

REGULAR PAPER

# Prediction of departure flight delays through the use of predictive tools based on machine learning/deep learning algorithms

J.G. Muros Anguita<sup>1</sup>  and O. Díaz Olariaga<sup>2</sup> 

<sup>1</sup>Escuela Internacional de Postgrado, Universidad de Granada, Granada, Spain and <sup>2</sup>Facultad de Ingeniería Civil, Universidad Santo Tomás, Bogota, Colombia

**Corresponding author:** O. Díaz Olariaga; Email: [oscardiazolariaga@usta.edu.co](mailto:oscardiazolariaga@usta.edu.co)

**Received:** 5 November 2022; **Revised:** 25 April 2023; **Accepted:** 26 April 2023

**Keywords:** Flight delay; Air traffic; Air transport; Airport; Machine learning; Deep learning

## Abstract

The objective of this research is to predict the delays in the departure of scheduled commercial flights through a methodology that uses predictive tools based on machine learning/deep learning (ML/DL), with supervised training in regression, based on the available flight datasets. Since the novel contribution of this work is, first, to make the comparison of the predictions in terms of means and statistical variance of the different ML/DL models implemented and, second, to determine the coefficients of the importance of the features or flight attributes, using ML methods known as permutation importance, it is possible to rank the importance of flight attributes by their influence in determining the delay time and reduce the problem of selecting the most important flight attributes. From the results obtained, it is worth mentioning that the model that presents the best performance is the ensemble or combinatorial method of random forest regressor models, with an acceptable prediction range (measured with the root-mean-square-error).

## Nomenclature

ACI	Airports Council International
ADAM	Adaptive Moment Estimation
AIBT	actual on-block time
ANN	artificial neural network
AOBT	actual off-block time
ATM	Air Traffic Management
BGR	bagging gradient
DL	deep learning
DTR	decision trees
ETR	extra trees
FAA	Federal Aviation Administration
FSC	full-service carrier
HBR	Huber Regressor
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
KNN	k-nearest neighbors
LCC	low-cost carrier
LIR	simple linear regression
LSTM	long short-term memory
ML	machine learning
MLP	multi-layer perceptron

NN	neural network
RFR	Random Forest
ReLU	rectified linear unit
RMSE	root-mean-square-error
RMSProp	root mean square propagation
SGD	stochastic gradient descent
SVM	support vector machine
XGR	extreme gradient boosting

## 1.0 Introduction

Demand for air transport worldwide has grown vertiginously in the last three decades [1, 2]. Airlines require more capacity and more availability from airports (in infrastructure) to operate routes and frequencies designed to respond to this demand. The global airport industry has also reacted by permanently expanding its infrastructure and even building new airports with a vast volume of public and private investment [3]. However, the demand and the response have not grown equally fast: the capacity offered or installed in many airports in the world cannot fully satisfy the needs of airlines. This situation has resulted, first, in airport congestion and, consequently, in the generation of delays [4–6]. For this reason, many studies, both technical and academic, state that delay is a concept that is closely linked to the capacity of airports [7–9].

The volume of delays has been increasing in the last two decades, not only in its magnitude but also in its negative economic impact, mainly in terms of costs [10–14]. In the aviation system of the United States in 2019, 18.87% of all flights in the system had a delay [15]; in Europe, in 2019, 46.4% of all takeoffs departed late (whatever the reason for the delay), and the average (time) of these delays (of all late departed flights, concerning time scheduled departure) was 28.4 minutes [16]; in Colombia, the country used as a case for this research, in 2019, 28.60% and 22.54% of domestic and international flights, respectively, departed with delay [17]. Regarding the economic impact of delays, according to Federal Aviation Administration (FAA) estimates [18], the cost generated in the United States aviation system in 2019 was USD 33 billion (including all items); in Europe, the total cost of delays for 2012 was estimated at EUR 11.2 billion (in all respects) [19]; in Colombia, the cost of delays, only for airlines (and only for passenger compensation), in 2019 was USD 6.5 million [17].

The objective of the present research is to predict the delays in the departures of scheduled commercial flights, using the data of the complete airport network of Colombia (country-case study: 58 airports) observing the data of 357,595 flight departures for the entire year 2018 (study period), including both domestic and international departures. For this, and as a methodology, predictive tools based on machine learning/deep learning (ML/DL) were used, with supervised training (that is, the delays of the dataset flights used in training and testing are known) in prediction in regression (which will result in an actual expected delay time for a flight departure). The novel contribution of this work is, on the one hand, that it compares the predictions in terms of mean and statistical variance according to the different ML/DL models implemented (ten in total), so that the optimum model can be chosen from among them and, on the other hand, that it determines the coefficients of importance of the flight attributes, using ML methods known as permutation importance [20, 21], which allows ranking the importance of flight attributes by their influence on the delay time and filtering out the less-important attributes of the flight. Finally, all ML/DL models are retrained to analyse their new features by using the reduced dataset with the most important attributes selected by the models to simplify dataset gathering and ML/DL model training.

## 2.0 Literature review

The two most frequent perspectives in which the delay of commercial flights (mainly scheduled flights) is studied are its prediction and its propagation. Rich literature can be found on this second topic [22–38]. Other lines of research review the relationship of delays with (a) flight efficiency and other economic

(for example, costs) and logistical aspects of delays [12, 14, 35, 39–46]; (b) aircraft ground operations [47]; (c) airport capacity [48–50]; (d) aspects related to clients [51, 52]; and finally, (e) its generation, characterisation, behaviour and structure [10, 41, 53–68].

On the other hand, prediction of scheduled flight delays has been analysed from different methodologies that can be grouped into at least six preponderant groups [54, 57, 69–71]: (1) statistical analysis (regression models, correlation analysis, econometric models, Monte-Carlo, parametric tests, non-parametric tests and multivariate analysis [MVA]); (2) probabilistic models (which encompass or include analysis tools that estimate the probability of an event based on historical data); (3) network analysis (mainly graph theory, although studies related to Bayesian network analysis are also included in this category); (4) operations research (comprising advanced analytical methods such as optimisation, simulations and queuing theory); (5) machine learning (ML) with regression prediction (in which the most used methods include: k-nearest neighbor, recurrent NN, cascade NN, deep learning [CNN and LSTM], SVM, fuzzy logic, and Random Forests); and finally (6) machine learning in prediction in binary classification (delayed/not delayed) and multiclass (different levels of delay), in which two commonly used classification methods include decision trees and Naive Bayes. Most ensemble models, such as the Random Forest, fall into the latter category.

