



Parametric estimation scheme for aircraft fuel consumption using machine learning

Mirza Anas Wahid¹ · Syed Hashim Raza Bukhari²  · Muazzam Maqsood³ · Farhan Aadil³ · Muhammad Ismail Khan^{4,5} · Saeed Ehsan Awan⁵

Received: 19 July 2021 / Accepted: 15 August 2023 / Published online: 16 September 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

The most efficient technique that is used for aircraft engine tuning is through mounting the engine on the engine test bench (ETB) to analyze, tune and monitor its variables through the ETB run. It is practically very difficult to unmount the engine from the aircraft and mount it on the ETB for analyzing and estimating a single variable such as fuel consumption or oil temperature as the unmounting process requires huge manpower and machinery. This problem can be resolved if the fuel consumption of an air vehicle is estimated without unmounting the engine from the aircraft through applying data analytics and machine learning models. Therefore, in this paper, the fuel consumption of an aircraft is analyzed and estimated through advanced data science techniques. The dataset went through data analyzing and preprocessing techniques before applying multiple machine learning models such as multiple linear regression (MLR), support vector regression, decision tree regression and deep learning algorithm RNN/LSTM. The performance of algorithms has been evaluated using model evaluation methods such as mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination. The models are evaluated in taxi, cruise and approach flight phases where the LSTM performs excellent among all other algorithms with RMSE 15.1%, 10.5% and 0.9%, respectively.

Keywords Machine learning · Engine test bench · Fuel consumption · Data analytic

1 Introduction

The aviation industry works as a catalyst for nation economy and social growth [1]. Both operations involving the transportation of goods and persons by air are part of this industry [2]. The report considers that the consumption of fuel is a big problem for the aviation industry. Due to high energy prices and increased competition, the use of fuel is quickly becoming a key element in the aviation sector [3]. One of the main parameters of direct operational costs in the aviation industry is the consumption of fuel [4]. Fuel for air transportation remains the most important and volatile operational cost factor, which for the air transport industry is an ongoing problem [5]. Airbus estimates that the fuel for the average family operator A320 in 2003 accounted for about 28% of gross running costs. In the near term, though, it will amount to over 45% of all aircraft operational costs [6]. A country's economy relies primarily on fuel prices increases in fuel usage have a diverse effect on airlines in which one is a significant influence on

✉ Syed Hashim Raza Bukhari
hashim.bukhari@mail.au.edu.pk

Mirza Anas Wahid
mirza-anas.wahid.1@ens.etsmtl.ca

¹ Department of Software and IT Engineering, École de technologie supérieure, Montreal, Canada

² Department of Electrical and Computer Engineering, Air University, Islamabad, Pakistan

³ Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan

⁴ Department of Electrical Engineering, City University of Hong Kong, Kowloon, Hong Kong

⁵ Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan

operational expenses and the other is a reduction in demand for passenger and air freight transport [7]. The number of flights has increased by 80% since 1990 and is projected to increase further by 45% by 2035. The European traffic forecast estimates that 12.8 million flights will be operated annually [8].

Fuel consumption represents about 20% of the aircraft systems' gross running costs. Monitoring the fuel needed for a particular flight service will improve the effectiveness of the operations of the aircraft, thereby saving aircraft management costs [9]. The aircraft performance parameters monitoring shall be used to calculate consumption and assess aircraft aerodynamic features such as drag and fuel consumption metrics [10]. Therefore, the advantages of fuel consumption retrofits can be quantified before and after using a measure by analyzing aircraft performance parameters monitoring. Statistical analysis of historical datasets and physical estimate models was mainly used for analyzing flight parameters such as fuel consumption [11]. The Gauges of aircraft performance determined by specific operating conditions are compared to book values from the operating manual of the flight team [12]. Figure 1 shows the fuel consumption in billions of gallons each year. From 2005 to 2021, trillion gallons of fuel consumed in the air business. The fuel consumption has a reduction in the COVID pandemic, but the trend shows the alarming condition.

The aircraft engine is the most important and essential aircraft system, and its reliability depends upon flight and flight crew safety [14]. The aircraft engine operates in rough environmental conditions, and its performance output steadily worsens during its lifetime [15]. A reliable and efficient method of engine optimization and tuning reduces the cost of engine overhaul cost; furthermore, it benefits the engine safety and reduces life cycle costs. The traditional method of aircraft engine parameters analysis is the manufacturing of the engine test bench facility to optimize and analyze engine different features and increase the efficiency of the engine in the flight [16]. The engine test bench (ETB) creates a virtual environment for

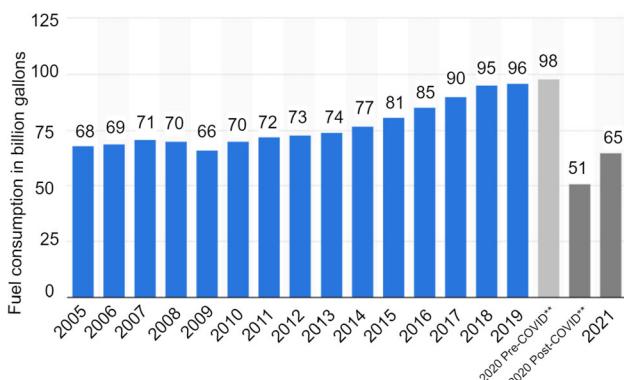


Fig. 1 Year-wise airlines fuel consumption [13]



Fig. 2 Engine test bench (ETB) of jet engine [22]

monitoring, analyzing and estimating engine parameters such as fuel consumption [17] as shown in Fig. 2.

With the help of an engine test bench facility, the efficiency of the engine increases and the rate of engine burnout accidents during flight reduces. The principal concern is the expense related to the ETB establishment, which generally adds up to millions of dollars. This is mainly a result of its complexity, involving numerous subsystems, assemblies and components. [18]. Mounting the engine on the engine test bench (ETB) and analyzing, tweaking and monitoring its variables throughout the ETB run is the most efficient way for aircraft engine tuning. Unmounting an engine from an airplane and mounting it on an ETB for the purpose of evaluating and estimating a single variable such as fuel consumption or oil temperature is almost impossible. The unmounting process necessitates a large amount of labor and machinery. This issue can be overcome by using data analytics and machine learning models to predict an air vehicle's fuel usage without removing the engine from the aircraft. In this regard, we use a data-driven approach to model fuel consumption in aircraft using flight data recorder (FDR) archives. The air vehicles have flight data recorder (FDR) which records complete mission profile and aircraft sensors data. The sensors are navigation, guidance, engine, fuel tanks, environment and payloads. The FDR data is used to analyze flight data, and it helps in analyzing aircraft systems, subsystems and components behavior [19]. The approach proposed in this study is to use FDR data instead of generating new ones and use FDR data to estimate engine fuel flow or fuel consumption by predicting it using advanced machine learning algorithms [20]. Therefore, in this study, machine learning algorithms such as multiple linear regression (MLR), decision tree regression (DTR), support vector regression and recurrent neural network/long short-term memory algorithms are used for the development of an aircraft fuel consumption model based on the FDR dataset. In this study, the focus is on supervised learning [21], because the dataset and problem statement are related to supervised learning. The dataset consists of independent variables and a dependent variable, which is fuel flow.

To the best of our knowledge, this is one of the few studies which is related to fuel consumption estimation with the application of machine learning. With the help of this research, the problem related to aircraft fuel consumption analysis and estimation without mounting engine on the ETB is resolved and it saves heavy machinery and labor need which requires for engine mounting and unmounting. Furthermore, to date, there is very little published research on the M/s Honeywell fuel flow dataset. The application and evaluation of machine learning and deep learning model for the prediction of fuel flow/fuel consumption of an aircraft in different phases have been done in this paper. The brief contributions of this paper are that we have organized and extracted the data from the flight data recorder of an aircraft and preprocessed the mission data with respect to multiple flight phases. Moreover, we have reduced the need for labor and heavy machinery by predicting fuel consumption using multiple ML algorithms. Furthermore, with the help of this study, fuel consumption analyzed and estimated without mounting engine on the ETB; therefore, the need of ETB may reduce.

The paper is structured into multiple sections which are described as follows: Section 2 covers the literature review, Sect. 3 presents the ML models' development and methodology, Sect. 4 presents ML models' results and discussion, and the paper concludes in Sect. 5 conclusion and future research.

2 Related work

Machine learning is the most emerging topic of the fourth industrial revolution (4IR). These studies are related to aircraft fuel consumption or engine fuel flow. The research focus is to predict aircraft fuel consumption based on an aircraft dataset. Many models for aircraft fuel estimation have been suggested in the current literature. They can be divided into three different categories: complex mathematical, data-driven and hybrid models. Mathematical models are built on complex math and have a greater physical interpretation of the structure. But, since they require expert experience, cannot take symbolic data into account and do not neglect nonlinearity, they can be difficult to enforce. In previous research, the fuel consumption of an aircraft was predicted using the energy-balance (EB) mathematical equation. The EB equation is a method that explains the changes in aircraft engine thrust and drag, which are continually balanced by fuel energy usage. Commercially, two mathematical models are used: One is simmod, which is built by the Federal Aviation Administration based on the advanced fuel consumption concept and is the most common EB model in [23, 24, 25], and

the other is base of aircraft data (BADA), which was developed by the European Organization for the Safety of Air Navigation [26]. In [27], the review on engine downsizing has been conducted and multiple aspects of researchers work on engine downsizing are compiled and categorized into different categories. In the following subsections, the literature regarding fuel consumption prediction or estimation of an aircraft using different machine learning techniques has been discussed. The related work is divided into data-driven and statistical and hybrid methods.

2.1 Data driven approach

The article referenced as [28] discusses a challenge in unsupervised machine learning. The focus of this research is on a dataset from a passenger airplane, including information gathered from air data units, engines and control systems. To examine the data, the researchers categorized it into three phases: climb, cruise and descent. They developed a fuel consumption model using the decision tree regression algorithm for each of these phases of flight. The study also looked at the implementation of artificial neural networks (ANNs), a more intricate deep learning algorithm that could increase accuracy. The decision tree model performed well, but ANNs showed higher precision, although at the cost of increased computational power and time. Decision trees, however, are quicker and use fewer computational resources. These models have many possible applications within the aviation industry, depending on the specific requirements. Decision trees can be used for fast estimation and prediction, while ANNs can be employed to design high-precision models where accuracy is critical. This research examines the implementation of machine learning algorithms to model fuel use in various phases of flight and for complete flight assignments dependent on data. Two nonparametric techniques, namely neural and decision-making networks, are applied to and contrasted with the publicly accessible flight activity data. Neural networks show the best performance for fuel profile modeling over an entire flight mission but have a strong computational commitment.

In [29], the authors discuss the use of flight operational data to estimate fuel consumption in piston-engine aircraft. The study uses statistical modeling techniques, including classification and regression tree (CART) and neural network (NN) approaches, to analyze the data and compare the performance of the two methods. The paper also discusses the impact of ADS-B data errors on vertical speed and the need for further investigation. The results show that both CART and NN methods are adaptive for nonlinear relationships among variables and generate nonparametric

models for selected variables. The CART method performs better than the NN model in fuel consumption estimation.

An airliner aircraft is focused, and its fuel consumption is predicted based on the flight dataset, passenger dataset and reservation dataset by the authors in [30]. The study has used a huge dataset, i.e., 54,000 flights and 9,900,000 passenger data used for predicting fuel consumption and calculating total flight delays. The flight dataset consists of the departure year, month, day of the week, day of the month, day of the year, scheduled departure time, scheduled arrival time, the airport of departure, the airport of arrival, domestic or international flight, aircraft ID, position ID, fuel on board and route time. In this paper, the research has been carried out on the feasible application of cost-effective airline companies' predictive modeling methods to save costs and increase service efficiency. In order to reduce costs and flight delays and improve services, the authors anticipated the amount of fuel consumed. Since airlines may take various actions according to the prediction timetable, the author has found three timetables for the prediction: the day before departure, the condition of nearly all bookings being taken was reflected; the week before departure, the state in which reservations were still not collected; and the state in which no bookings were made five months before departure. For predicting fuel consumption, three algorithms are used: random forest regression, XGBoost and deep neural networks. It is a supervised problem, and random forest regression performs well as compared to other machine learning and deep learning algorithms. The root mean absolute error (RMSE) method is used for evaluation metrics. The performance evaluation, algorithms and parameters used in discussed literature are given in Table 1 Literature Review.

