machine learning method with wide application areas.

2.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a decision tree-based ensemble machine learning algorithm. Gradient boosting decision tree (GBDT) algorithm is a XGBoost version developed to increase speed and prediction performance. It performs well in both classification and regression problems. The XGBoost algorithm uses a mathematical function expressed as the sum of the predictions ($\hat{y}_i$) of the

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

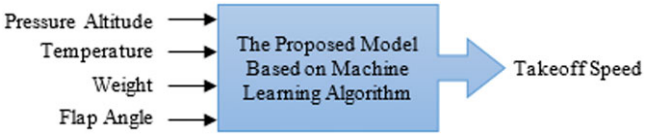


Figure 2. *Block diagram of the proposed models for takeoff speed prediction.*

total k trees in an ensemble, as given in Equation (4). In this function, the prediction of each tree for observation number i is represented by $f_k(x_i)$ [49].

$$\hat{y}_i = \sum_{k=1}^{k} f_k(x_i), f_k \in F \tag{4}$$

The main purpose of the algorithm is to improve the tree created from continuously splitting trees according to the residual error between the prediction results of the previous one and the true value. The new prediction value is obtained by adding it to the previous prediction value [49, 50]. As a result, the trees are connected to each other and a tree tries to correct the residual errors of the predictions of the previous trees.

3.0 Problem definition

In this section, the steps followed for takeoff speed prediction using machine learning algorithms and the proposed models are discussed.

First, the data set was determined. For this purpose, the aircraft takeoff data shared by Boeing was used. This data set consists of the pressure altitude in feet (ft) during aircraft takeoff, the outer environmental temperature in Fahrenheit (F), the aircraft's weight in pounds (lb), the angle of the flaps on its wings in degrees (°) and the aircraft's speed in knots (kn). The lower and upper limits given in the data set for these variables are 0-9000 ft for pressure altitude, (-40)-(160) F for temperature, 75–140lb for weight, 1°–15° for flap angle, and 98–162kn for takeoff speed, respectively [22].

In the model proposed in this study, the pressure altitude, the outer environmental temperature, aircraft's weight and flap angle were selected as independent variables, and the takeoff speed was selected as the dependent variable. Analytical methods are not very suitable for takeoff speed prediction using this complex relationship consisting of independent-dependent variables. For this reason, artificial intelligence techniques are used to solve problems with complex relationships that do not have a direct relationship between them. In this study, machine learning, one of the artificial intelligence techniques, was preferred in the takeoff speed prediction with complex relationships. The block diagram of the proposed method is presented in Fig. 2.

After determining the data, LR, SVR, CART, RF and XGBoost algorithms were selected from machine learning algorithms for training and test the data. Seventy per cent of the data set was used for training and 30% for testing, and the models were created accordingly. These algorithms were modeled using Anaconda Python software.

The root mean square error (RMSE) expressed by Equation (5), the mean square error (MSE) expressed by Equation (6) and the mean absolute error (MAE) expressed by Equation (7) were used as performance criteria to evaluate the simulation performance. The smaller the error value, the higher the model accuracy is considered. Therefore, these metrics were prefered to objectively evaluate the performance of the proposed models [51].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{5}$$

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

Table 1. *The RMSE, MSE and MAE values of the proposed base models*

	Base models		
Model	RMSE	MSE	MAE
LR	1.731	2.995	1.247
SVR	2.104	4.428	1.636
CART	1.284	1.649	0.524
RF	1.162	1.350	0.51
XGBoost	**0.592**	**0.35**	**0.282**

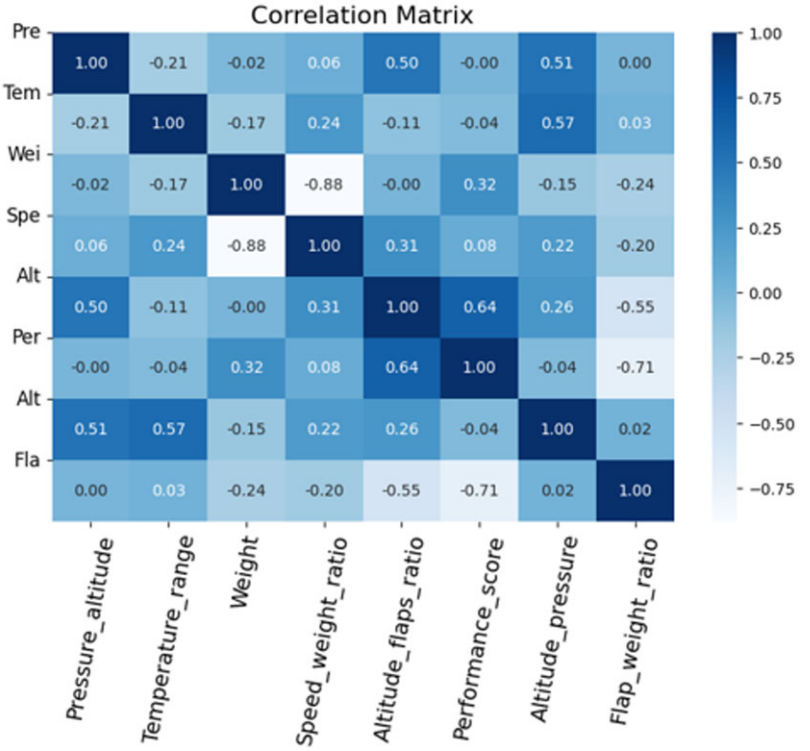


Figure 3. *The correlation map of the numerical data created with these new variables.*

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}i\right)^2 \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{7}$$

4.0 Simulation results

In this section, the results of the study are presented in detail.