### 3.0 Application case

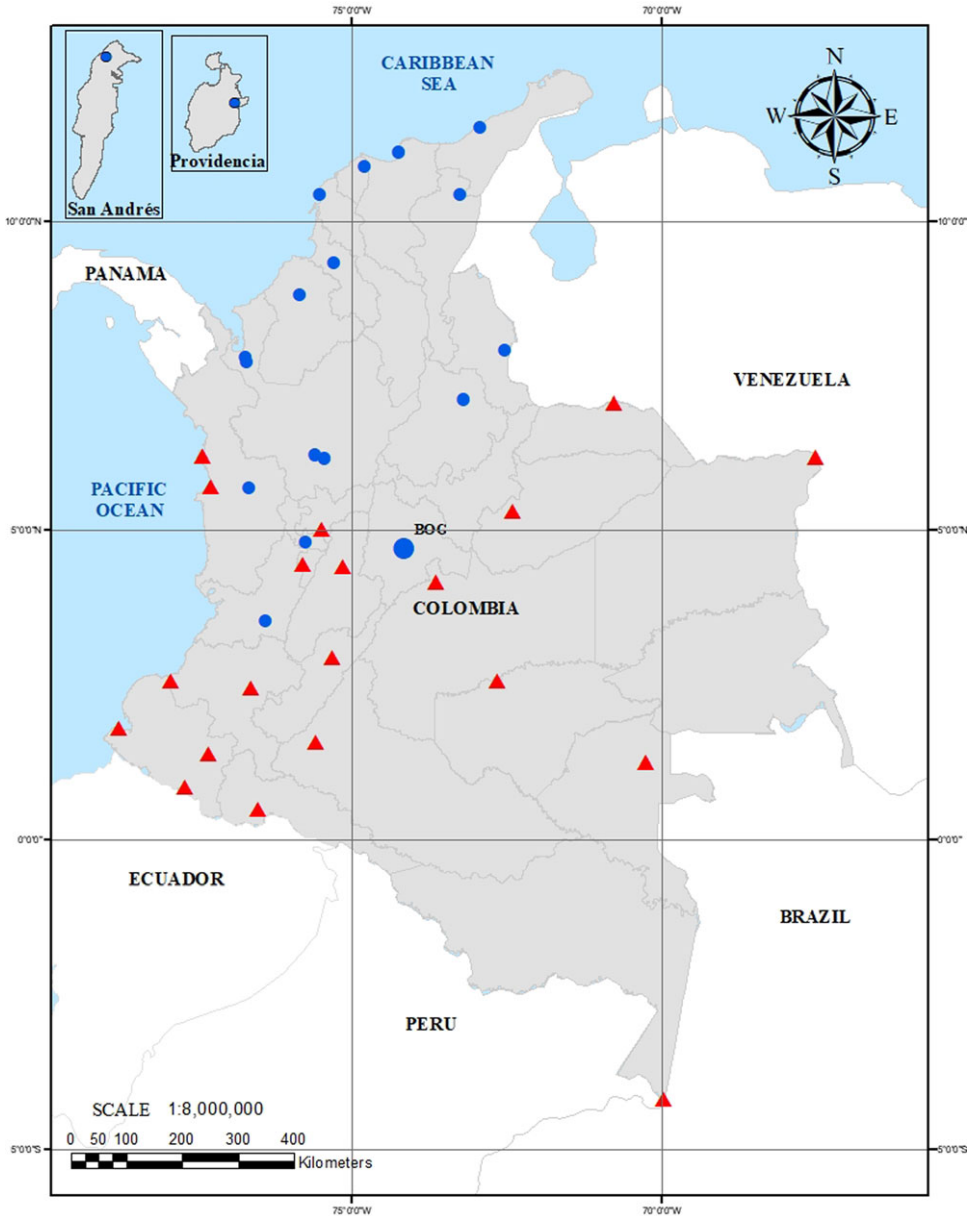
The data for the development of the present research are obtained from the country-case of application (or study) Colombia, currently the third largest air market in the Latin American subcontinent, and fifth in the Americas, by volume of traffic handled [108, 109].

In Colombia, the air transport industry was liberalised in the early 1990s. This liberalisation brought structural reforms in both the airport and airline sectors, through an uninterrupted battery of public policies (still in force today), which include not only normative and regulatory aspects, but also aggressive public and private investment programs in infrastructure and technology [110]. Regarding the management of airport infrastructure, Colombia has followed the regional trend of granting the administration of said infrastructures to the private sector. Since the mid-1990s, and in various temporary phases called generations, the Colombian government has granted concessions to 19 airports, which manage the bulk of air traffic in the entire network, including the largest and most important in the country [111]. As a result of public policies – both privatisation and public and private investment in airport infrastructure – and deregulation policies of the commercial aviation sector – airfares have been fully liberalised since 2012 [112] – since the beginning of the liberalisation of the industry in 1991 and until 2019, the total number of passengers transported grew almost 800% [17].

On the other hand, the entry into the market in Colombia of aero-private operators with a full-service business model (full-service carrier [FSC]), occurred very soon after the liberalisation of the sector (mid-1990s). But the entry into the market of low-cost airlines (low-cost carrier [LCC]) was many years after liberalisation (2012). In another order, it can be affirmed that air transport in Colombia has practically no competition (at the domestic level) with other means of transport, especially for medium and long distances, due to two determining situations: first, the complex geography of the country (crossed from southwest to northeast by three mountain ranges of the Andes mountain range), and, second, to the deficiency (in infrastructure, coverage, capacity and technology) of the terrestrial (road) and rail communication systems [113]. Figure 1 shows the geographical location of the most important airports (by volume of managed traffic) of the Colombian airport network (and participants in this research). Figure 2 shows the development of passenger air traffic in the last four decades.

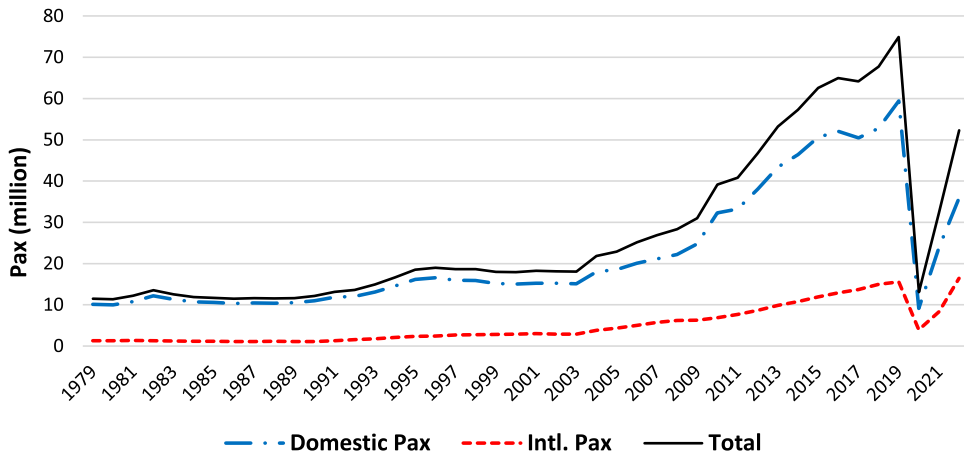
The resounding drop in air traffic verified in 2020 (see Fig. 2) is due exclusively to the complete suspension of commercial passenger air traffic (throughout the country's airport network) between mid-March and early August 2020, due to the health emergency generated by the COVID-19 pandemic. The reopening and reactivation of the commercial aviation sector was very slow, in stages and by airports, and developed throughout the last four months of 2020.