A simpler model to estimate aviation fuel consumption using a support vector algorithm is presented in this [31] article. The technique developed here can be used for rapid simulation in the aviation industry. The dataset used in this study for the prediction of aircraft fuel consumption is from the route data and the aircraft performance manual. The performance of the trained models is evaluated using receiver operating characteristics (ROC) curve. This approach can be applied with certainty to any aircraft type, including piston and turboprop. The data used in this analysis was specific to CFM56 engine-driven Boeing 737–800 aircraft. The predicted fuel consumption values are compared with the actual values that are provided in the flight manual and were considered accurate for rapid model implementation. In this study, flight data is not divided into different flying phases; full flight data is used for developing the SVM model. The findings of this study demonstrate that a ROC optimization support vector model can efficiently represent complex fuel consumption features for the entire flight phase.

2.2 Statistical approach

The authors in [32] also relate the performance to engine fuel flow rate prediction of airliner aircraft, i.e., the A320-214. If the engine is manual and not electronically controlled, then the engine fuel flow rate parameter is also known as aircraft fuel consumption, but if the engine is EFII modified, then the engine fuel flow rate is not aircraft fuel consumption. In this study, the dataset consists of 10 airline aircraft's flight parameters. The flight dataset is divided into different flight phases, i.e., ascent, climb out, cruise, descent and landing. The problem statement is a supervised problem, and the dataset is divided into three sets: a training set, a validation set and a test set. The CART (classification and regression trees) and least squares boosting (LSB) algorithms are used for engine fuel flow prediction through regression modeling. The model results are compared with airliner aircraft models, i.e., BADA and ICAO Databank. The mean error (ME) and percentage coverage (PC) methods are used for model performance evaluation. This paper examined how machine learning methods are used to predict the fuel flows of aircraft engines. The CART and LSB nonparametric methods were employed. The collection of regressive variables is aided by physical knowledge of flight principles. Actual FDR data from 10 different aircraft models were evaluated using the regression techniques. At the cost of increased complexity and some interpretability, CART provided easy and interpretable fuel flow models, while LSB models found themselves to have higher predictive ability than CART models. Both approaches worked better in the fuel flow estimation process than the ICAO database and the BADA process. The models also provide estimates of uncertainty concerning fuel flow rates, which measure the magnitude of the variability in actual flight activities. A US-based airline dataset fuel consumption prediction is performed for reducing the airliner's fuel consumption [33]. The dataset consists of airline fuel data, weather data and traffic data. The machine learning methods which are used for fuel prediction are quantile regression, gradient boosting and random quantile forests. It is a supervised learning problem in which the dataset is divided into three sets, i.e., training set, validation set and test set. The training set is 60% of the total data, while the validation set is 20% of the total data and the test set is also 20% of the total data. The model is assessed through the goodness-of-fit measure method. The random quantile forest value after the goodness-of-fit method is higher among all, and the quantile regression method value is lowest among all.

Table 1 Literature summary of fuel consumption analysis of aircrafts

S. no.	Ref.	Algorithms	Dataset	Performance evaluation
1	[28]	Decision tree regression and ANN	Air data unit, engine and control inputs	ANN-MRE = 0.8%, $r^2 = 0.9995$
2	[32]	CART and LSB	Flight trajectory, velocity, acceleration, time and altitude	CART Results Ascent: ME = 1.4, cruise: ME = 2.8 Descent: ME = 12.1, approach: ME = 13.7 LSB result Ascent: ME = 0.7, cruise: ME = 2.0 Descent: ME = 6.8, approach: ME = 6.5
3	[30]	Random forests regression, XGBoost and deep neural networks	Flight time, date, schedule and fuel onboard	Random Forest-RMSE = 0.100 XGBoost-RMSE = 0.094 Deep neural networks-RMSE = 0.120
4	[35]	Hierarchical density-based spatial clustering of applications with noise (HDBSCAN) and principal component analysis (PCA)	Flight operational data	RMSE less than 5 % Aggregate errors less than 2%
5	[33]	Quantile regression (QR), gradient boosting machine (GBM) and random quantile forests (RQM)	Past traffic, air vehicle types, takeoff hour, time and weather	Goodness-of-fit measure RQF = 0.250 GBM = 0.237 QR = 0.231
6	[34]	CART and SS-ANOVA	Flight altitude, ground, speed and the vertical speed dataset	CART-MAE of 1.244 SS-ANOVA -MAE of 1.086
7	[31]	Support vector regression (SVR) Radial basis function network	Data from aircraft performance	AUC value of ROC_SVM is 0.89; 0.83 of ROC_RBFN
8	[36]	Deep neural network with physics-based optimization Linear regression (LR), support vector regression neural network	Aircraft aerodynamics, flight navigation sensors and guidance data	DNN-MAE: 127, MAPE: 1.521 NN-MAE: 172, MAPE: 1.897 LR-MAE: 201, MAPE: 2.695 SVR-MAE: 195, MAPE: 2.677
9	[37]	Feed-forward back-propagation (NN) and Levenberg–Marquardt algorithms with the genetic algorithm optimization	Real flight dataset flight altitude, airspeed and fuel flow	Climb: BP; R = 0.98, LM; R = 0.984 Cruise: BP; R = 0.954, LM; R = 0.948 Descent: BP; R = 0.954, LM; R = 0.99097
10	[41]	Generalized linear rules model (GLR) and prediction rule ensemble (PRE)	This research is on the same dataset that we are using in this research [42] but applied on 1000 flight instances and takeoff phase only	Takeoff GLR: R2 = 80.59 RMSE: 342.23 MAE: 280.43 Prediction rule ensemble 94.88 RMSE: 252.26 MAE: 194.03
11	[29]	Classification and regression tree (CART) and neural network (NN)	Operational data collected from FDR of cessna 172 aircraft	MAPE Using CART average MAPE = 6.30 Using NN average MAPE = 14.3
12	Proposed scheme	MLR, SVR, DTR and RNN/LSTM		MAE, RMSE and r^2

2.3 Hybrid approach

In [34], the fixed-wing aircraft, i.e., the Cirrus SR-20 piston-engine aircraft fuel flow rate, is predicted. Cirrus SR-20 fleet operating flight data from Purdue University was used for training and testing of the CART algorithm. This study analyzes two nonparametric models in each phase of the landing and takeoff (LTO) process to predict the fuel flow rate for the GA piston-engine aircraft. With the implementation of ML algorithms. In this study, the dataset used for machine learning modeling consists of flight operational data, including the aircraft altitude, ground speed and vertical speed. The CART algorithm seems to be stable for aircraft mission data. The CART model intuitively explains the fuel flow rate in in-flight phases and shows the clear logic of fuel flow prediction. Moreover, the ANOVA smoothing spline was designed to remodel the fuel flow rate and examine the relationship between variables further. The two models demonstrate high fuel flow accuracy and show that the aircraft's altitude has less impact on the fuel flow in this analysis for piston-engine aircraft. The findings of the study are considered rational from the point of view of current flight operations. Furthermore, for the approach phase, it is notable that the fuel flow rate predicted in the CART model differs significantly from the derived value dependent on the general aviation landing and takeoff (LTO) process. The possibilities are flight data is provided by flight training aircraft operated by pilot students; inexperience in approach operations could produce a lot of noisy data; and errors and inaccuracies in the CART model could occur in the way the average fuel flow rate was derived from initial flight data.

In [35], a two-step approach is used to estimate fuel consumption for new flight sectors, which involves unsupervised learning and a regression model. The unsupervised learning step uses hierarchical density-based spatial clustering of applications with noise (HDBSCAN) to cluster the principal component analysis (PCA)-reduced data. The regression model is derived for each cluster with fuel burn as the output. The proposed approach is demonstrated using actual operational data shared by an airline partner. The results show that the proposed approach is effective and efficient for fuel budgeting purposes, including new sectors.

In this article [36], the aircraft fuel consumption is predicted with the help of a physics-based deep learning model. In this study, the main algorithm is a deep learning algorithm with multiple hidden layers aided by physics-based optimization mathematical algorithms, or we can say that in this study, a hybrid deep learning model was developed for the prediction of aircraft fuel consumption. It is a supervised learning problem, and the dataset used in

this article consists of aircraft aerodynamic features such as body axis states, environmental states, performance states and dynamic states. The dataset was extracted from the quick access recorder (QAR) of a major European flag carrier airline. With the hybrid deep learning algorithm, other algorithms are also developed, such as linear regression (LR), support vector regression (SVR) and neural networks (NNs). All the model's results are evaluated using mean absolute error (MAE) and mean absolute percentage error (MAPE). After the development of a deep neural network, i.e., a hybrid model, LR, SVR, NN and BADA, the results are evaluated using MAE and MAPE and compared with each other. It is concluded that the deep neural network, i.e., the hybrid model, has the minimum MAE and MAPE values. In this study [37], a hybrid approach is followed for the prediction of aircraft fuel consumption. The neural network (NN) architecture is developed using a genetic algorithm. The feed-forward back-propagation and Levenberg–Marquardt algorithms are implemented for the prediction of fuel consumption, and a genetic algorithm is developed to optimize the overall model in terms of time and accuracy. In this article, the Levenberg–Marquardt algorithm performs excellently in the descent phase, while back-propagation results are good in the other two flight phases, which are climb and cruise. The variation in airspeed and altitude is also incorporated into the model. The model is evaluated using the mean percentage error. The real flight data used in this study for the aircraft fuel consumption prediction is rather than flight manual datasets. In [38, 39] and [40], base of aircraft data (BADA) model is used for the prediction of aircraft fuel consumption, but the BADA model loses its accuracy in the terminal flight phases, whereas it is considered to be more accurate in the cruise flight phase. Moreover, in this article [41] the same dataset [42] was used, but in this research, the generalized linear rules model and prediction rule ensemble methods are used for the estimation of fuel flow rate on the takeoff flight phase for 1000 flight instances. Similarly, in [43] the same dataset is used for visualizing the industrial big data of the flight data recorder to discover the factors affecting the aircraft fuel efficiency, but no estimation method was applied.

3 Research methods

The research methodology related to ML algorithms implementation for fuel flow prediction is discussed in this section as shown in Fig. 3. The methodology section covers the data preprocessing steps which are followed for fuel consumption research. Furthermore, the ML algorithms that are used in this study are discussed.