The first step of the study is to determine the data. For this purpose, the aircraft takeoff data shared by Boeing was used. The parameters of pressure altitude, the outer environment temperature, aircraft's weight, the angle of the flaps were selected as independent variables and takeoff speed was selected

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

Table 2. *The RMSE, MSE and MAE values of the proposed new models with feature engineered*

	Models with feature engineered		
Model	RMSE	MSE	MAE
LR	1.355	1.836	1.027
SVR	1.974	3.898	1.257
CART	0.756	0.572	**0.09**
RF	0.733	0.538	0.115
XGBoost	**0.615**	**0.379**	0.144

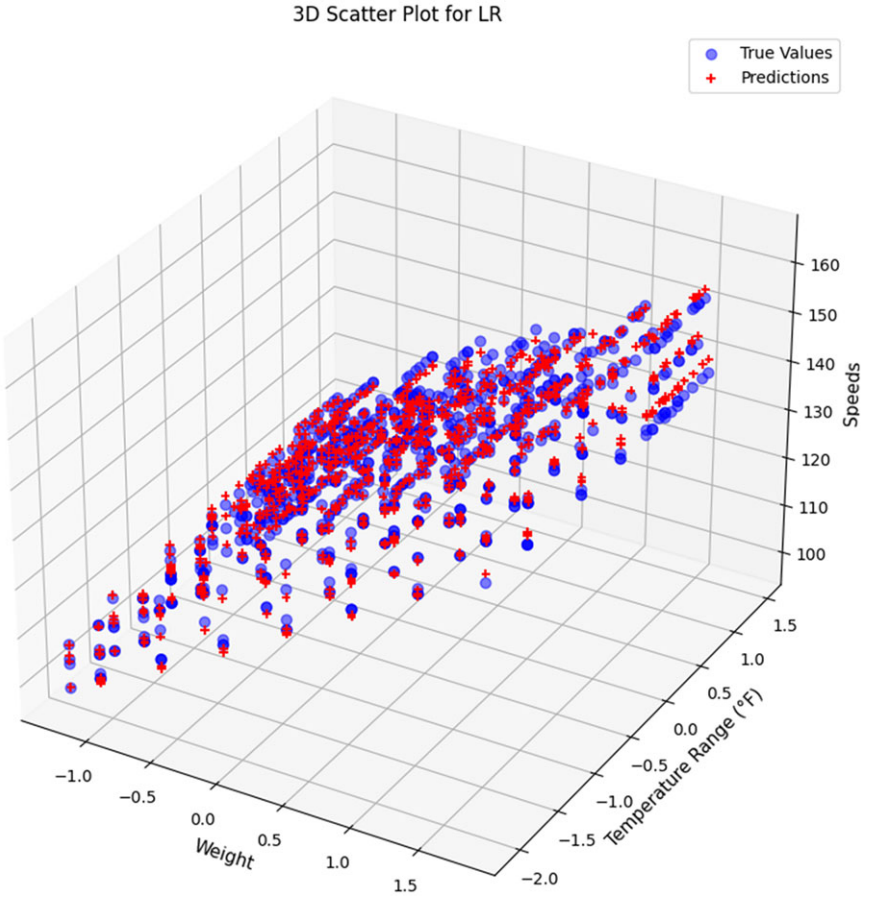


Figure 4. *Comparison of actual takeoff speed values with predicted takeoff speed values for LR.*

as dependent variable. Then, the data pre-processing step was performed. In this step, missing values, outliers and variable types were checked and necessary corrections were made. The StandardScaler operation in the Scikit-Learn library was applied to scale the data. Thus, each variable was standardised. The standardised data was separated as 70% train data and 30% test data with the holdout method. Models were generated using LR, SVR, CART, RF and XGBoost algorithms from machine learning algorithms. The generated models were trained with the training data and evaluated with the test data.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