Finally, and exclusively for the present study, the Colombian Airport Authority uses the Standard IATA Delay Codes-AHM 730 [72] to encode and record the causes of all the delays that occur in the



**Figure 1.** Geographical location of the most important airports (by volume of managed traffic) of the Colombian airport network (and participants in this research). Description: blue circle, privately governed airports; red triangle, airports with public governance. BOG is the IATA code for the country’s main airport, Bogotá-El Dorado, located in the capital city of Colombia. Source: Ref. [17].

departures of all scheduled flights in the entire national network of airports open to commercial traffic every day of the year. In Colombia, the standard accepted by the national Airport Authority to measure the punctuality of the departure (or arrival) is up to 15 minutes 59 seconds (positive or negative) compared to the scheduled time. The exact moment to measure the real time of departure is when towing and push-back of the aircraft is initiated from its parking position on the apron.



**Figure 2.** Development of passenger air traffic in Colombia (complete airport network), period 1979–2022. Source: Ref. [17].

#### 4.0 Operational fundamentals of flight delays

What exactly is meant by delay, and how is it measured? Before providing an answer to this question, it is worth noting that the air transport system has several actors or stakeholders. The three most important (from an operational point of view) are: airports, airlines and the air traffic management system (ATM). Each one of these actors has its own policies and operational strategies [114, 115]. For this reason, the problem of delay cannot be attributed solely to airport operation (or management of its capacity). Furthermore, the problem of delay cannot be attributed only to the operation of these three actors, since there are other external factors that influence the appearance or aggravation of delays, one of which (the most important) is adverse weather [62, 116, 117].

On the other hand, each country in the world has its own regulatory policy (to control and monitor activity) of its civil aviation system, as well as management policies for its air transport infrastructure. All of this makes it difficult, if not impossible, to count on a universal standard that defines and regulates the concept of delay equally throughout the world [118]. Likewise, and out of necessity, each actor in the air transport system (in each country) manages the problem of delays according to their own operational strategies. This has led to the adoption of a globally coordinated approach on how to manage delay, following certain examples or initiatives. This way, delay can be scalable between airports, both locally/nationally and internationally. In addition, national regulators of commercial aviation (or control and monitoring bodies) increasingly monitor the quality of air service. It is at this point that the concept of punctuality is introduced and serves as a performance indicator [69].

Adopting recommendations or suggestions from national or supranational industry reference organisations, airport authorities (or similar bodies) in many countries have implemented the criterion that a commercial (scheduled) flight is punctual if it starts (or ends) in a range of 15 minutes (positive or negative) compared to the programmed time. If not, that flight is considered to depart or arrive late [16, 119, 120]. Some countries may apply this standard slightly differently in terms of the exact interpretation of ‘15 minutes’. That is, in some cases punctuality is accepted up to 15 minutes 59 seconds and in others, punctuality is limited to 14 minutes and 59 seconds. Early departures (sometimes called ‘negative delays’) cannot be considered a delay but depending on the timing of such anticipation they can cause certain problems in airport operations. Unexpected departures can disrupt the departure flight sequence and early arrivals can influence apron parking allocation [61].

Finally, and to complete the answer to the initial question, it remains to be determined from which event or operation the time that leads to a flight being considered punctual or delayed is counted. Again, globally, there are several criteria for when counting starts. Each national regulatory body and even

each specific airport authority defines its policy in this regard. A widely used first standard is that, for departures, the scheduled (reference) time is compared with the exact (and real) time of towing or push-back. That is, at which point the tractor or trailer starts the withdrawal of the aircraft from its parking position on the apron. Another standard is the comparison of the scheduled time with the actual off-block time (AOBT). On the other hand, the arrival delay is usually calculated as the difference between the scheduled arrival time and the actual on-block time (AIBT, literally in-block) at the parking space on apron [121–123].

As mentioned, delays can have different origins, so it is very important for airport planners and managers, as well as for national regulators of the air transport service, to know the reasons for the delays. This knowledge helps, on the one hand, airport managers to generate operational strategies for their mitigation and/or minimisation and, on the other, regulators or control agents of the air transport service to identify those responsible for the delays and the actors harmed and, where appropriate, apply suitable penalties or mandatory financial compensation (or similar) to the affected actors. For this, it has been common for many countries to adopt (or adapt) the IATA Delay Codes-AHM 730 Standard [72], in which all the possible reasons that generate a delay are specified and codified. All the possible causes related to a delay are coded there, associated (by groups) with the operations of the airline, the airport, and air traffic flow management (ATFM), as well as the meteorological and other diverse causes (security, immigration, customs, government guidelines, etc.) [72].

## 5.0 Data and methodology

### 5.1 Data

For this research, a database with departures of scheduled commercial flights, domestic and international, of the entire Colombian airport network (58 airports) for the whole year 2018 was used (of the data provided by the Colombian aeronautical authority, only delays in flight departures are available). Such database includes statistical information related to 357,595 flights, with a total of 5,721,520 data entries (each flight data contains 16 associated flight attributes). Each flight attribute is associated with the following relevant information: (a) type of traffic (domestic or international); (b) airline; (c) origin airport (always national); (d) destination airport (national or international); (e) flight number (linked to the airline, route and frequency of said flight); (f) scheduled departure date; (g) scheduled departure time; (h) towing date (push-back); (i) towing time (push-back); (j) delay expressed in fraction of an hour; (k) delay expressed in minutes; (l) flight situation; (m) departure status (according to the standard of the country's airport authority); (n) delay code (according to the Standard IATA Delay Codes-AHM 730 [72]), in which each code (89 possible) is associated with the reason that generates or causes said delay; (o) reason for the delay; (p) observations associated with the flight and/or its delay (by the airport authority).

Machine learning models require all input and output variables to be numeric. Encoding is a required pre-processing step when working with categorical data, as it is the case of our different flight's attributes. Then, all the flight attributes or inputs must be converted into number via ordinal encoders functions. In addition, data scaling is a recommended pre-processing step when working with many machine learning algorithms, as it is this case with the ten models' analysis. Unscaled target variables on regression problems can result in exploding gradients causing the learning process to fail. Also, data scaling is a recommended pre-processing step when working with deep learning neural networks as it is the case of MLP model. Therefore, the target flight attribute data called here time (flight) delay (DEM) must be standardised, that involve rescaling the distribution of data values (of DEM), then the mean of observed values is 0 and the standard deviation is 1. It is sometimes referred to as 'whitening'. This can be thought of as subtracting the mean value or centring the data. Like normalisation, standardisation can be useful, and even required in some machine learning algorithms when the input data has values with differing scales. Standardisation assumes the data fit a Gaussian distribution with a well-behaved mean and standard deviation, as it is in this case. But even this standardised Gaussian pre-processing it is recommended to be applied, if this expectation is not met, to get more reliable results.



## 5.2 Typology of implemented models

ML/DL models mathematically represent some types of universal approximators [73] of the function  $f$ , such that it fulfills:  $Y = f(X)$ . So, in the present work ten ML/DL models are proposed and used to map the  $Y$  output (delay time) with a finite and arbitrary (available attributes) set of  $X$  inputs (flight attributes, which can be up to ten in its most extensive version of the dataset), although of these ten models dare defined in Table 1, only a single DL-type model is developed and used, the so-called multi-layer perceptron (MLP) [74].