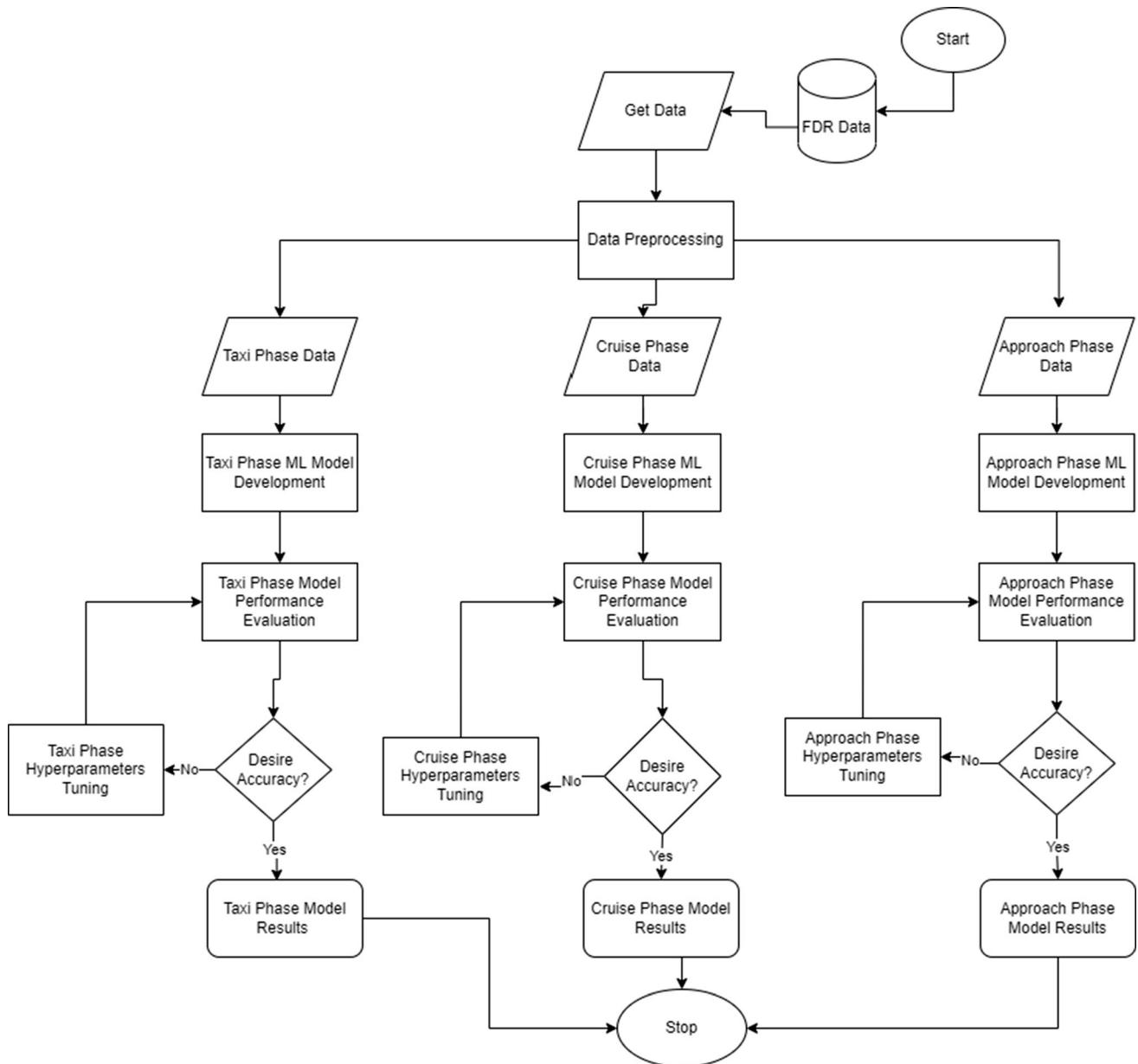


Fig. 3 The research methodology

3.1 Data preprocessing

In this paper, the algorithm development process consists of data preprocessing before model training. The data preprocessing consists of sub-processes that are data extraction and integration, flight phases separation, features selection, data description and visualization, data splitting and features scaling. The data preprocessing method is the most time-consuming process in the whole research. The dataset has gone through different data preprocessing steps and techniques, as shown in Fig. 4. The dataset

downloaded from the online source consists of several different CSV files of flights. In this research, a dataset of 200 flights is used. As all the flight's data features are on a separate CSV file, all the flight CSV files that are extracted from the online source must be integrated or merged. After integration, the dataset contains 33 flight tail numbers where each flight tail is a different aircraft and has 25 flying instances. In many data projects, data preparation is time-consuming, and it could be repeated many times over the course of analysis as new problems and ideas come to light. The provided data comes in 831 CSV files and a relatively large volume of 10.2 GB, including readings of numerous sensors on aircraft—detailed dynamics, system

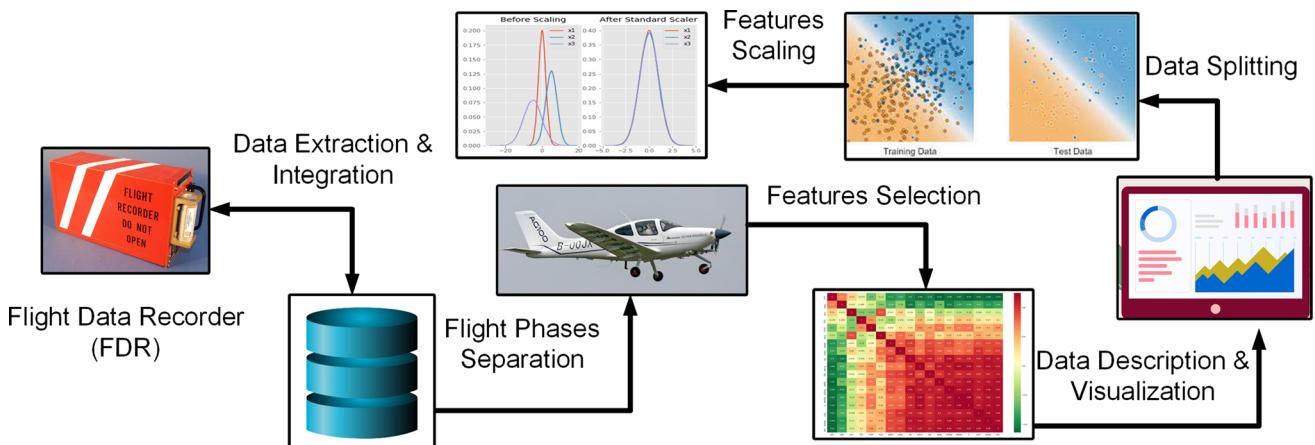


Fig. 4 Data preprocessing sequence flow

performance and other engineering parameters, such as altitude, speed, acceleration and position measurements. The data is parsed and populated into an open-source database. The processed dataset had a dimension of more than 200 columns and over 5 million rows. With efficient Python packages and an optimized workflow, the data preprocessing is carried out on a laptop computer. All the flight data was extracted from the online CrowdANALYTIX database and integrated the 200 flight data for the sake of this research.

In the data extraction and integration step, the data of 200 flights are in combined form. The combined data comprises 200 different flight records stored in a single.csv file. Therefore, it is needed to separate different flight phases from the file for this research. The feature “PH” value indicates the flight phases, 0 is equal to “Unknown” phase, 1 is equal to “Preflight,” 2 is equal to “Taxi,” 3 is equal to “Takeoff,” 4 is equal to Climb, 5 is equal to Cruise, 6 is equal to Approach and 7 is equal to Rollout. It is necessary to separate different phases in the early stage of data preprocessing to reduce algorithm complexity and computation power. The dataset is divided into different flight phases, i.e., preflight taxi, ascent, cruise, descent and post-flight taxi. All the phases are divided concerning GPS altitude measurements.

The taxi flight phase is when the aircraft moves on the ground before takeoff or after landing. It involves navigating from the gate to the runway or from the runway to the parking area. Cruise is the level portion of aircraft travel where flight is the most fuel-efficient. It occurs between the ascent and descent phases and is usually the majority of a journey. Technically, cruising consists of heading (direction of flight) changes only at a constant airspeed and altitude. It ends as the aircraft approaches the destination where the descent phase of the flight commences in preparation for landing. For most commercial passenger aircraft, the cruise phase of the flight consumes

the majority of fuel. As this lightens the aircraft considerably, higher altitudes are more efficient for additional fuel economy. The approach phase of flight starts when an airworthy aircraft under the direction of the flight crew descends below 5,000 feet AGL intending to approach and terminate when the aircraft reaches the approach end of the landing runway (the threshold of the runway) or at the start of a maneuver.

Feature selection is among the fundamental terms in machine learning that has an immense influence on the functioning of one ML model. The data features are used to train one machine learning model to have a significant impact on the performance. In this study, Pearson’s correlation was used for features selection as shown in equ 1. It is evidently the most frequently used measure for linear relations between two normal distributed variables, often referred to as the “correlation coefficient.” Generally, the Pearson coefficient is obtained using a lowest squares fit, and a value of 1 defines a positive perfect high relation, -1 defines a perfect negative relationship, and 0 indicates the absence of a relationship between the variables.

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad (1)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (2)$$

In this paper, the Pearson correlation method is used for feature selection as all the independent features have a linear relationship with the dependent feature. All the flight phases which are taxi, cruise and approach features are selected using the correlation method. As total features are 226, top 09 highly correlated features with dependent features, i.e., “Fuel Flow,” are used in this research to reduce the computational power and time.

In this research, the datasets for each flight are described utilizing the Python library `panda`. The `describe()` function from the `pandas` library quickly provides basic statistical details of the selected features. It describes some basic statistical details such as percentile, mean and standard deviation of a data frame or a series of numeric values. In this study, the `describe()` function from the `pandas` library is employed to generate a comprehensive description of all flight phases within the dataset. The dataset splitting process is used to split the dataset into two sets, i.e., a training set and a test set. The training set is used for training the machine learning model, which consists of 80% of the total data, and the test set usually consists of 20% of the total dataset, which is used to test the machine learning model. In this paper, the dataset is divided into different flight modes, i.e., ground mode and air mode. Therefore, the dataset splitting function will be performed for each flight phase independently. After splitting the dataset, the data transformation process is used for all flight phases. In this process, the scaling, normalization or standardization methods are implemented so that all the values are generalized. The dataset was first split into training and test sets, and then, we applied the transformation method to avoid original information loss.

The data is transformed after splitting the dataset into training and test sets to avoid information leakage in the test set. The main two feature scaling techniques indeed put all features on the same scale in which one of them is standardization which consists of subtracting each value of the feature by the mean of all values of a feature and then dividing by the standard deviation which is the square root of the variance, this will put all the values of the feature between around minus three and three. Standardization of datasets is a common requirement for many machine learning estimators implemented in `scikit-learn`; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with zero mean and unit variance. The x sample standard score is calculated in Eq. 3.

$$z = (x - u)/s \quad (3)$$

where u is the mean of the training samples or zero if `with_mean = False` and s is the standard deviation of the training samples or one if `with_std = False`. Another technique used for feature scaler is normalization. It consists of subtracting each value of the feature from the minimum value of the feature and then dividing by the difference between the maximum value of the feature and the minimum value of the feature; since this is positive, this is always larger, which means that all the values of features will become between 0 and 1. Therefore, by using standardization get values between minus three plus three more

or less and by using normalization get all the values of features between 0 and 1. The normalization is defined in this study, and the standard scaler is used for feature scaling. The normalization is defined in Eq. 4.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

Here, the question arises that one should go for standardization or normalization. Normalization is recommended when there is a normal distribution in most of the features. This will be a great feature scaling technique, while standardization works well all the time. Therefore, it is recommended to use standardization as it is less prone to outliers.

3.2 Machine learning models development

In this paper, four different algorithms are used for fuel consumption prediction: MLR, DTR, SVR and stacked LSTM. In subsections, the mathematical details of each algorithm are discussed.

3.2.1 Multiple linear regression model methodology

MLR machine learning technique uses several independent features to predict y dependent feature, i.e., fuel flow. The aim of MLR is to model a linear relationship between the flight dataset's independent variables and the fuel flow's dependent variable. In Eq. 5, the y is the "fuel flow" dependent variable and x is the independent variables that are independent features of different flight phases. The mathematical model of MLR is shown in Fig. 5.

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon \quad (5)$$

The β_0 and β_k weights are updated using `sklearn.linear_model` class. MLR error ϵ is calculated using least square method, and this method seeks to minimize the sum of the squared residuals. This means from the given data distance calculated from each data point to the regression line, square it, and the sum of all of the squared errors together. The same `sklearn` Python ML library is used for MLR model development. After initializing the linear model class, the next process is to train the training set. The training takes time depending upon the length of dataset instances. The training time increases as the number of instances grows. Similarly, the MLR technique has been used for each flight phase. After training the model, the next process is to test the model with the test set.