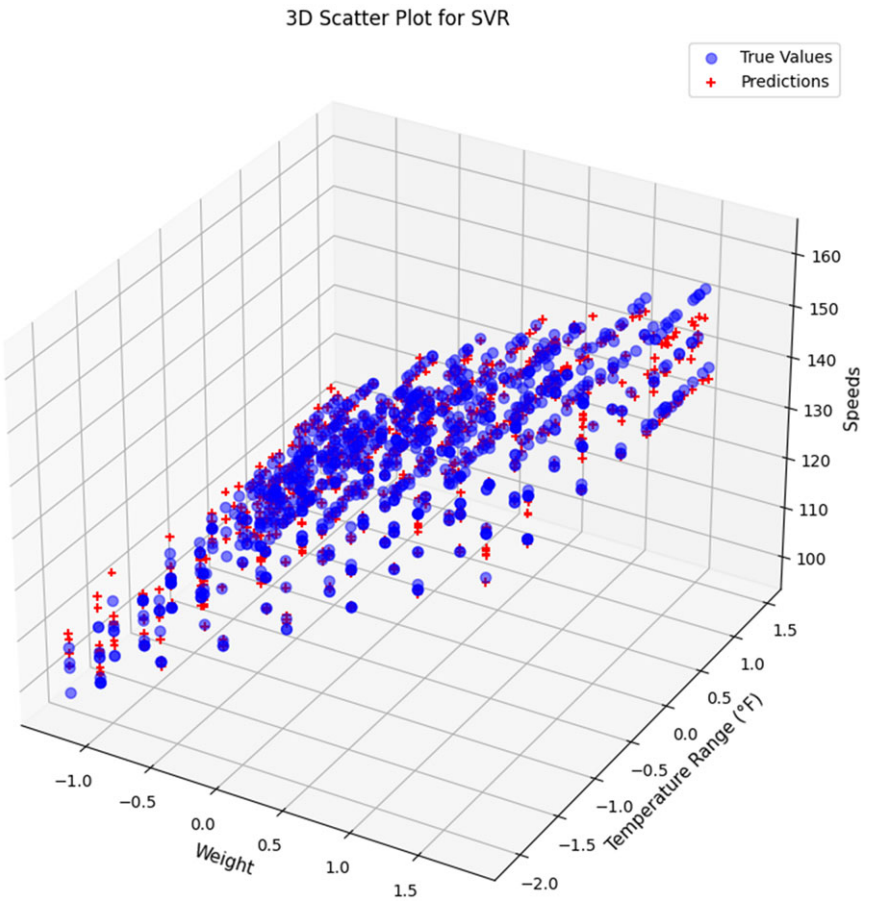


Figure 5. *Comparison of actual takeoff speed values with predicted takeoff speed values for SVR.*

For takeoff speed estimation, the values of the performance criteria of the simulation results made with base models are presented in Table 1.

When Table 1 is examined, the largest error metric value is obtained in the SVR model, while the smallest error metric value is obtained in the XGBoost model. Considering the error values of the proposed base models to improve the performance of the selected models, feature engineered steps were applied and new variables were added. The new variables derived in this study represent important factors that evaluate the performance of the aircraft. Speed_weight_ratio measures agility of the aircraft by showing the ratio of speed to weight. Altitude_flaps_ratio evaluates the relationship between pressure altitudes and flaps. Performance_score measures overall performance by considering the interaction between speed, weight and flaps. Altitude_pressure reflects the interaction between pressure altitude and temperature. Flap_weight_ratio affects manoeuverability by determining the ratio of flaps to weight. These characteristics help understand the aircraft's efficiency. The correlation map of the numerical data created with these new variables is presented in Fig. 3.

When the correlation map of the numerical data in Fig. 3 is examined, since the flap angle is a categorical variable, it is not directly specified in the correlation map. However, new variables were generated using the flap angle data. Remodelings and predictions were made using the updated data set with the addition of new variables. For this purpose, scaling was first applied to the updated data. Then, the scaled data was separated into training and test data at the same rates in the base model using the

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

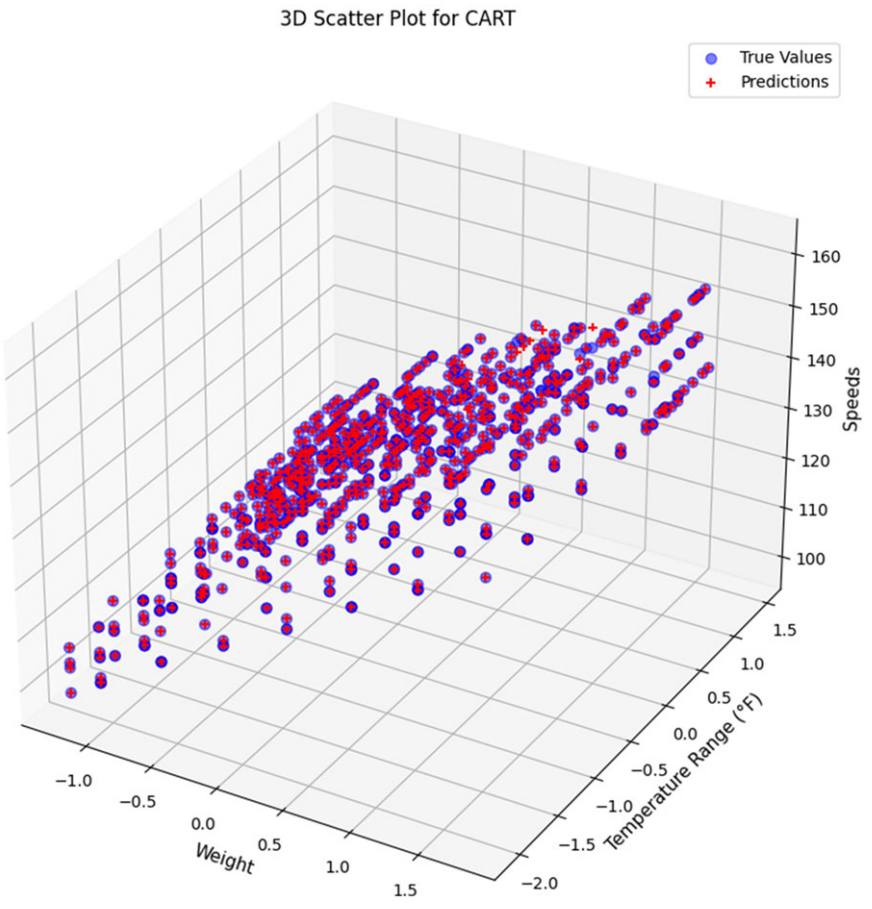


Figure 6. *Comparison of actual takeoff speed values with predicted takeoff speed values for CART.*

holdout method. Models were generated again using LR, SVR, CART, RF and XGBoost algorithms. The new models generated were trained with training data and evaluated with test data. For takeoff speed prediction, the RMSE, MSE and MAE values of the simulation results made with the new models obtained by applying feature engineered are presented in Table 2.