The MLP model, whose architecture is shown in Fig. 1, has been generated using the keras/tensorflow libraries, which consist of two dense layers (connections of all neurons in the previous layer with all neurons in the posterior layer) of 9 neurons in the first layer and 36 neurons in the second, using the activation function ReLU (rectified linear unit) [75–78], so that neurons can learn the non-linear relationships between the output and the input (of each neuron). The ReLU function is defined as:  $y = \max(0, x)$ , in which  $y$  represents the neuron output and  $x$  the neuron input. The ReLU function has several advantages: its computation is easy and, therefore, training times are short; it also converges quickly and does not present saturations (a problem that does occur when other activation functions such as sigmoid and hyperbolic tangent are used by having a linear output); then ReLU neurons activation do not present the vanishing gradient problem (which prevent changing or updating the model weights because the derivatives of weights sensitivity are nearly zero); they are sparsely activated (sparsity feature), which simplifies the existing connectivity or couplings between all neurons; and, finally, allows the ML/DL model to learn the non-linear relationships between the input  $x$  and the output  $y$ .

The rest of the models developed in this work, and referred in the Table 1, are the following: the support vector machine (SVR) model, a very robust model in both linear and non-linear predictions because the dimensional space of the features has been increased [79]; the Random Forest (RFR) model, which is an ensemble-type method that improves the results of any of its simplest constituent algorithms by combinations of these [21, 80]; the simple linear regression (LIR) model [81]; the bagging gradient (BGR) model, which improves predictions by producing multigroup by combinations and repetitions [82–84]; the extreme gradient boosting (XGR) model, which from weaker models generates stronger models using the descending gradient as an optimising method [85, 86]; the Huber Regressor (HBR) model, which is also very robust in training the data with outliers (or atypical values) by penalising them through *ad hoc* cost functions [87, 88]; the extra trees (ETR) model, which implements many more decision trees to better fit learning than simple decision tree (DTR) methods [89, 90]; the DTR model, which is the simple model of decision trees, which is reused due to its self-explanatory capacity of the patterns learned by the ML [91]; the k-nearest neighbors (KNN) model, which predicts closest neighbours based on its K parameter [92]. The ‘R’ at the end of model names indicates that the model has been implemented for regression analysis. Table 1 lists, at the summary level, the models used.

## 5.3 Training ML/DL models

Training the models is part of the core ML process. In this work, each of the ten ML/DL models has been trained separately with the dataset under four possible scenarios: the first scenario with the ten full flight attributes (9 inputs + 1 output), the second scenario reducing scenario one to only six flight attributes (5 inputs + 1 output), the third scenario considering only flight attributes that could be known in advance (5 inputs + 1 output), and finally the fourth scenario reducing the third one to only four flight attributes (3 inputs + 1 output) (see Fig. 3). The reduced scenarios are aimed at evaluating whether the initial flight attributes of the dataset can be reduced. Another aspect to highlight from this paper is having carried out the statistical model training by implementing the k-fold and cross-validation techniques, to ensure that there are no biases in the training due to having chosen special subsets from the dataset for training/test, resulting in a more robust evaluation of the performance of ML/DL models.

The training has been implemented by dividing the dataset into batches of 128 samples each, subsequently, the ML/DL model processes the data forward (from the inputs to the output layer), to predict or estimate the model output values. At this point, with the error obtained by measuring the predicted value

**Table 1.** *Models used in this research*

Model	Acronym	Main characteristic
Multi-layer perceptron	MLP	Supervised machine learning model consisting of fully connected neurons of feedforward artificial neural network (ANN) for regression analysis of deep learning model type.
Support vector machine	SVR	Supervised machine learning model based on constructing a set of hyperplanes in a high dimensional space searching to maximise different data distance for regression analysis.
Random Forest	RFR	Supervised machine learning model as an ensemble learning method that operates by constructing a multitude of decision trees as conditional for regression analysis.
Simple linear regression	LIR	Supervised machine learning model approaching linearly the relationship between a scalar response and several explanatory variables (also known as independent variables) for regression analysis.
Bagging gradient	BGR	Supervised machine learning model for regression by bootstrap aggregation, also known as bagging, is an ensemble method works by training multiple models independently and combining later (using some deterministic averaging method) resulting in a strong model. This final model is more robust than those individual weak learners' components.
Extreme gradient boosting	XGR	Supervised machine learning model for regression is an ensemble constructed from weak decision trees model. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. It is scalable.
Huber Regressor	HBR	Supervised machine learning model for a robust regression using the Huber loss function that is less sensitive to outliers in data, via penalising the outlier data training and learning.
Extra trees	ETR	Supervised machine learning model for regression consisting of extremely randomised trees, or extra trees for short, is an ensemble method that combines the predictions from many unpruned decision trees. Predictions are made by averaging the prediction of the decision trees.
Decision trees	DTR	Supervised machine learning model for regression, based on decision support hierarchical models that uses a tree-like model of decision and their possible consequences, including chance event outcomes, resource costs and utility. It is one way to display an algorithm that only contains conditional control statements.
K-nearest neighbours	KNN	Supervised machine learning model for regression is a non-parametric method (not making any assumption on underlying data) and one of the simplest algorithms. The target or output is predicted by local interpolation of the targets associated of the KNN in the training set.

versus the target value, the training method goes backward (from the output layer to the input layer) to update the model weights (or connections between neurons) of all neurons of all layers of the model, through the backpropagation method [93–95]. This method of backward propagation of the error starts obtaining the total error through the appropriate defined cost function. Here, the root-mean-square-error (RMSE) applied metric results from comparing the values (of the batch) estimated by the model, with



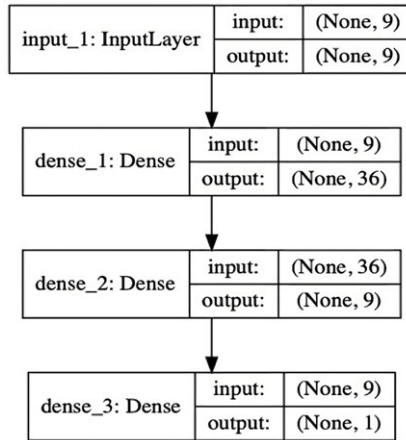


Figure 3. Layers and neurons type architecture of the developed MLP model. Source: authors.

the real or target values collected from the known supervised features of the dataset. This core process of ML can be explained as the process distribution of the total error between the different contributions made by all the weights or nodes of the model, which is conducted in turn by the stochastic gradient descent (SGD) method [96, 97]. The SGD method calculates the partial derivatives of the error versus each of the weights, which is a way of measuring the model sensitivity of the error compared to each of the weights of the artificial neural network. The adjective ‘stochastic’ of the SGD method comes from the fact that the gradient is calculated by randomly chosen batches and not for the entire dataset.