3.2.2 Support vector regression model methodology

In this paper, the SVM version for regression SVR is used for the prediction of fuel consumption. The SVM was

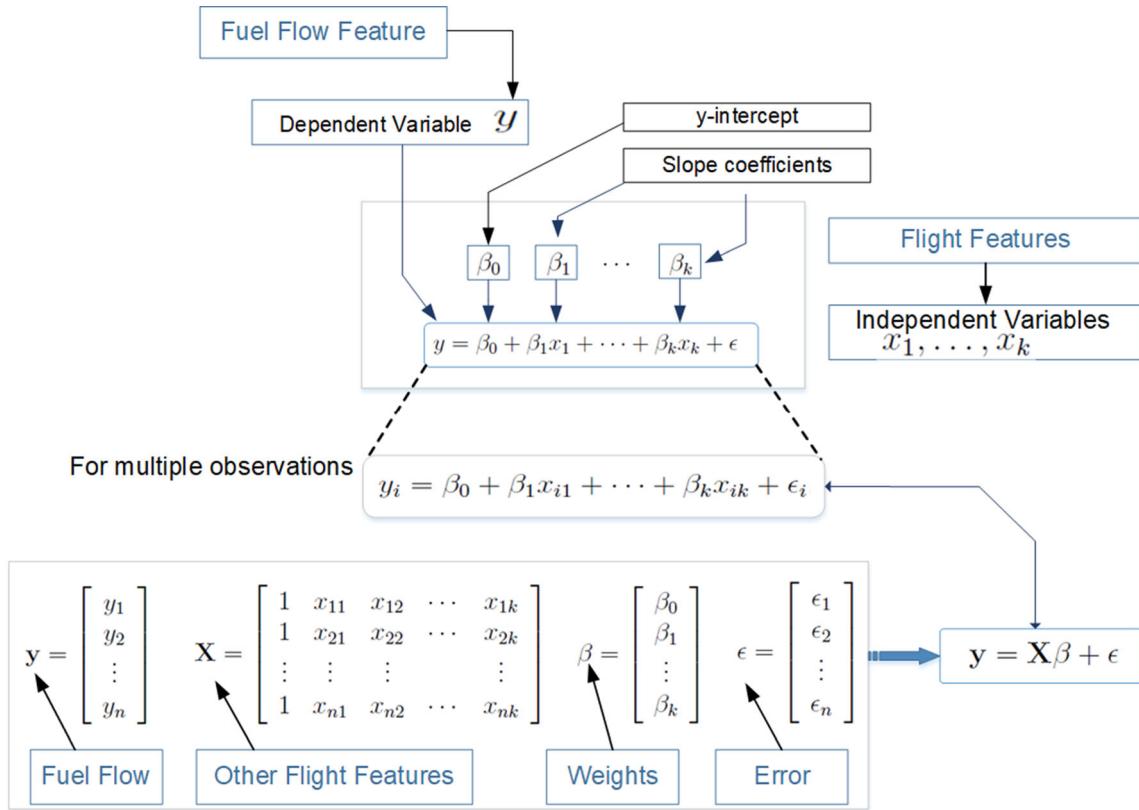


Fig. 5 MLR mathematical model

developed and introduced for controlling the structural risks [44]. Nowadays, SVM is used for COVID time-series, technical and power load prediction data. For time-series prediction of flight data recorder (FDR), Eq. (6) presents the FDR input at a given instance i , where y_i shows the output data with element N [45].

$$D = (X_i, y_i), 1 \leq i \leq N \quad (6)$$

Hence, the regression Eq. (7) is given as follows:

$$f(X_i) = W^T \phi(X_i) + b \quad (7)$$

Equation (7) represents the bias and weight as b and W correspondingly. The input vector X can be mapped by (X) into a higher-dimensional space. The value of W and b can be found using optimization problem as shown in Eqs. (8) and (9).

$$\text{Min} \frac{1}{2} \|W\|^2 + c \sum_{i=1}^N (\epsilon_i + \epsilon_i^*) \quad (8)$$

The same conditions apply:

$$\begin{aligned} y_i - W^T(\phi(x)) - b &\leq \xi + \epsilon_i \\ W^T(\phi(x)) + b - y_i &\leq \xi + \epsilon_i^* \\ \epsilon_i, \epsilon_i^* &\geq 0 \end{aligned} \quad (9)$$

It should reflect the trade-off between the simplicity of a model and the ability of generalization. The ξ_i and ξ_i^* are the slack variables, and they are used to measure error costs. As the nonlinear data can be mapped from the original vector space to higher-dimensional space using kernel mapping where linear regression model can be used, the linear model for SVR can be found using Eq. (10):

$$y_i = f(X_i) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(X_i, X_j) + b \quad (10)$$

The α_i and α_i^* are the lagrange variables multipliers. Gaussian radial function (RBF) is used for all flight phases. It is the most widely used kernel with a width of σ as given in Eq. (11).

$$K(X_i, X_j) = \exp(-\|X_i - X_j\|^2 / (2\sigma^2)) \quad (11)$$

After initializing the SVR class, the next process is of selecting or initializing the hyperparameter like kernel “C.” After hyperparameter initialization, the next process is to train the training set. The training takes time depending upon the length of dataset instances. In this research, the same SVR methodology is used for each flight phase.

3.2.3 Decision tree regression model methodology

In this paper, the decision tree regression (DTR) is selected because, unlike the traditional decision tree, it can predict the numeric outcomes of the dependent variable. In a regression problem, let $X = X_1, X_2, \dots, X_{pn}$ be independent variables and p_n is the total number of independent variables that are flight data independent features. Let n and fuel flow instances that are $Y = Y_1, Y_2, \dots, Y_n$ be the number of observations and a target variable that takes continuous values, respectively. vf is a feature variable and th is a threshold value. Let t and $\gamma = (vf, th)$ be a node and candidate split, respectively.

$$Q_l(\gamma) = (x, y) | x_{vf} \leq th_t \quad (12)$$

Equation (12) shows that Q_l that is the left side in the decision tree is found by splitting the data into γ candidate split.

$$Q_r(\gamma) = (x, y) | x_{vf} > th_t \quad (13)$$

Equation (13) shows that Q_r that is the right side in the decision tree is found by splitting the data into γ candidate split. Also, Eq. (13) can be declared as $Q_r(\gamma) = Q/Q_l(\gamma)$. Let n_l and n_r be the number of a sample left and right child (side), respectively.

$$I(Q, \gamma) = \frac{n_l}{n} S(Q_l(\gamma)) + \frac{n_r}{n} S(Q_r(\gamma)) \quad (14)$$

I is an impurity function, and Eq. (14) shows how the impurity function is calculated. The impurity function is minimized by considering Q and γ parameters. In this paper, all the above-discussed mathematical work is done by the sklearn ML library in which `sklearn.tree.DecisionTreeRegressor` class is used for DTR model development. Once the DTR class is initialized, the subsequent step involves training the DTR model, calculating its fitness and testing it on the fuel flow test dataset.

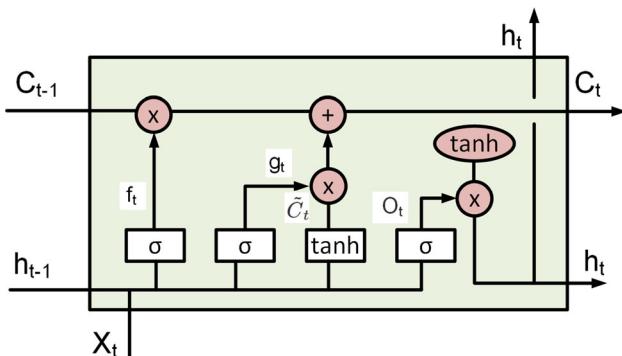


Fig. 6 Structure of LSTM cell

3.2.4 LSTM model methodology

In this paper, stacked long short-term memory (LSTM) is used that is followed by a linear regression to handle the aircraft fuel consumption prediction of different flight phases. LSTM is an advanced form of recurrent neural network (RNN) as it resolves the vanishing gradient problem. The basic LSTM cell structure is shown in Fig. 6.

The idea behind the LSTM algorithm is to transfer cell state from the previous time step to the current time step. With the addition of a gate structure, the simple linear transformation of the cell state can be defined. One approach is to use the sigmoid function whose outputs are 0 to 1, followed by a multiplication operation, which decides whether to remember or forget the previous knowledge as given in Eq. (15)

$$\sigma(x) = \frac{1}{1 + e^{-1}} \quad (15)$$

This mechanism is commonly known as the forget gate and used in LSTM for voice recognition, control theory of robotics and many other technical problems. The input contribution percentages should be measured in addition to the forget gate. A \tanh function is used in this measurement as shown in Eq. (16)

$$\tanh(x) = \frac{e^x - e^{-1}}{e^x + e^{-1}} \quad (16)$$

Since the \tanh value will drop from -1 to 1 , it is a good activation function to describe the active and negative contribution. In the LSTM cell structure, the \tanh layer is a contribution gate. In general, the inputs (tensors) are usually passed through the LSTM. With respect to Fig. 6, the forget gate f_t is defined in Eq. (17) as follows:

$$f_t = \sigma(W(h_{t-1}, X_t) + b_f) \quad (17)$$

where h_{t-1} and X_t first go through the processes of linear transformation, and then, it inputs a sigmoid transformation to get a 0 and 1 binary result, which can be used in deciding whether the information from the previous time step will contribute in the next cell state or not. It is a basic forget gate structure. Once the forget gate chooses to consider cell state contribution from the previous C_{t-1} , the next step is to determine whether or not cell state input from the current time step will contribute. With the concepts of recurrent neural networks (RNNs), the output of the previous time step should be taken into account. Hence, the current input X_t can then be paired with the previous output h_{t-1} . In Fig. 6, functions i_t and C_t are given and defined in Eqs. (18) and (19), respectively:

$$i_t = \sigma(W(h_{t-1}, X_t) + b_i) \quad (18)$$

$$C_t = \tanh(W(h_{t-1}, X_t) + b_c) \quad (19)$$

Hence, in Eq. (18) i_t is the sigmoid function which decides whether the information will transfer or not while function as shown in Eq. (19). C_t estimates the new cell state value using \tanh transformation. As shown in Fig. 6, the multiplication function proceeds between i_t and C_t . The actual significance of this function is which inputs will be retained and how much percentage those inputs will be held. After this step, cell status can be updated accordingly as shown in Eq. (20).

$$C_t = f_t C_{t-1} + i_t x C_t \quad (20)$$

The result of the output shall be derived from the cell unit. The o_t and h_t in Fig. 6 are defined as shown in Eq. (21) and Eq. (22).

$$o_t = \sigma(W(h_{t-1}, X_t) + b_o) \quad (21)$$

$$h_t = o_t \tanh(C_t), \quad (22)$$

where o_t gate selects the part of input which will be selected as output composition and h_t is derived by determining how much effect cell state will work on those selected inputs. Although one LSTM cell unit will be enough for memory handling, in this paper, for achieving maximum accuracy the stacked LSTM cell structure is used in which the same LSTM cell stacked over each other as shown in Fig. 7.

The stacked LSTM algorithm is the most complex and challenging algorithm to implement among all other algorithms considered in this paper. The stacked multi-layered LSTM is used for the prediction of fuel flow. After

data preprocessing of the fuel flow feature, the fuel flow is divided into 80% training set and 20% test set, and develop data structure with 100-time steps and 01 output. After creating the data structure, the X_{train} is reshaped into three dimensions for the LSTM model. After initializing RNN, the first layer of LSTM with a 0.2 dropout layer was added to the model. After the first LSTM layer, the second LSTM layer was added also with a 0.2 dropout layer, and similarly, the third LSTM layer was added with a same 0.2 drop layer. After developing three LSTM layers stacked, the RNN is compiled with “adam” optimizer and loss of mean square error. After compiling the RNN, the fourth LSTM layer is added to the model with the same dropout layer of 0.2 and 60 LSTM units and the end output dense layer is added with unit 1. After developing a four-layer stacked LSTM, the model is compiled again with the same “adam” optimizer and loss function, i.e., mean square error. After developing the RNN/LSTM model, the next step is to fit the RNN/LSTM model to the training set. To avoid overfitting, the dropout layers are used with the early stop function in the RNN/LSTM fitting. Initially, 100 epochs are set with 32 batch sizes. The early stop function helps in avoiding overfitting as the model training automatically stops when the model is not improving further. After model training, the next step is to test the model with the test set as the model has not seen the test data and the test data is unknown to the model. Therefore, in this paper, the results are plotted between actual data and predicted data, while the model is evaluated. The deep learning models are prone to higher errors, but they require a significantly large amount of training time and high computational time and resources.