When Table 2 is examined, in the takeoff speed prediction made with the new feature engineered models, the largest error metric value is obtained in the SVR model, while the smallest error metric value is obtained in the XGBoost model. Although there is a minimal increase in the error values in the feature engineered model compared to the base model for XGBoost, when Table 2 is compared to Table 1, it is generally seen that there is an improvement in the error values of the proposed models. In other words, it is seen that the models generated with the feature engineered data perform better than the base models. The comparison of the actual takeoff speed values and the takeoff speed values estimated with models established with the feature engineered data is given from Figs. 4 to 8, respectively, for the LR, SVR, CART, RF and XGBoost algorithms.

The comparison of the true takeoff speed values with the predicted takeoff speed values is presented in the first column (A) of Fig. 9. The graphs of true versus predicted values make it easier to visually evaluate the model accuracy and variability with the regression line and 95% confidence interval. The graphs of the residuals versus predicted values in the second column (B) of Fig. 9 analyse the distribution of errors in the model predictions. In these graphs, the red line representing zero error provides a reference for evaulating the accuracy of the predictions.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

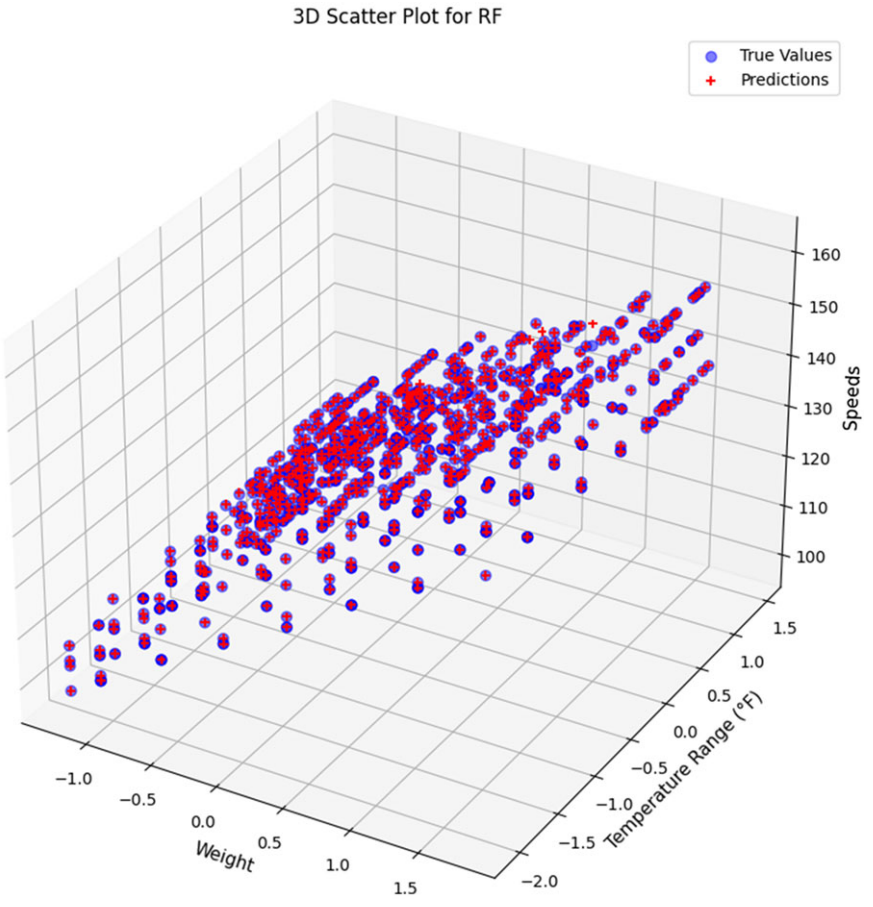


Figure 7. *Comparison of actual takeoff speed values with predicted takeoff speed values for RF.*

When the figures comparing the actual takeoff speed values with the estimated takeoff speed values are examined, satisfactory results are observed. However, another method that can be done to increase the model performance is hyperparameter optimisation. In this study, when Table 2 is taken as reference, it is seen that the XGBoost algorithm has a smaller error value compared to the other models. Therefore, hyperparameter optimisation was applied to the XGBoost algorithm to increase the model performance. For this purpose, learning_rate, max_depth, n_estimators, colsample_bytree hyperparameters were selected for optimisation. learning_rate limits the contribution of each tree and its default value is 0.3. max_depth controls the maximum depth of the trees and its default value is 6. n_estimators affects the complexity of the model and its default value is 100. colsample_bytree determines the percentage of random selection of features for each tree and its default value is 1. The value ranges selected for hyperparameter optimisation in this study are given in Table 3.