The equations corresponding to MLP [74, 98] are those associated with neuron layers, in which a first layer corresponds to the input (represented by the  $X$  tensor), and the last output layer (associated with the  $Y$  tensor). Between them, the hidden layers to be implemented are inserted. Each layer contains a variable number of neurons (or nodes) that can be activated with the activation functions, which can be of various types [99]. The input and output a neuron are shown schematically in Fig. 4.

In Fig. 4,  $x_i$  represents the inputs of feature  $i$  of the input layer,  $w_{ij}$  represents the weight of feature  $i$  of the previous layer in connection with neuron  $j$  of the posterior layer,  $\Sigma$  represents the transfer function (sum of all the inputs coming from the outputs of all the neurons of the previous layer with neuron  $j$  corrected with the  $b_j$  bias of layer  $j$ , which produces the net output (transfer function + bias)  $Net_j$ . The chosen activation function  $\sigma$  is then applied to each neuron (for example, *tanh* or *sigmoid*, although *ReLU* has been chosen here), which in turn produces the output  $O_j$ . From here, the connections of the output of this neuron are established again, together with all the neurons of its layer, with the neurons of the next layer, until reaching the last output layer  $Y_j$  (which can be several neurons of output, although in the case of regression prediction it is constituted by a single neuron). This inflow-outflow is called ‘forward flow’ or ‘propagation’ (from the input layer through all the neurons in the intermediate layers, to the final output layer).  $\theta_j$  is the threshold value of neuron  $j$ , to activate its output function (if it exists).

In this work, the RMSE has been used as the cost function. Equation (1) defines this error with the norm 2  $\| \cdot \|$ , in which  $E$  is the error, and  $n$  is the number of samples in the dataset,  $y(x)$  is the actual output (provided by the dataset under supervised learning method) and  $\hat{y}(x)$  is the output estimated by the model.

$$E = \sqrt{\frac{1}{n} \sum_x \|(y(x) - \hat{y}(x))\|^2} \tag{1}$$

In ML, once the cost function is obtained, the fundamental process for the automatic adjustment of the weights of the model is implemented via BP [93]. Such method calculates the gradient of said total error (in this case the RMSE) compared to each of the weights or connections of the neurons of each

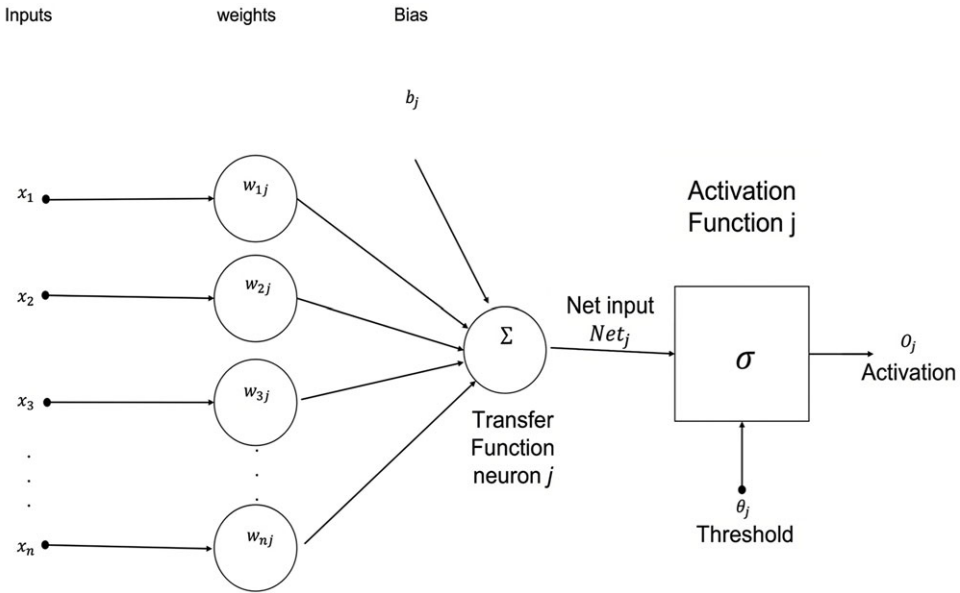


Figure 4. Neuron connections of MLP networks. Source: Ref. [100].

layer of the model, for which the rule of chain derivation applies, to propagate said error backward or to distribute the error between all the model weights, neuron by neuron, layer by layer. If the method converges, the new adjustment of weights of the model is sought to produce a smaller error when it is iterated with the next batch of input samples  $X$ . This is the essence of ML, minimising some loss function (cost or error), adjusting the contributions of the weights of the model neurons connections per each dataset batch and repeating the BP process adjustment to all dataset iterations implemented. Learning that is sufficiently guaranteed when using the SGD method by the theorems of the existence of the null derivation at the points of minimum error. The following group of Equations (2) shows the derivation chain rule in artificial neuron networks (ANN) [101, 102].

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \tag{2a}$$

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \left( \sum_{k=1}^n w_{kj} o_k \right) = \frac{\partial}{\partial w_{ij}} w_{ij} o_i = o_i \tag{2b}$$

The problems of searching the cost function minimums are solved through different optimisers, generically known as the SGD [99]. These optimisers are responsible for avoiding the solution from getting trapped in any of the local minimums that the cost function presents when it is not a convex function, for which each of them has implemented different strategies to find the global optimum. In the present work, the adaptive moment estimation (ADAM) optimiser has been used because it is one of the most efficient optimisers in terms of its high convergence speed and robust stability [100], which is an improved version of the RMSProp that implements a variable/adaptive speed (or learning step) and a moment or inertia that tries to avoid getting caught in the local minima of the cost function. The following group of Equations (3) describes the update of the weights by the ADAM optimiser [100].

$$m_w^{(t+1)} \leftarrow \beta_1 m_w + (1 - \beta_1) \nabla_w l^{(t)} \tag{3a}$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1 - \beta_2) (\nabla_w l^{(t)})^2 \tag{3b}$$

**Table 2.** List of flight attributes, and their participation in the different scenarios of the model

Flight attribute	Description	Scenario in which the attribute participates
TRF	Type of traffic (domestic/international)	1, 3
AEL	Airline	1, 2, 3, 4
ORI	Origin airport name (departure)	1, 2, 3, 4
DES	Destination airport (name)	1, 3
NVU	Flight number	1, 2, 3, 4
FPS	Scheduled departure date	does not participate
HPS	Scheduled departure time	does not participate
FRM	Push-back date	does not participate
HRM	Push-back time	does not participate
DEA	Time delay (according to Standard IATA Delay Codes-AHM 730)	does not participate
DEM	Actual time delay (in minute)	1, 2, 3, 4
EVU	Flight status (can be on time or delayed)	does not participate
CDE	Delay code (according to Standard IATA Delay Codes-AHM 730)	1, 2
MDE	Reason for delay (according to Standard IATA Delay Codes-AHM 730)	1
OBS	Observations (from the airport authority)	1
EAC	Flight status (can be on time, delayed, early or canceled)	1, 2