4 Results and discussion

In this section, the ML model results are compiled and discussed. The ML fuel flow (FF) predictive model results are defined concerning flight phases which are taxi, cruise and approach phases. Results section covers the details of the dataset, data preprocessing results, ML models graphical plots and model evaluation metrics. Graphical representation aims to show the difference between actual values vs predicted values. This section also provides all characteristics of all features and numerical results of the estimation experiment. The method of ML algorithms has been run on a computer that has a 64-bit Windows 10 operating system, 2.3 GHz processor and 12-GB RAM and 4 GB NVIDIA GEFORCE GTX graphic card. The performed method has been implemented in the Python programming language.

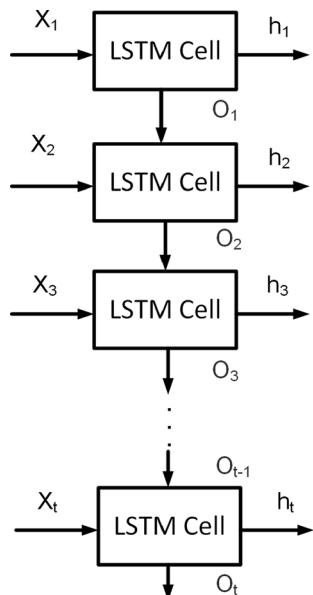


Fig. 7 Stacked LSTM structure

4.1 Dataset study

The dataset was extracted from the flight data recorder and published online by US-based company honey well in the competition hosted by CrowdANALYTIX to discover the factors affecting fuel efficiency using the flight recorder data [41, 42]. This dataset consists of 06 zip files, and each zip file contains 200 flight instances and a total of 1000 flight instances. The dataset consists of different independent features and fuel flow-dependent features. The dataset consists of different flight phases, but the focus of this paper is on three different flight phases that are taxi, cruise and approach. Each phase encountered data preprocessing separately. The details of flight features are given in Table 2. The dataset consists of a total of 226 features, but the most relevant features are discussed here. The table covers the flight operational data with abbreviation details and units.

4.2 Data preprocessing results

The very first step after data cleaning is to visualize the linear relationship between variables. Hence, the correlation bar plots of concern flight phases (taxi, cruise and approach) are developed using the seaborn Python library as shown in Figs. 8, 9 and 10, respectively. The linear relationship between variables is measured through it, and the variables' or features' significance for model training is decided through this procedure. Therefore, the features are selected on the basis of correlation. Out of 266 features, the

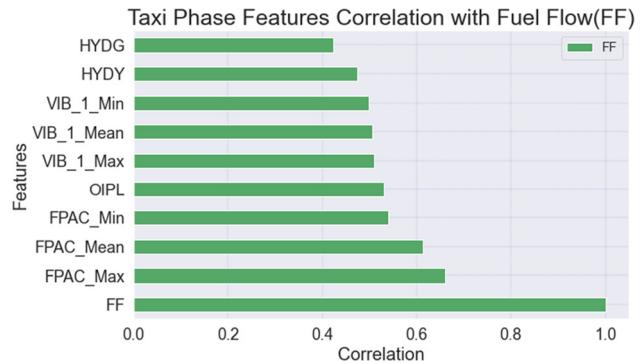


Fig. 8 Taxi phase correlation

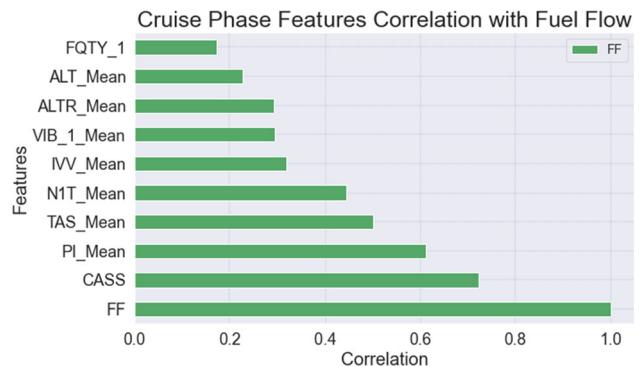


Fig. 9 Cruise phase correlation

top highly correlated features are selected. As each flight phase data has its own characteristics, the selected features of each phase are different. In this study, for analyzing data

Table 2 The dataset features description

Features	Descriptions	Unit	Sampling rate
ALTR_Mean	Altitude rate	ft/min	4
IVV_Mean	Inertial vertical speed	ft/min	16
TAT	Total air temperature	Deg	1
SAT	Static air temperature	Deg	1
LONG_Max	Longitudinal acceleration	g	4
PS_Mean	Static pressure	in	2
PSA_Mean	Average static pressure	MB	2
PT_Mean	Total pressure	MB	2
HYDG	Low hydraulic pressure green	N/A	1
HYDY	Low hydraulic pressure yellow	N/A	1
VIB_1_Mean	Engine vibration 1	IN/SEC	4
OIPL	Low oil pressure all engines	N/A	1
FQTY_1	Fuel quantity tank 1	lbs	1
FPAC_Min	Flight path acceleration	g	16
ALT_Mean	Pressure altitude LSP	ft	4
N1T_Mean	N1 target LSP	%RPM	4
TAS_Mean	True airspeed	Knots	4
CASS	Selected airspeed	knobs	1
PI_Mean	Impact pressure LSP	MB	2

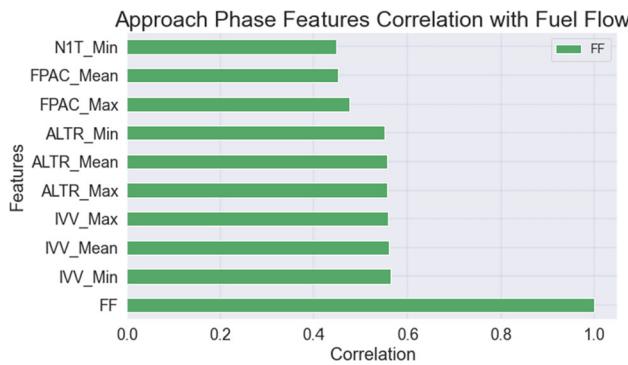


Fig. 10 Approach phase correlation

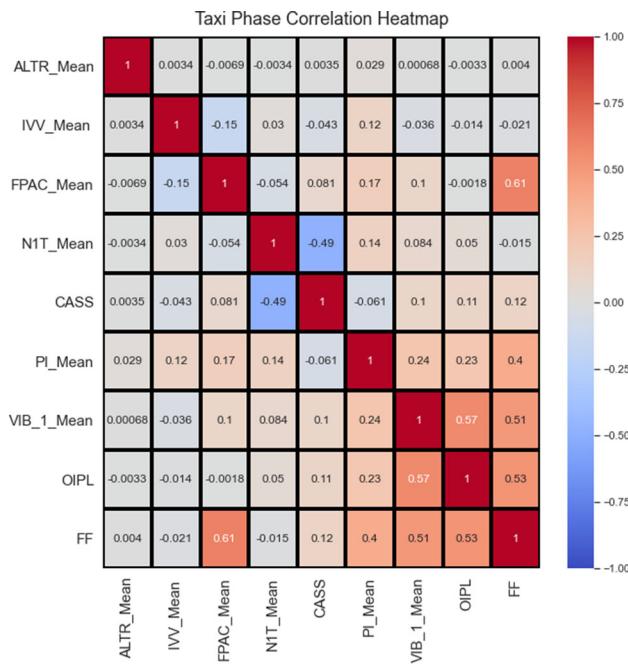


Fig. 11 Taxi phase multicollinearity test

duplication features correlation heatmap was developed using the Python data science library seaborn. The heatmap of taxi, cruise and approach phases is shown in Figs. 11, 12 and 13, respectively. The visualizing process is easier to understand and better than reading the huge tabular values; therefore, correlation heatmap helps in understanding features multicollinearity and dependencies upon each other. The cruise phase heatmap shows that "CASS" is the most highly correlated feature and it will dominate the model most. Furthermore, the multicollinearity can also be visualized and observed between "CASS" and "PI_Mean." Similarly, the "ALTR_Mean" and "IVV_Mean" have multicollinearity. In the Taxi phase, the "VIB_1_Mean" and "OIPL" are slightly collinear, while in the approach phase, several variables have inter-dependencies or multicollinearity among them. Therefore, we use the scikit-learn library to resolve the multicollinearity issue. We performed

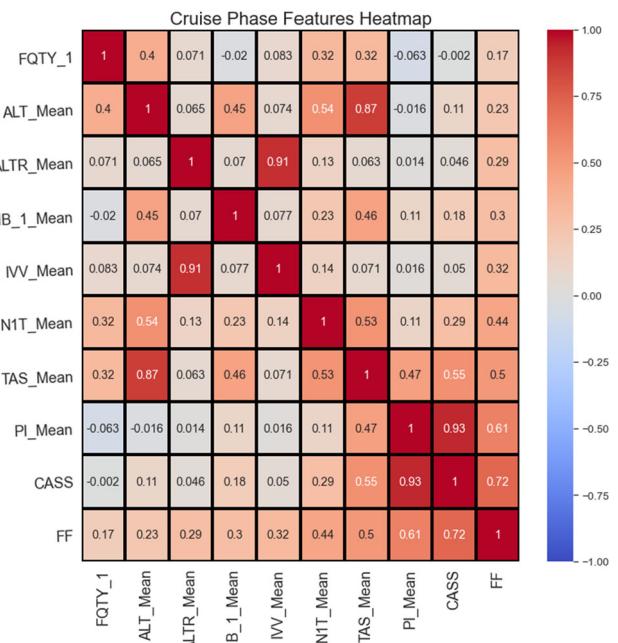


Fig. 12 Cruise phase multicollinearity test

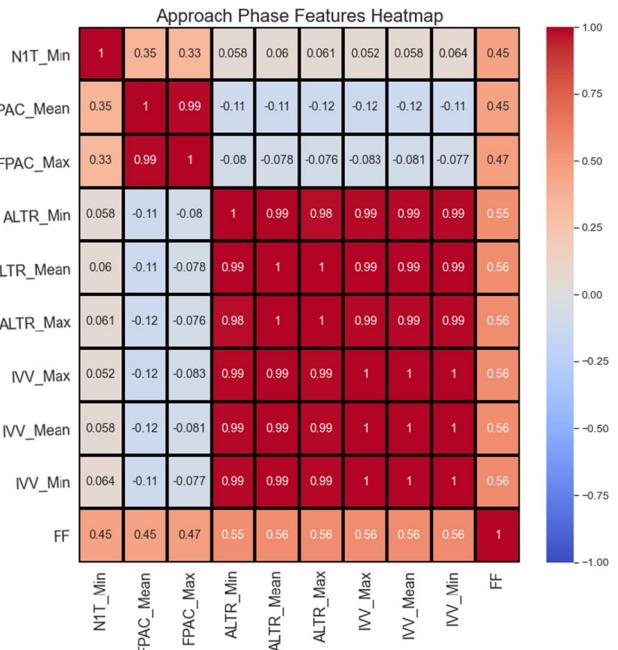


Fig. 13 Approach phase multicollinearity test

data statistics on taxi, cruise and approach flight phases as described in Tables 3, 4 and 5, respectively. We have calculated the mean, standard deviation, minimum values, maximum values, etc., for individual features. The results show that the approach phase "ALTR_Mean" has a higher value of standard deviation, i.e., 759.8, while the taxi and cruise phase has a low standard deviation as compared to the approach phase which is 40.09 and 135.46. Similarly,

Table 3 Taxi phase features description

	ALTR _Mean	IVV _Mean	FPAC _Mean	N1T _Mean	CASS	PI _Mean	VIB _1_Mean	OIPL	FF
count	236211.0	236211.0	236211.0	236211.0	236211.0	236211.0	236211.0	236211.0	236211.0
mean	-8.018627	-0.15155	-0.0002338	92.044	127.3660	0.30087	0.071430	13.428	1318.41
std	40.090201	6.4586936	0.0180770	2.3205	7.567386	0.364036	0.04492	3.598263	739.5049
min	-860.0	-64.4375	-0.245654	83.9531	99.0	0.0	0.0	0.0	0.0
25%	-20.0	-2.5625	-0.000427445	90.75	120.0	0.078125	0.0625	15.0	1334.0
50%	-8.0	-0.25	0.0	92.453125	131.0	0.15625	0.0625	15.0	1392.0
75%	4.0	1.4375	0.0	93.75	133.0	0.390625	0.09375	15.0	1436.0
max	916.0	64.875	0.366252869	96.484375	142.0	7.359375	0.3125	15.0	12364.0