Using GridSearchCV, hyperparameter optimisation on the XGBoost model was performed with using the value ranges in Table 3. he optimisation result was obtained as learning_rate=0.05, max_depth=6, n_estimators=1000, colsample_bytree=1. As a result of the simulation performed with the new XGBoost model established with the obtained optimum hyperparameter values, the RMSE value was found as 0.564, the MSE value as 0.319 and the MAE value as 0.143. The comparison of the actual takeoff speed values and the takeoff speed values estimated with the new XGBoost model is given in Fig. 10.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

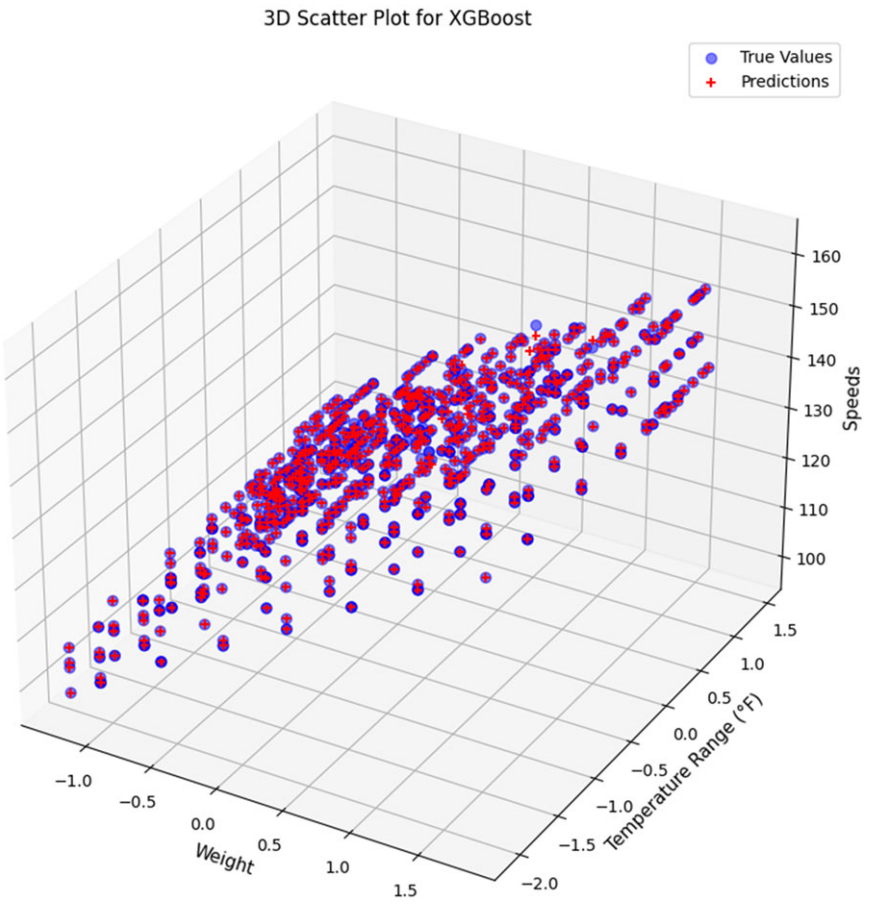


Figure 8. *Comparison of actual takeoff speed values with predicted takeoff speed values for XGBoost.*

When Fig. 10 is examined, it is seen that the XGBoost algorithm shows better performance in takeoff speed prediction than previous models with the determined hyperparameters.

5.0 Conclusions

The development of technology has led to significant developments in the aviation industry, as in many other areas. For different missions, aircraft with different features and sizes are designed. In these designs, the performance and safety of the aircraft are the main issues. Although there are many factors that affect aircraft safety, one of these factors is the takeoff and landing phases of the aircraft. The speed changes of the aircraft during takeoff and landing are important for the safety of the aircraft and are an issue that must be controlled.

In this study, takeoff speed prediction using flight takeoff data belonging to Boeing company is presented for the first time in the literature by using different machine learning algorithms and comparing them with each other. It is important to consider the parameters that are not directly related to each other simultaneously and with equal importance in takeoff speed prediction. In this study, LR, SVR, CART, RF and XGBoost algorithms were selected from machine learning algorithms for takeoff speed prediction using variables that are not directly related to each other. Takeoff speed prediction was made with these machine learning algorithms and presented comparatively. Then, feature engineered method was applied to reduce the error values of the algorithms and takeoff speed was obtained with lower error