$$\hat{m}_w = \frac{m_w^{(t+1)}}{1 + \beta_1^{(t+1)}} \tag{3c}$$

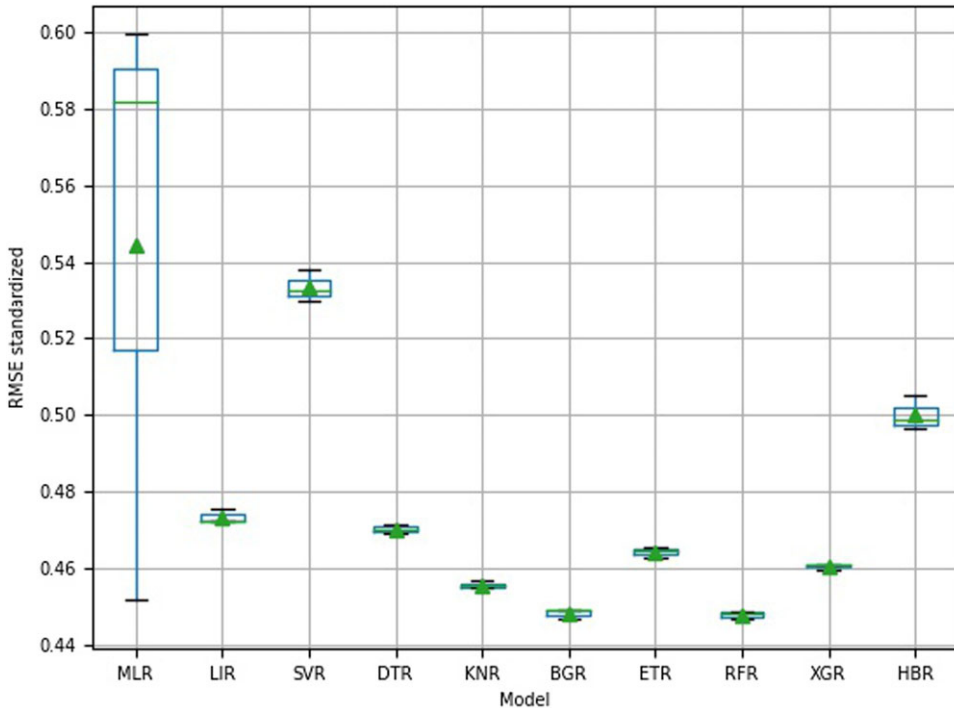
$$\hat{v}_w = \frac{v_w^{(t+1)}}{1 + \beta_2^{(t+1)}} \tag{3d}$$

$$w^{(t+1)} \leftarrow w^t - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w + \epsilon}} \tag{3e}$$

In the Equations (3),  $w^{(t+1)}$  represents the weights in iteration  $t + 1$ ;  $\eta$  represents learning speed or learning step;  $\nabla_w l^{(t)}$  represents the gradient of the cost function  $l$  with respect to the weights;  $\epsilon$  is a minimum value to avoid dividing by 0 (in keras  $10^{-7}$  is used);  $\beta_1$  y  $\beta_2$  are the forgetting factors for the gradient or first moment and the second moment respectively;  $m_w^{(t+1)}$  is the moving mean of the weights or first moments in the iteration  $t + 1$ ;  $v_w^{(t+1)}$  is the moving variance (or second moment) of the weights in the iteration  $t + 1$ . The variables with *hat* indicate that they are estimators and the superscripts with the variable  $t$  correspond to each of the iterations (or learning cycles). This is the process of fitting the MLP/DL model. On the other hand, in models based on decision trees, the adjustment is made mainly by induction and pruning methods of the decision trees [101, 102].

The dataset learning process is repeated in this work 150 times (iterations) for the entire dataset and only for the MLP/DL model (considered sufficient to achieve an asymptotic or stable RMSE value), which marks the maximum fine adjustment possible of the model in the learning of the dataset, or until a sufficient adjustment of the error is obtained for the rest of the non-DL models, characterised by belonging to the ensemble or boosting or stacking type, in which they improve the simplest models of decision trees by their appropriate combinations [82].

To give more robustness to the value of the metrics obtained from the RMSE, k-folds techniques have been implemented [103, 104]. Specifically, three groups have been applied to divide the dataset between

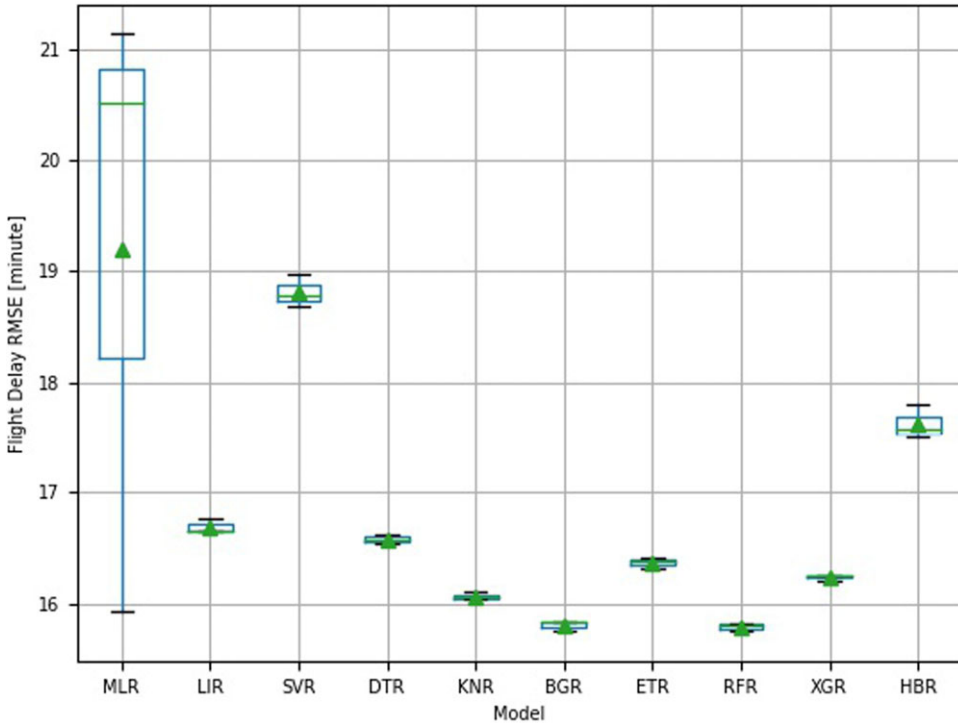


**Figure 5.** RMSE comparison resulting from the learning obtained for each of the ten models, transforming the delay time into a standard distribution (0 mean and 1 for standard deviation), referred to scenario 1 resulting from choosing ten flight attributes, nine inputs (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) and one output (DEM). Source: authors.

three different ways of grouping the training and test data subsets. The training data, as its name implies, is used to train the model and the test data to evaluate the learning performance of the model. However, the model can never see the test data during the learning process to make an objective evaluation. In addition, with cross-validation techniques this training is repeated a variable number of times. Because target delay flight attribute was standardised as necessary step of data pre-processing, the loss or cost function RMSE of each ML/DL model it is also characterised statistically, in terms of a mean and a standard deviation, resulting all of this in a more robust way performance measured compared to the case when a single configuration of training and test subset is used. It should be emphasised that all these methods are stochastic, due both to the use of the SGD method, and to the own stochastic nature of the dataset, so their results are not deterministic, but rather are probabilistic.