Table 4 Cruise phase dataset description

	ALTR_Mean	IVV_Mean	N1T_Mean	CASS	TAS_Mean	PI_Mean	VIB_1_Mean	FF
Count	273730.0	273730.0	273730.0	273730.0	273730.0	273730.0	273730.0	273730.0
Mean	-18.1886	-11.556	85.48656	272.4889	401.166	125.448	0.15210	5018.08
Std	135.462	133.2844	15.31182	17.79751	44.8321	16.3299	0.04379	631.165
Min	-2404.0	-2627.75	0.0	149.0	222.89	58.84375	0.0	1166.0
25%	-48.0	-23.5	85.238	265.0	391.14	118.5625	0.125	4822.0
50%	-8.0	-1.0625	90.789	274.0	420.60	124.90625	0.15625	5112.0
75%	28.0	20.625	92.636	285.0	429.46875	136.09375	0.1875	5358.0
Max	3160.0	3419.0625	96.988	300.0	448.21875	158.328125	0.25	8944.0

Table 5 Approach phase dataset features

	ALTR _Mean	IVV _Mean	FPAC _Mean	N1T _Mean	CASS	TAS _Mean	PI _Mean	VIB _1_Mean	OIPL	FF
Count	224081.0	224081.0	224081.0	224081.0	224081.0	224081.0	224081.0	224081.0	224081.0	224081.0
Mean	-1154.9	-1145.15	-0.012	46.46	228.52	285.7	96.61	0.10	15.0	2968.7
Std	759.8732	760.8475	0.023777	28.075213	58.56467	96.21708	41.4078	0.038833	0.0	1116.877
Min	-6080.0	-6241.4375	-0.40136	0.0	80.0	0.0	8.328125	0.0	15.0	1176.0
25%	-1696.0	-1692.0	-0.01783	24.78125	180.0	204.015625	59.28125	0.09375	15.0	1842.0
50%	-1088.0	-1067.625	-0.00763	55.148	250.0	284.421875	105.109375	0.09375	15.0	2936.0
75%	-632.0	-617.875	-0.000366	67.60	280.0	370.453125	134.28125	0.125	15.0	3934.0
Max	1308.0	1252.75	0.1495	96.65	300.0	446.859375	158.71875	0.3125	15.0	8416.0

the “IVV_Mean” has a 760.84 standard deviation which is relatively more as compared to both other flight phases. The independent variable is “FF” which has a mean value of 1318.41 gallons per hour and a standard deviation of 739.50 in the taxi flight phase. Similarly, in the cruise phase “FF” has a mean value of 5018.08 gallons per hour and a standard deviation of 631.16. But in the approach phase the “FF” has a mean value of 2968 and a standard deviation value of 1116.87. The taxi phase has 2,36,211 instances. The taxi phase is the flight phase during which

an aircraft moves on the surface of an airfield under its own power, omitting takeoff and landing. The cruise phase has 2,73,730 number of instances, the cruise phase is the longest phase in the flight, and most of the time aircraft are in the cruise phase. Cruising takes up the majority of a flight and may include changes in heading (direction of flight) while maintaining a constant velocity and altitude. The approach can be defined as the phase of flight that begins when an airworthy aircraft under the control of the flight crew descends below 5,000 feet AGL with the

intention of conducting an approach and ends when the aircraft crosses the approach end of the landing runway (runway threshold) or when a go-around maneuver is initiated. Moreover, it has 224,081 instances.

4.3 Flight phases model results

After data examination and data preprocessing, the datasets of all flight phases are ready for machine and deep learning model development. In this study, as discussed above the flight phases considered are taxi, cruise and approach. Hence, MLR, DTR, SVR and LSTM models are separately applied in each flight phase. The plots visualization is for understanding the difference between actual and predicted data values.

The metrics that are used for evaluation are mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination r^2 . The MAE, RMSE and r^2 are calculated for each flight phase separately. The fundamental difference between the real or true values and the expected values is an error. Absolute discrepancy means if there is a negative sign in the outcome it is neglected. Hence, MAE is the difference between the true and predicted values. MAE takes the average of this error from each sample and provides the result as shown in Eq. 24. MAE method is implemented using sklearn's mean absolute error method. RMSE is the standard deviation of the errors that take place when a dataset prediction is made. It is the same as MSE, except while evaluating the accuracy of the model, the root of the value is taken into consideration as shown in Eq. 23. RMSE method is implemented by taking the root of the sklearn mean square error function.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (23)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (24)$$

The coefficient of determination measures how well a model matches a given dataset. It shows how close the regression line (i.e., the expected values plotted) is to the real data values. The r^2 value ranges from 0 to 1, with 0 indicating that the model does not match the given data and 1 indicating that the model matches the data perfectly as mentioned in Eq. 25.

$$r^2 = MSS/TSS = (TSS - RSS)/TSS \quad (25)$$

where RSS is the sum of squares of residuals and TSS is the total sum of squares. MAE, RMSE and coefficient of determination r^2 score are used for each flight phase model ML evaluation.

The very first machine learning model that is applied to flight phases dataset is MLR. We use the scikit-learn Python library to train and test the models. The MLR model deployment takes the minimum time. It is acceptable when one just wants to get an idea of the trend line. MLR model is not so accurate, but it can be applied when the application is related to trend prediction or basic prediction. It is necessary to evaluate the machine learning models; therefore, the evaluation metrics selected in this study are frequently used for the assessment of general machine learning problems. The scatter plots of actual vs predicted are one of the effective forms of data visualization. We can tell pretty much everything from it. Hence, we plot the multiple scatter plots of actual vs predicted to visualize the difference between the actual readings and the predicted ones. The MLR model plots of flight phases are shown in Figs. 14, 15 and 16. One can easily visualize the difference between the actual and the predicted one in the scatter plots. In Fig. 15 of the cruise phase, the predicted values and actual values are not so close; few of the predicted values are diverted. Similarly, in Fig. 16 most of the actual values are not overlapping the predicted one; we can conclude through this visualization that the approach phase does not perform satisfactorily. Meanwhile, in Fig. 14 of the taxi phase, we can notice that the predicted values are very close to the actual one; therefore, we presume that the MLR model performs relatively well in the cruise phase as compared to other flight phases. In the case of engine estimation, the MLR model can only be used for basic laboratory testing. The MLR models are evaluated in Tables 6, 7 and 8 in all discussed flight phases that are taxi, cruise and approach phases, respectively. The RMSE values are **0.5456**, **0.6177** and **0.5860**, MAE are **0.29,475**, **0.3837** and **0.47,180**, r^2 are **0.690**, **0.6119** and **0.6555** and the training time are 0.023899s, 0.0041s and 0.0105899s. In the taxi phase, the error is minimum as compared to the cruise and approach phases as the features selected for each

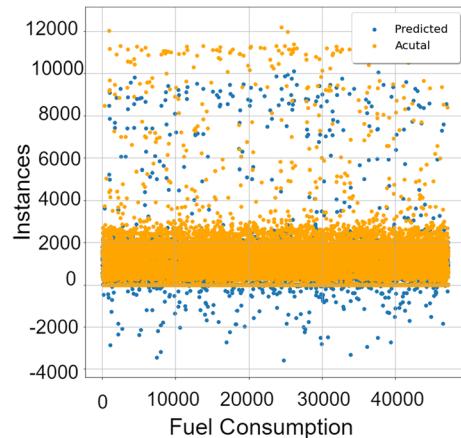
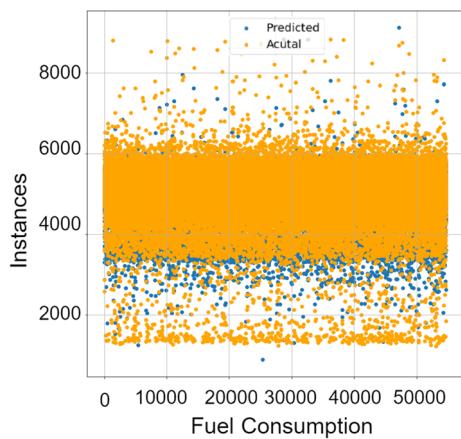
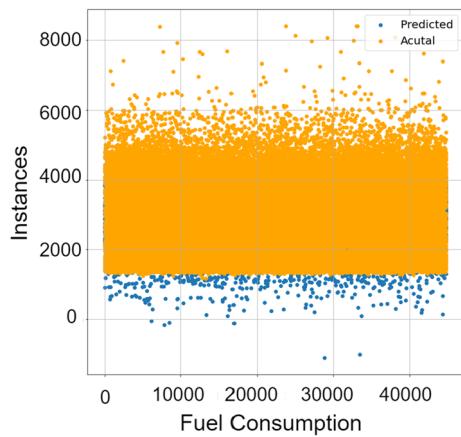


Fig. 14 MLR taxi phase

**Fig. 15** MLR cruise phase**Fig. 16** MLR approach phase**Table 6** Taxi phase performance evaluation matrix

Methods	Taxi phase performance			
	RMSE	MAE	r^2	Training Time
MLR	0.545	0.294	0.690	0.023899 s
DTR	0.459	0.171	0.791	0.35 s
SVR	0.431	0.159	0.815	1 hr 37 min 54 s
LSTM	0.151	0.082	0.939	48 hr 2 min 5 s

flight phase are different based on data preprocessing. The results of the cruise and approach phase are relatively close.

The DTR model plots of flight phases taxi, cruise and approach are shown in Figs. 17, 18 and 19, respectively. In Tables 6, 7 and 8, the DTR models of taxi, cruise and approach flight phases, respectively, are evaluated. The RMSE values are **0.459**, **0.1955** and **0.49,380**, MAE are **0.17,175**, **0.0644** and **0.2992**, r^2 are **0.791**, **0.9611** and **0.7554** and the training time are 0.35 s, 1.23 s and 1.55 s.

Table 7 Cruise phase model performance evaluation matrix

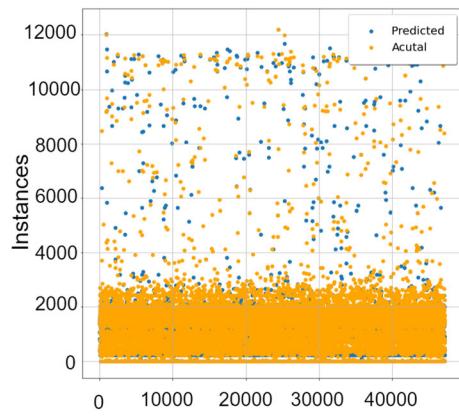
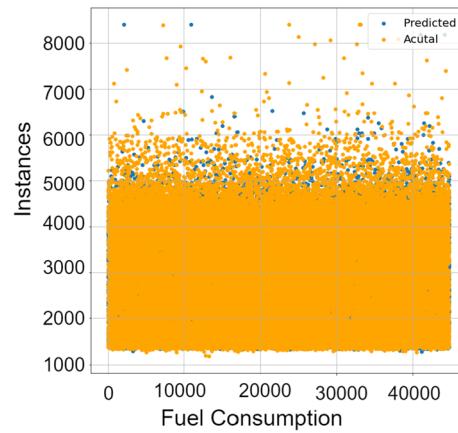
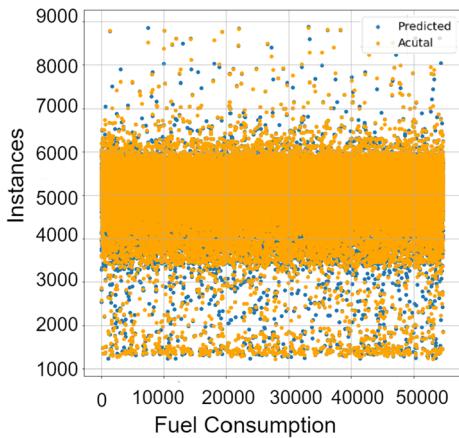
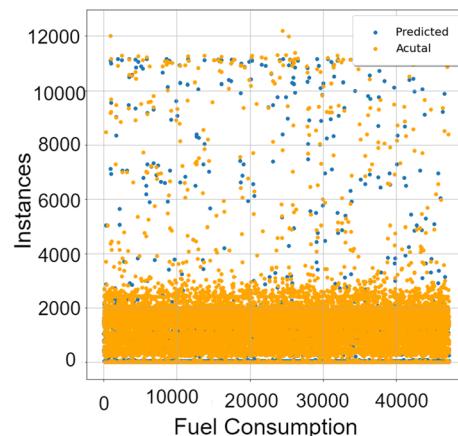
Methods	Cruise phase performance			
	RMSE	MAE	r^2	Training Time
MLR	0.6177	0.3837	0.6119	0.0041 s
DTR	0.1955	0.0644	0.9611	1.23 s
SVR	0.27745	0.11445	0.9217	1 hr 58 m 4 s
LSTM	0.105	0.0531	0.97,687	41 hr 12 min 2 s