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

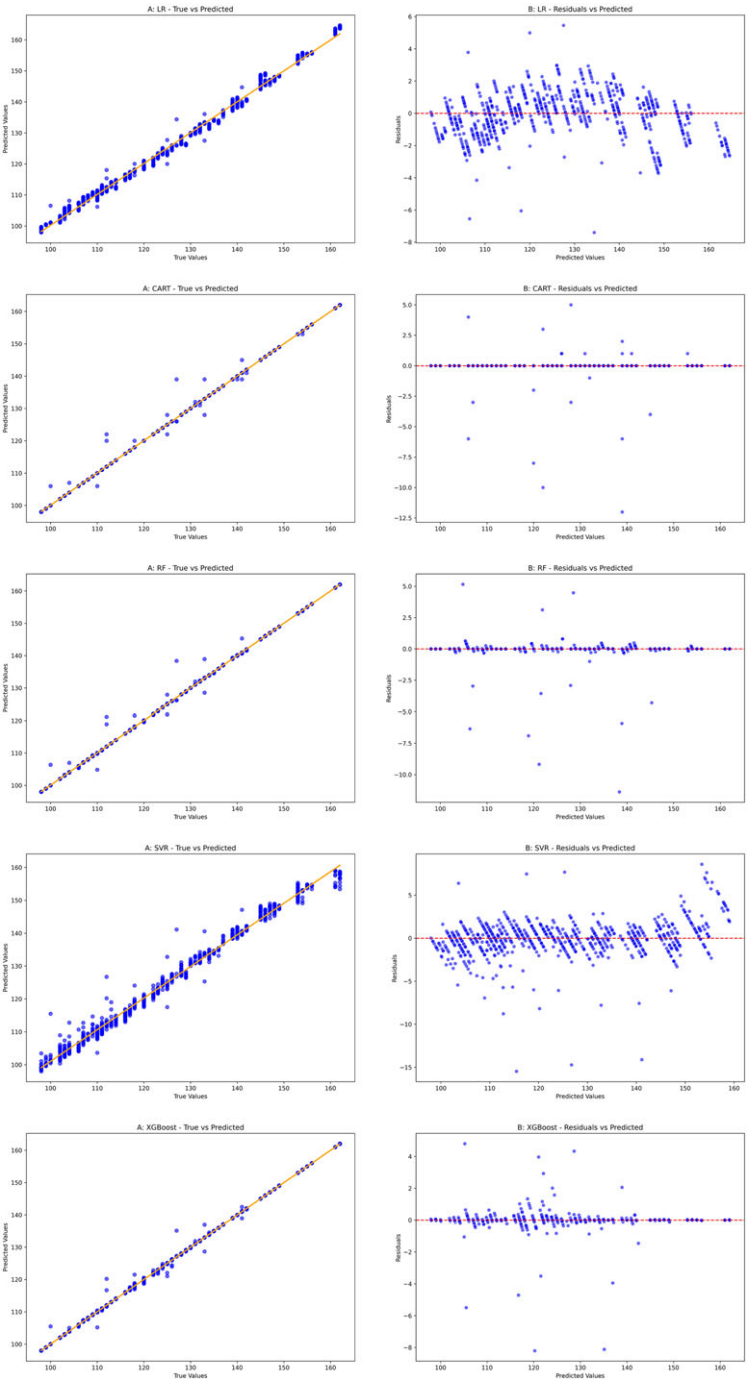


Figure 9. *First column A: Comparison of actual takeoff speed values with predicted takeoff speed values, second column B: Comparison of the residuals versus predicted values, for LR, CART, RF, SVR and XGBoost algorithms.*

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

Table 3. *The value ranges selected for hyperparameter optimisation*

Hyperparameters	Range values
learning_rate	0.01–0.5
max_depth	3–12
n_estimators	700–1,100
colsample_bytree	0.7–1.1

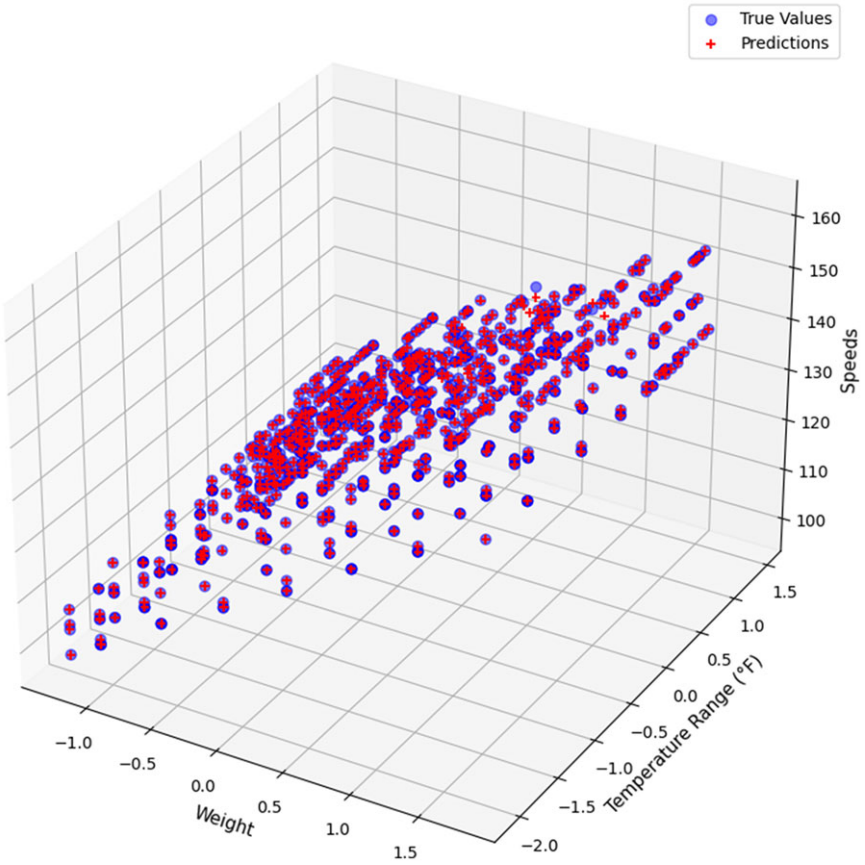


Figure 10. *Comparison of actual takeoff speed values with predicted takeoff speed values for XGBoost with hyperparameter optimisation.*