#### 5.4 Statistical examination

Before training, different examinations or statistical controls were applied to have a feature insight. For this purpose, different tests has been implemented to determine if flight attribute distributions are Gaussians (or normal) (Shapiro-Wilk, D'Agostino and Anderson-Darling tests), or whether flight attributes are not correlated with the output (linear Pearson, Spearman Rank, Kendall Rank, and chi-squared tests), or whether the flight attributes are not stationary (ADFfuller test), or whether flight attributes have equal distribution as the output (student's test, paired student, ANOVA, Mann-Whitney U, Kruskal-Wallis H, and Friedman tests). These tests, of statistical characterisation of flight attributes, serve to know amongst other things, whether their learning process will be easy or which



**Figure 6.** Comparative analysis of the ten models’ performance (RMSE), in its original scale of minutes of delay, referred to as scenario 1, choosing ten flight attributes, nine inputs (TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC) and one output (DEM). Source: authors.

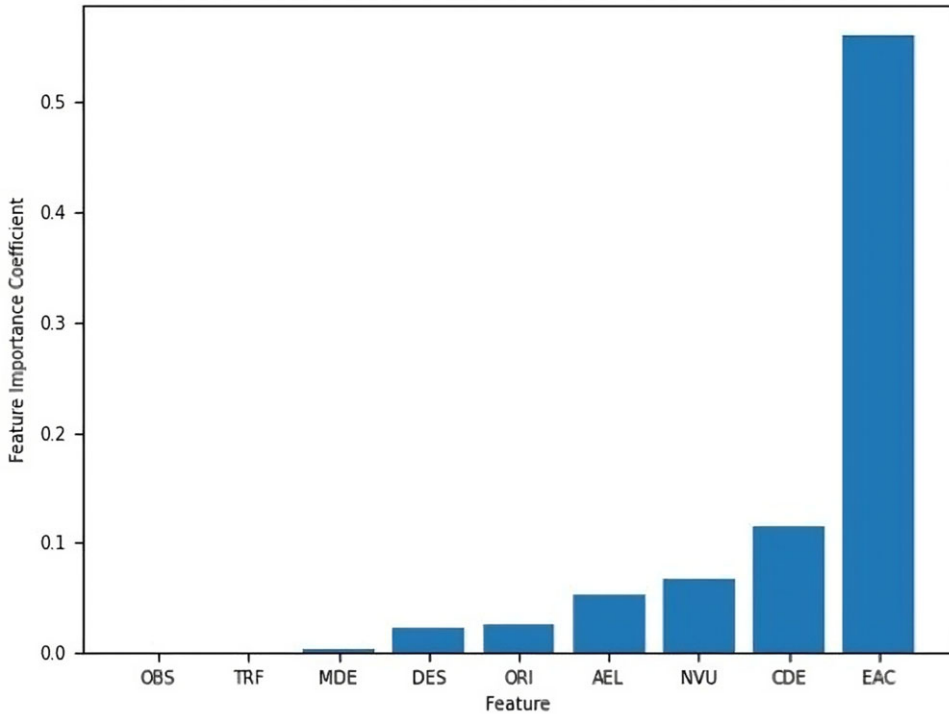
data transformation must be required to facilitate it. These techniques are an enhancement of dataset preprocessing to facilitate ML training.

### 5.5 Determination of the importance of the flight attributes

To rank the importance of flight attributes in determining flight delay time, the permutation importance (PI) method [20, 21] has been implemented. This method consists of applying inspection techniques to trainable estimators; each model proposed and configured for this work offers a different response regarding the selection of the most significant attributes of the flights. The PI method consists of randomly permuting the features among themselves and measuring in each case how the output (delay time) varies through the training of the new dataset (permuted or sometimes called corrupted in the scientific literature) with each ML/DL model, which allows establishing the coefficient of importance or the sensitivity of the output (time delay) to each feature  $j$  called  $c_j$  (see Equation (4)).

$$c_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j} \tag{4}$$

In Equation (4)  $s$  is the result of training the ML model when it is trained with the  $D$  dataset without permuting. Next, each feature  $j$  associated with a column of dataset  $D$  is randomly permuted with the other features to produce a permuted version of the dataset  $\tilde{D}(j, k)$ , in which the subscript  $k$  represents the repetitions used to manage the features,  $K$  is the maximum number of permutations, and  $s_{k,j}$  is the new result of training the ML model for each feature  $j$  repeated  $k$  times with the permuted (or corrupted) dataset  $\tilde{D}(j, k)$ .



**Figure 7.** Relative importance coefficients in increasing order, of the nine full flight attributes considered as inputs of scenario 1, when determining the flight delay time (DEM) as output. Source: authors.

## 6.0 Results

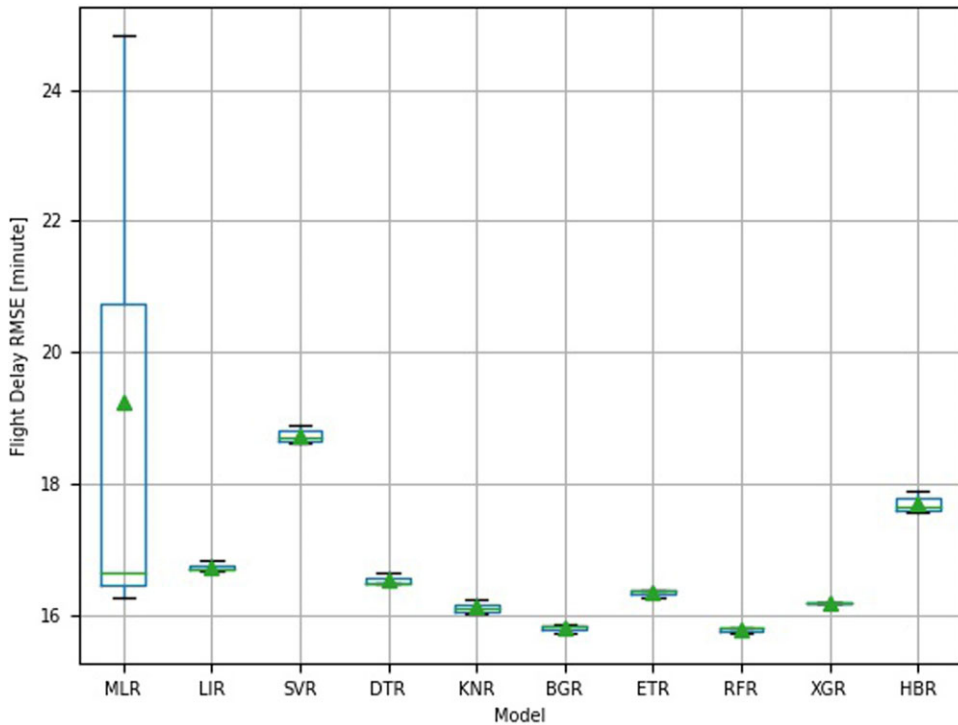
Before training the ten ML/DL models implemented, the outliers (or atypical data values) are removed from the dataset. If they are retained, the models are forced to learn marginal patterns of said flight attributes, which is detrimental to a good generalisation in the prediction of the flight delay time. In this work, four methods have been used to identify and remove outliers, two of which are properly unsupervised ML methods, such as the local outlier factor (LOF) and one class SVM (OCS) methods, which cluster normal data and mark as outliers those that are outside these perimeters of normal values. The third method is Isolation Forest (IF), which applies binary decision tree theory that does not require pre-grouping within standard profiles, making it faster and more efficient in detecting outliers. The fourth method, the elliptic envelope (EE) [105], calculates the outliers by the method of the minimum covariance determinant, which must meet the applicability condition that the number of samples (flights) must be higher than the square of flight attributes, which is true in this case.