The cruise phase have less error as compared to other, and the taxi phase and approach phases errors are slightly closed, but the overall performance of DTR is good as it requires less computational power and time. DTR algorithm is recommended when one requires fast and more accurate results as compared to MLR. According to model evaluation results, it seems like DTR which is most suitable for engine tuning laboratory applications as it requires very little computational power and time. It can be deployed in the field as on-field computational power is very limited. Figure 17 shows the actual vs predicted values scatter plot of the taxi flight phase and it is easily noticeable that the scatter plot of the DTR taxi phase shows better performance as compared to the MLR taxi phase. The blue dots of predicted values are very close to the orange dots of actual values. Similarly, in Fig. 18 the cruise phase scatter plot shows that the DTR results in the cruise phase are relatively adequate as compared to MLR results of the cruise phase. In the MLR cruise phase model, the upper and lower orange dots of actual data do not intersect with the predicted one but in the DTR cruise phase plot the upper and lower actual values dots are very close to each other, which shows the performance of DTR in the cruise phase is exemplary. Figure 19 shows the DTR approach phase scatter plot and it is noticeable that the blue dots of predicted values do not proceed lower to the actual value, while in Fig. 16 it is observable that the predicted values go lower to the actual value. It indicates that the DTR performance is better as compared to MLR.

The support vector regression (SVR) scatter plots of a representation of taxi, cruise and approach flight phases are shown in Figs. 20, 21 and 22, respectively. Figure 20 shows the scatter plot of the SVR taxi phase model. It shows that the values of predicted and actual are alike in DTR taxi phase plots, which indicates that the performance of SVR and DTR in the taxi phase is close. Similarly, the cruise and approach phase scatter plots of actual vs predicted in Figs. 21 and 22 show that the performance of SVR is near to DTR. In Tables 6, 7 and 8, the SVR performance is evaluated in taxi, cruise and approach flight phase, respectively. The RMSE values are **0.431**, **0.27,745** and **0.4089**, MAE are **0.159**, **0.11,445** and **0.2802**, r^2 are

Table 8 Approach phase performance evaluation matrix

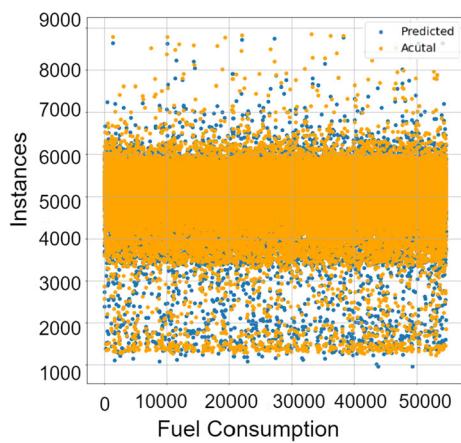
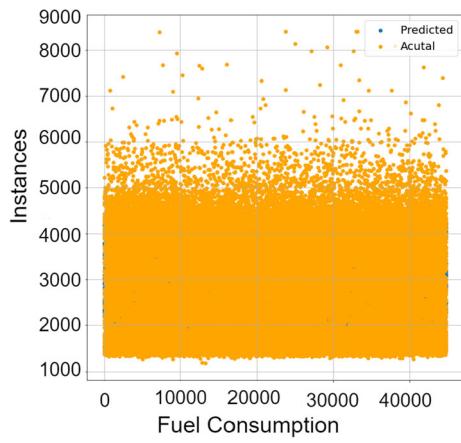
Methods	Approach phase performance			
	RMSE	MAE	r^2	Training Time
MLR	0.5860	0.47180	0.6555	0.0105899 s
DTR	0.49380	0.2992	0.7554	1.55 s
SVR	0.4089	0.2802	0.8323	1 hr 14 min 13 s
LSTM	0.0998	0.07,955	0.96,821	38 hr 1 min 3 s

**Fig. 17** DTR taxi phase**Fig. 19** DTR approach phase**Fig. 18** DTR cruise phase**Fig. 20** SVR taxi phase

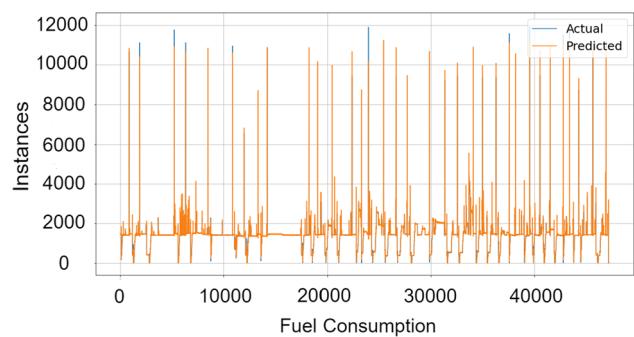
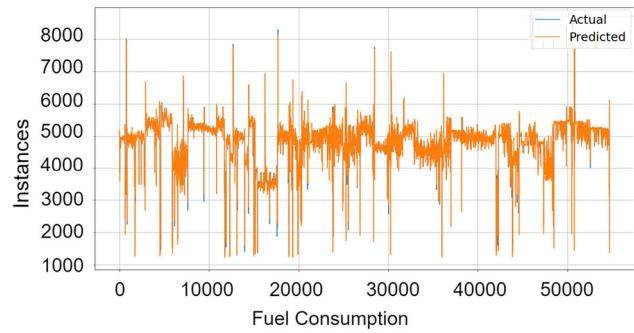
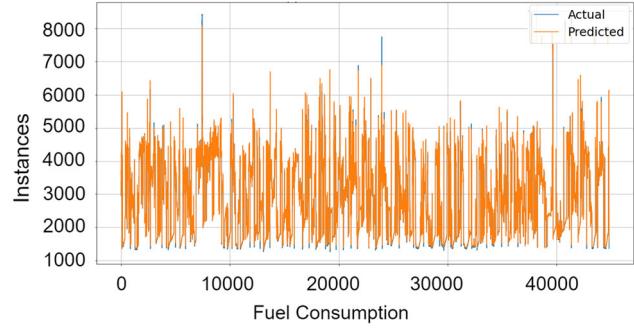
0.815, **0.9217** and **0.8323** and the training time are 1 h 37 min 54 s, 1 h 58 m 4 s and 1 h 14 min 13 s. In cruise, the results of SVR are much better as compared to others. In the approach phase, the results are not that good, while in the taxi phase, the results are better as compared to the approach phase. In this paper, SVR takes greater training time as compared to MLR and DTR. SVR model training requires high computational power as compared to MLR and DTR models. SVR models also require more

deployment time relative to MLR and DTR. Therefore, SVR can be used for ETB fuel consumption estimation, but it requires relatively high computational systems to train multiple flights of data.

In this paper, the game-changing algorithm is LSTM as it outclasses all other algorithms in terms of accuracy, but it requires high computational power and hours of continuous model training. As a stacked-based LSTM model is

**Fig. 21** SVR cruise phase**Fig. 22** SVR approach phase

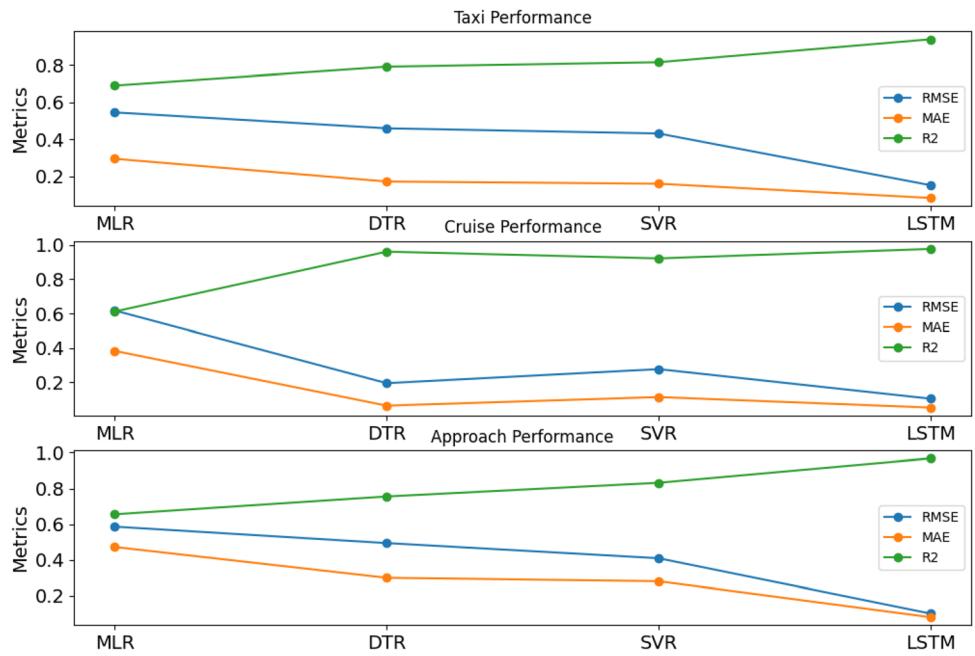
developed in this paper, it requires more training time as compared to the conventional LSTM model. The training time of taxi, cruise and approach phases is 48hr 2min 5 s, 41hr 12min 2 s and 38hr 1min 3 s, respectively. It is a multi-layered LSTM model with 32 batches sizes, 100 epochs, adam optimizer and 0.2 dropout layer used after each LSTM layer. With the help of the early stopped function, the LSTM model training stopped on 93 epochs, 76 epochs and 69 epochs to avoid overfitting. The pros of the LSTM model are its accuracy, while the cons of LSTM are its training time. One can use any of the algorithms in the flight phases depending on the application and case scenarios. If the engine tuning team has high computation capability and requires high accuracy, then they need to go for the LSTM model; otherwise, decision tree model is acceptable for low computational systems. Most of the time, the aircraft is in the cruise phase. Therefore, the cruise phase is also a linear phase. There are no such maneuvers in the cruise phase. Once the aircraft enters the cruise phase, the autopilot gets on and the pilot relatively relaxes as compared to other phases. In the cruise phase

**Fig. 23** LSTM taxi phase**Fig. 24** LSTM cruise phase**Fig. 25** LSTM approach phase

scenario, DTR is recommended as its accuracy is close to LSTM and has very little training time. Figure 23 shows the line plot of actual values compared to predicted values of the taxi flight phase. It is quite noticeable that the orange line of actual values are overlapping the predicted values, which reveals that the performance of LSTM is better than MLR, DTR and SVR flight phases. Similarly, Figs. 24 and 25 show the line plots of actual vs predicted values. The line plot and both graphs indicate that the orange line of actual values is quite close to the predicted one.

The LSTM model performance is also evaluated in taxi, cruise and approach flight phases as discussed in Tables 6, 7 and 8, respectively. In the taxi flight phase, the LSTM RMSE is **0.151**, MAE is **0.082**, the coefficient of determination is **0.939** and the training period is 48hr 2min and

Fig. 26 Comparative performance of machine learning techniques across multiple flight phases



5 s. In the cruise phase, the RMSE of the LSTM model is **0.105**, MAE is **0.05,311**, coefficient of determination is **0.9768** and training time is 41hr 12 min 2 s. Similarly, in approach phase RMSE is **0.0998**, MAE is **0.07,955**, coefficient of determination is **0.96,821** and training time is around 38hr 1min and 3 s.