values with the developed new models. When the simulation results were compared, the compatibility of the estimated values obtained with the developed models with the real takeoff speed values was satisfactory. The proposing of base models for takeoff speed estimation, then the reduction of error values by applying feature engineered to these models and the comparison of the results are presented for the first time in the literature. Simulation results using models developed using variables that have no direct relationship with each other for takeoff speed prediction emphasise that machine learning algorithms can be used as an effective and alternative method to assist pilots in takeoff speed estimation.

In next studies, new models can be developed using different data sets and different artificial intelligence techniques. With appropriate integration systems, the proposed models can be integrated into aircraft for use in real-time flight.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

References

[1] Dhaliwal, S.S. Control Of An Unconventional Double-Ducted Fan VTOL UAV For Obstructed Environments Using The MARC System. University of Calgary, Yüksek Lisans Tezi, 168, 2009, pp 2009.

[2] Puri, A. Survey Of Umanned Aerial Vehicles (UAV) For Traffic Surveillance. Internal Report, University of South Florida, 2005.

[3] Watts, A.C., Perry J.H., Smith, S.E., Burgess, M.A., Wilkinson, B.E., Szantoi, Z., Ifju, P.G. and Percival, H.F. Small unmanned aircraft systems for low-altitude aerial surveys, *J. Wildlife Manag.*, 2010, **74,** (7), pp 1614–1619.

[4] Zongjian, L. UAV for mapping-low altitude photogrammetric survey, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII(Part B1), 2008.

[5] Min, A.T.W., Sagarna, R., Gupta, A., Ong, Y.S. and Goh, C.K. Knowledge transfer through machine learning in aircraft design, *IEEE Comput. Intell. Mag.*, 2017, **12,** (4), pp 48–60.

[6] Daniel, K. and Wietfeld, C. Using public network infrastructures for UAV remote sensing in civilian security operations, *Homeland Secur. Affairs Suppl.*, 2011, 3.

[7] Martinez, D., Fernández, A., Hernández, P., Cristóbal, S., Schwaiger, F., Nunez, J.M. and Ruiz, J.M. Forecasting unstable approaches with boosting frameworks and lstm networks, 9th SESAR Innovation Days, 2019.

[8] Mattingly, J.D., Heiser, W.H. and Pratt, D.T. *Aircraft Engine Design*, American Institute of Aeronautics and Astronautics, 2002.

[9] Kuznetsov, A., Shevchenko, A. and Solonnikov, J. The methods of forecasting some events during the aircraft takeoff and landing, *IFAC Proc. Vol.*, 2013, **46**, (19), pp 183–187.

[10] Puranik, T.G., Rodriguez, N. and Mavris, D.N. Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning, *Transp. Res. Part C Emerging Technol.*, 2020, **120**, p 102819.

[11] Diallo, O.N. A predictive aircraft landing speed model using neural network, 2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC), IEEE, 2012, pp 3D2-1.

[12] Hu, C., Zhou, S.H., Xie, Y. and Chang, W.B. The study on hard landing prediction model with optimized parameter SVM method, 2016 3, 5th Chinese Control Conference (CCC), IEEE, 2016, pp 4283–4287.

[13] Tong, C., Yin, X., Wang, S. and Zheng, Z. A novel deep learning method for aircraft landing speed prediction based on cloud-based sensor data, *Future Gener. Comput. Syst.*, 2018, **88**, pp 552–558.

[14] Zhang, H. and Zhu, T. Aircraft hard landing prediction using LSTM neural network, in Proceedings of the 2nd International Symposium on Computer Science and Intelligent Control, 2018, pp. 1–5.

[15] Borup, K.T., Fossen, T.I. and Johansen, T.A. A machine learning approach for estimating air data parameters of small fixed-wing UAVs using distributed pressure sensors, *IEEE Trans. Aerospace Electron. Syst.*, 2019, **56**, (3), pp 2157–2173.

[16] Kovarik, S., Doherty, L., Korah, K., Mulligan, B.,Rasool, G., Mehta, Y., Bhavsar, P. and Paglione, M. Comparative analysis of machine learning and statistical methods for aircraft phase of flight prediction, International Conference on Research in Air Transportation 2020, 9th International Conference, 2020.

[17] Jarry, G., Delahaye, D. and Feron, E. Approach and landing aircraft on-board parameters estimation with LSTM networks, 2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT), IEEE, 2020, pp 1–6.

[18] Kang, Z., Shang, J., Feng, Y., Zheng, L., Liu, D., Qiang, B. and Wei, R. A deep sequence-to-sequence method for aircraft landing speed prediction based on QAR data, Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20–24, 2020, Proceedings, Part II 21, Springer International Publishing, 2020, pp. 516–530.