4.4 Discussion

The comparison of different models is graphically presented in Fig. 26. Among the algorithms considered in both the existing literature and this paper, the LSTM consistently outperforms the others. However, it is important to note that the LSTM requires significant computational power and longer training times. In the taxi phase, the second best algorithm is SVR, as its accuracy is comparable to that of DTR. While LSTM achieves superior performance, its computational demands and training time make SVR a viable alternative. During the cruise phase, the DTR algorithm demonstrates accuracy comparable to LSTM. However, it is worth mentioning that LSTM's computational requirements and training time remain considerably higher. In this phase, DTR outperforms SVR in terms of accuracy. In the approach phase, the SVR algorithm outperforms the DTR algorithm in terms of accuracy. However, it requires relatively longer training time compared to DTR. Nevertheless, the LSTM algorithm consistently delivers exceptional performance across all flight phases. It is important to note that when developing the LSTM model, the training and test set data should not be shuffled. Since the LSTM algorithm relies on time-series prediction, preserving the sequence of the data is

crucial. In the case of LSTM, finding suitable hyperparameters is a challenging task. Determining the optimal number of hidden layers, units, dropout layers, optimizer and other hyperparameters requires careful consideration. In this study, a hard-coded method was employed to find suitable hyperparameters. However, future research should focus on developing an automatic method for tuning LSTM hyperparameters.

While this paper successfully models individual flight phases (taxi, cruise, approach) with minimal errors, it is important to acknowledge that when developing a model for the entire flight mission, errors may increase. It is an area that warrants further investigation in future research.

Decision tree regression (DTR) is known to have larger errors compared to other algorithms. However, its advantage lies in requiring less computational time. Thus, DTR can be a suitable choice for applications such as fuel consumption estimation where high computational power may not be readily available.

As discussed above, it is practically very difficult to unmount the engine from the aircraft and mount it on the ETB for analyzing and estimating a single variable such as fuel consumption or oil temperature. The unmounting processes require huge manpower and machinery. This problem can be resolved if the fuel consumption of an air vehicle is estimated without unmounting the engine from the aircraft by applying data analytics and machine learning models. Therefore, in this paper, we have applied multiple machine learning models to overcome the problem.

5 Conclusion and future research directions

The main focus of this paper was to estimate the fuel consumption of an aircraft with the help of multiple machine learning algorithms in which the LSTM and DTR algorithms perform well among all other applied algorithms, but both have trade-offs of computation time, power and performance. The DTR has an RMSE value of **0.459** in the taxi phase, in the cruise phase, the RMSE value is **0.1955**, while in the approach phase, the RMSE value is **0.4938**. But the RMSE value of the LSTM model in the taxi phase is **0.151**, in the cruise phase is **0.105** and in the approach phase is **0.0998**. It is concluded that there is no need to unmount an engine from aircraft and place it on ETB for the purpose of engine fuel estimation. In this paper, we have used the flight data recorder data instead of ETB ground run data to verify that we can estimate the fuel consumption of aircraft without going into ETB and spending millions of dollars on manpower and machinery on mounting and unmounting aircraft engines from aircraft to ETB just for estimating a single variable such as fuel consumption. This article will serve as a reference point for potential work in this field. In this research, the ML model's hyperparameters are tuned through manual methods, especially in RNN/LSTM model. In the LSTM model, the hidden layers, dropout layers, optimizer, units and batch size iteratively varied manually to achieve maximum accuracy. In future work, it is recommended that one shall implement the automatic or smart method of hyperparameters selection or tuning to achieve maximum accuracy. In this consideration, algorithms need to automate for finding the appropriate values for hyperparameters. Furthermore, in future work concerns, it is also necessary to model the aircraft fuel flow algorithm for multiple flights of data and in-flight phases such as climb, descent and landing. Moreover, research on AI-based hardware for aircraft parameter estimation is also a future research goal.

Data availability The data used in this research is sourced from the Honeywell dataset, which is publicly available. The dataset can be accessed via ref. [42].

Declarations

Conflict of Interest The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Sultana S (2018) Economic growth of airlines industry: an overview of domestic airlines in Bangladesh. *J Manag Res Anal* 5
2. Zhang F, Graham DJ (2020) Air transport and economic growth: a review of the impact mechanism and causal relationships. *Transp Rev* 40(4):506–528
3. Wei H, Liu W, Chen X, Yang Q, Li J, Chen H (2019) Renewable bio-jet fuel production for aviation: a review. *Fuel* 254:115599
4. Tanzil AH, Brandt K, Wolcott M, Zhang X, Garcia-Perez M (2021) Strategic assessment of sustainable aviation fuel production technologies: yield improvement and cost reduction opportunities. *Biomass Bioenergy* 145:105942
5. Martinez-Valencia L, Garcia-Perez M, Wolcott MP (2021) Supply chain configuration of sustainable aviation fuel: review, challenges, and pathways for including environmental and social benefits. *Renew Sustain Energy Rev* 152:111680
6. Baumann S, Neidhardt T, Klingauf U (2021) Evaluation of the aircraft fuel economy using advanced statistics and machine learning. *CEAS Aeronaut J* 12(3):669–681
7. Trejo-Pech CO, Larson JA, English BC, Yu TE (2019) Cost and profitability analysis of a prospective pennycress to sustainable aviation fuel supply chain in southern usa. *Energies* 12(16):3055
8. Doganis R (2013) Flying off course: the economics of international airlines. Routledge
9. Badykov RR, Panshin RA, Tremkina OV, Prokofieva AA (2021) Utilization of low-potential energy through to the use in aircraft fuel system. In: IOP conference series: materials science and engineering, vol 1102(1). IOP Publishing, p 012007
10. Eguea JP, da Silva GPG, Catalano FM (2020) Fuel efficiency improvement on a business jet using a camber morphing winglet concept. *Aerosp Sci Technol* 96:105542
11. Manna S, Biswas S, Kundu R, Rakshit S, Gupta P, Barman S (2017) A statistical approach to predict flight delay using gradient boosted decision tree. In: 2017 International conference on computational intelligence in data science (ICCIDIS). IEEE, pp 1–5
12. Khadilkar H, Balakrishnan H (2012) Estimation of aircraft taxi fuel burn using flight data recorder archives. *Transp Res Part D Transp Environ* 17(7):532–537
13. Mazareanu E (2021) Commercial airlines: worldwide fuel consumption 2005–2022. Oct. [Online]. Available: <https://www.statista.com/statistics/655057/fuel-consumption-of-airlines-worldwide/>
14. Gouveia OR, Borges A, Costa D, Lopes P, Coelho D, Ferreira C, Serrano L. Development of a low-cost test bench for heavy-duty combustion engines
15. Mattingly JD, Heiser WH, Pratt DT (2002) Aircraft engine design. American Institute of Aeronautics and Astronautics
16. Dalon T. Surrogate-based optimization for optimal automatic calibration of modern automotive combustion engines at engine test bench
17. Kondratenko O, Deyneko N, Vambol S (2015) Engine test bench as a source of danger factors in experimental researches. Ph.D. dissertation, NTU
18. Wang Y, Shi Y, Cai M, Xu W, Pan T, Yu Q (2019) Study of fuel-controlled aircraft engine for fuel-powered unmanned aerial vehicle: energy conversion analysis and optimization. *IEEE Access* 7:109 246–109 258
19. Dimiduk DM, Holm EA, Niegzoda SR (2018) Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr Mater Manuf Innov* 7(3):157–172
20. Jasra S, Gauci J, Muscat A, Valentino G, Zammit-Mangion D, Camilleri R (2018) Literature review of machine learning techniques to analyse flight data
21. Muhammad I, Yan Z (2015) Supervised machine learning approaches: a survey. *ICTACT J Soft Comput* 5(3):1

22. eyeofbill, F-16 jet engine test at full afterburner in the hush house. Mar 2017. [Online]. Available: <https://www.youtube.com/watch?v=Oj4w7i-TqsE>
23. Collins BP (1982) Estimation of aircraft fuel consumption. *J Aircr* 19(11):969–975
24. Kapeller H, Dvorak D, Šimić D (2021) Improvement and investigation of the requirements for electric vehicles by the use of hvac modeling. *HighTech Innov J* 2(1):67–76
25. Qerimi D, Dimitrieska C, Vasilevska S, Rrecaj AA (2020) Modeling of the solar thermal energy use in urban areas. *Civ Eng J* 6(7):1349–1367
26. Nuic A (2011) User manual for the base of aircraft data (bada). Eurocontrol Experimental Centre, Cedex, France, revision, vol. 3
27. Namar MM, Jahanian O, Shafaghat R, Nikzadfar K (2021) Engine downsizing: global approach to reduce emissions: a world-wide review. *HighTech Innov J* 2(4):384–399
28. Baumann S, Klingauf U (2020) Modeling of aircraft fuel consumption using machine learning algorithms. *CEAS Aeronaut J* 11(1):277–287
29. Huang C, Cheng X (2022) Estimation of aircraft fuel consumption by modeling flight data from avionics systems. *J Air Transp Manag* 99:102181
30. Horiguchi Y, Baba Y, Kashima H, Suzuki M, Kayahara H, Maeno J (2017) Predicting fuel consumption and flight delays for low-cost airlines. In: Proceedings of the thirty-first AAAI conference on artificial intelligence, pp 4686–4693
31. Wang X, Chen X (2014) A support vector method for modeling civil aircraft fuel consumption with roc optimization. In: 2014 enterprise systems conference. IEEE, pp 112–116
32. Chati YS, Balakrishnan H (2016) Statistical modeling of aircraft engine fuel flow rate. In: 30th congress of the international council of the aeronautical science
33. Kang L, Hansen M (2017) Quantile regression based estimation of statistical contingency fuel. In: Twelfth USA/Europe air traffic management research and development seminar (ATM2017)
34. Huang C, Xu Y, Johnson ME (2017) Statistical modeling of the fuel flow rate of ga piston engine aircraft using flight operational data. *Transp Res Part D Transp Environ* 53:50–62
35. Yanto J, Liem RP (2022) Cluster-based aircraft fuel estimation model for effective and efficient fuel budgeting on new routes. *Aerospace* 9(10):624
36. Uzun M, Demirezen MU, Inalhan G (2021) Physics guided deep learning for data-driven aircraft fuel consumption modeling. *Aerospace* 8(2):44
37. Baklacioglu T (2016) Modeling the fuel flow-rate of transport aircraft during flight phases using genetic algorithm-optimized neural networks. *Aerosp Sci Technol* 49:52–62
38. Pagoni I, Psarakis-Kalouptsidi V (2017) Calculation of aircraft fuel consumption and CO_2 emissions based on path profile estimation by clustering and registration. *Transp Res Part D Transp Environ* 54:172–190
39. Yanto J, Liem RP (2018) Aircraft fuel burn performance study: a data-enhanced modeling approach. *Transp Res Part D Transp Environ* 65:574–595
40. Senzig DA, Fleming GG, Iovinelli RJ (2009) Modeling of terminal-area airplane fuel consumption. *J Aircr* 46(4):1089–1093
41. Srivastava I, Moharir AK, Yadam G (2020) Learning interpretable rules contributing to maximal fuel rate flow consumption in an aircraft using rule based algorithms. In: 2020 IEEE international conference for innovation in technology (INOCON). IEEE, pp 1–8
42. Predict fuel flow rate of airplanes during different phases of a flight (2021) [Online]. Available: <https://www.crowdanalytix.com/contests/predict-fuel-flow-rate-of-airplanes-during-different-phases-of-a-flight>
43. Li M, Zhou Q (2017) Industrial big data visualization: a case study using flight data recordings to discover the factors affecting the airplane fuel efficiency. In: 2017 IEEE Trustcom/BigDataSE/ICESS. IEEE, pp 853–858
44. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
45. Sousa JC, Jorge HM, Neves LP (2014) Short-term load forecasting based on support vector regression and load profiling. *Int J Energy Res* 38(3):350–362

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.