[19] Ozkat, E.C., Bektas, O., Nielsen, M.J. and la Cour-Harbo, A. A data-driven predictive maintenance model to estimate RUL in a multi-rotor UAS, *Int. J. Micro Air Veh.*, 2023, 15. https://doi.org/10.1177/17568293221150171

[20] Ozkat, E.C. Vibration data-driven anomaly detection in UAVs: A deep learning approach, *Eng. Sci. Technol. Int. J.*, 2024, **54**, p 101702.

[21] Karaburun, N.N., Hatipoğlu, S.A. and Konar, M. SOC estimation of Li-Po battery using machine learning and deep learning methods, *J. Aviat.*, 2024, **8,** (1), pp 26–31.

[22] Boeing 737 Operations Manual, 737-300 AOM, Meridian Airlines Academy, Dade Collier Training and Transition Airport Florida, Boeing Document D6-7620-3lO (MA), 2000.

[23] Mitchell, T.M. *Machine Learning*, McGraw Hill, 1997.

[24] Kotsiantis, S.B., Zaharakis, I. and Pintelas, P. Supervised machine learning: A review of classification techniques, *Emerging Artif. Intell. Appl. Comput. Eng.*, 2007, **160**, (1), pp 3–24.

[25] Celebi, M.E. and Aydin, K. (Eds). *Unsupervised Learning Algorithms*, Springer, 2016, Cham.

[26] Sutton, R.S. Introduction: The challenge of reinforcement learning, in *Reinforcement Learning*, Springer US, 1992, Boston, MA.

[27] Alzubi, J., Nayyar, A. and Kumar, A. Machine learning from theory to algorithms: an overview, *J. Phys. Conf. Ser.*, 2018, **1142**, p 012012.

[28] Grossberg, S. Nonlinear neural networks: Principles, mechanisms, and architectures, *Neural Networks*, 1988, **1,** (1), pp 17–61.

[29] Han, J., Pei, J. and Tong, H. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2022.

[30] Kuhn, M., Johnson, K., Kuhn, M. and Johnson, K. An introduction to feature selection, in *Applied Predictive Modeling*, pp 487–519.

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press

[31] Farhat, R., Mourali, Y., Jemni, M. and Ezzedine, H. An overview of machine learning technologies and their use in e-learning, 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies"(OCTA), IEEE, 2020, pp 1–4.
[32] Liang, H. and Song, W. Improved estimation in multiple linear regression models with measurement error and general constraint, *J. Multivar. Anal.*, 2009, **100**, (4), pp 726–741.
[33] Su, X., Yan, X. and Tsai, C.L. Linear regression, *Wiley Interdiscip. Rev. Comput. Stat.*, 2012, **4,** (3), pp 275–294.
[34] Fitzmaurice, G.M. Regression, *Diagn. Histopathol.*, 2016, **22,** (7), pp 271–278.
[35] Aghelpour, P., Mohammadi, B. and Biazar, S.M. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA, *Theoret. Appl. Climatol.*, 2019, **138**, (3), pp 1471–1480.
[36] Cortes, C. and Vapnik, V. Support-vector networks, *Mach. Learn.*, 1995, **20**, pp 273–297.
[37] Awad, M., Khanna, R., Awad, M. and Khanna, R. Support vector regression, in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and system Designers*, 2015, pp 67–80.
[38] Deka, P.C. Support vector machine applications in the field of hydrology: a review, *Appl. Soft Comput.*, 2014, **19**, pp 372–386.
[39] Cortez, P. and Morais, A.D.J.R. A data mining approach to predict forest fires using meteorological data, 2007.
[40] Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. **2**, Springer, 2009, New York, pp 1–758.
[41] Harrington, P. *Machine Learning in Action*, Simon and Schuster, 2012.
[42] Breiman, L. *Classification and Regression Trees*, Routledge, 2017.
[43] Bharti, J.P., Mishra, P., Moorthy, U., Sathishkumar, V.E., Cho, Y. and Samui, P. Slope stability analysis using Rf, gbm, cart, bt and xgboost, Geotech. Geol. Eng., 2021, 39, pp 3741–3752.
[44] Breiman, L. Bagging predictors, *Mach. Learn.*, 1996, **24**, pp 123–140.
[45] Breiman, L. Random forests, *Machine Learn.*, 2001, **45**, pp 5–32.
[46] Cutler, A., Cutler, D.R. and Stevens, J.R. Random forests, in *Ensemble Machine Learning: Methods and Applications*, 2012, pp 157–175.
[47] Cox, D.R. Regression models and life-tables, *J. R. Stat. Soc. Ser. B (Methodol.)*, 1972, **34**, (2), pp 187–202.
[48] Quinlan, J.R. *C4. 5: Programs for Machine Learning*, Elsevier, 2014.
[49] Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp 785–794.
[50] Li, J., An, X., Li, Q., Wang, C., Yu, H., Zhou, X. and Geng, Y.A. Application of XGBoost algorithm in the optimization of pollutant concentration, *Atmos. Res.*, 2022, **276**, p 106238.
[51] Chicco, D., Warrens, M.J. and Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Comput. Sci.*, 2021, **7**, p e623.

Cite this article: Karaburun N.N., Arık Hatipoğlu S. and Konar M. (2025). Aircraft takeoff speed prediction with machine learning: parameter analysis and model development. *The Aeronautical Journal*, **129**, 1534–1549. https://doi.org/10.1017/aer.2024.164

https://doi.org/10.1017/aer.2024.164 Published online by Cambridge University Press