Subsequently, the learning of the ten ML/DL models has been evaluated for four flight attributes selection scenarios, to show the prediction sensitivity of the models depending on which flight attributes are chosen. Table 2 shows the list of flight attributes and indicates in which model scenario each flight attribute participates.

The models' comparative results of the RMSE performances are shown for full scenario 1. This full scenario 1 removes 6 of the 16 initial flight attributes (because some of the them are redundant), resulting in the ten full following flight attributes: TRF, AEL, ORI, DES, NVU, CDE, MDE, OBS, EAC and DEM, in which the first 9 are inputs and the last one is the output to be predicted (DEM is the delay time measured in minutes) (Fig. 5).

Figure 5 shows the performances, measured with the RMSE metric, are very similar for all models and their value is below 0.5, except for the behaviour of two more dispersed models (MLR and SVR).





**Figure 8.** Comparative results of the RMSE obtained by the 10 ML/DL models when trained under reduced scenario 2, with only five flight attributes. It can be seen how the model learning performance (in terms of RMSE) is roughly identical to full scenario 1. Source: authors.

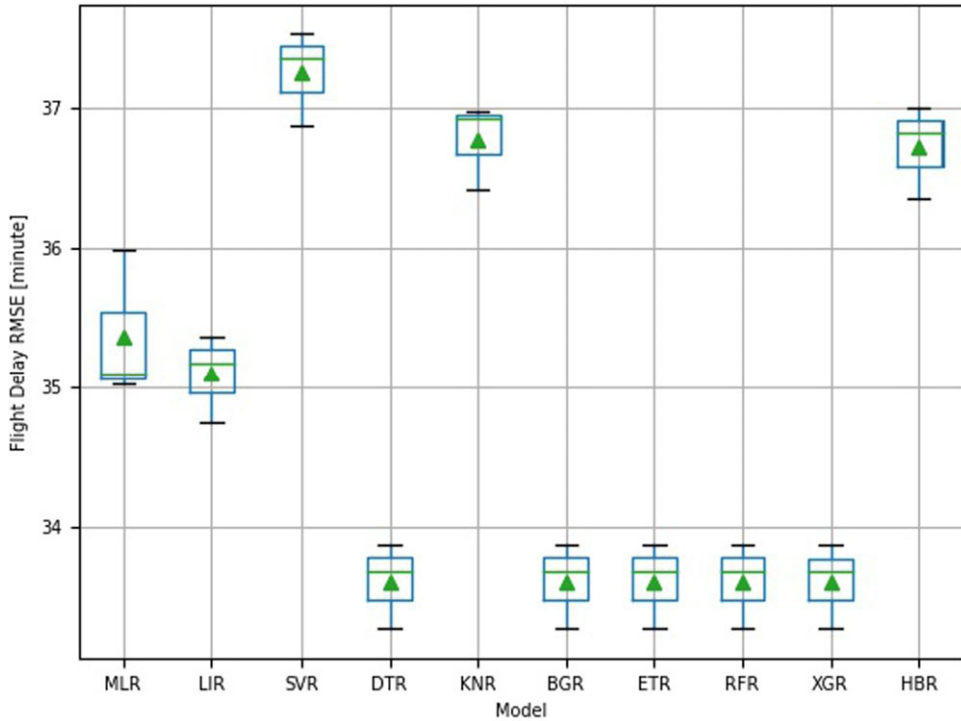
The RFR model is the one with the highest performance with an RMSE of 0.449 and a standard deviation value of 0.0013, which evidence that the model is robust.

It is already explained on subsection 5.1 (Data), that the DEM flight attribute has been standardised to facilitate the learning of the ML/DL models. For this reason, Fig. 6 shows the results of the ML models performance (measured with the RMSE metric), but in its original scale of delay times in minutes (which is done by reversing the standardisation process applied to the DEM), and evidence that the prediction errors (on their natural minute scale) are less than 16 minutes (for the best RFR model). That is, when introducing the nine flight attributes, a delay time is predicted that has an error of fewer than 16 minutes, which is a remarkable result for the features considered.

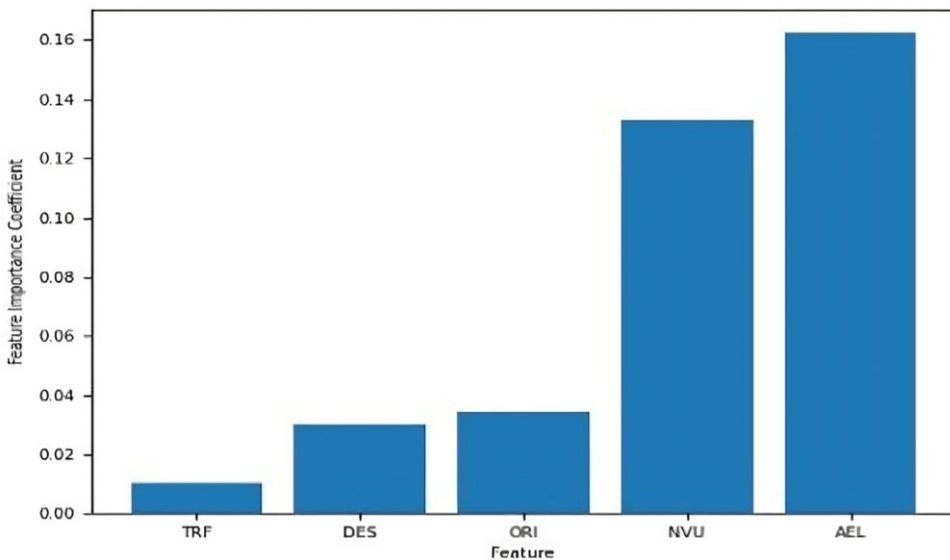
Next, reduced scenario 2 is generated, after applying PI method whereby each of the ten ML/DL models auto-selects the five best flight attributes that better determine the delay time, from among the nine full flight attributes inputs of the scenario (the tenth flight attribute is the output). That is, each model ranks the flight attributes inputs by their importance in determining the delay time, assigning them a (relative) importance coefficient.

Figure 7 shows the ranking of importance assigned by the RFR model to the nine input features, in which the last five are the most important (ORI, AEL, NVU, CDE, EAC) in increasing order.

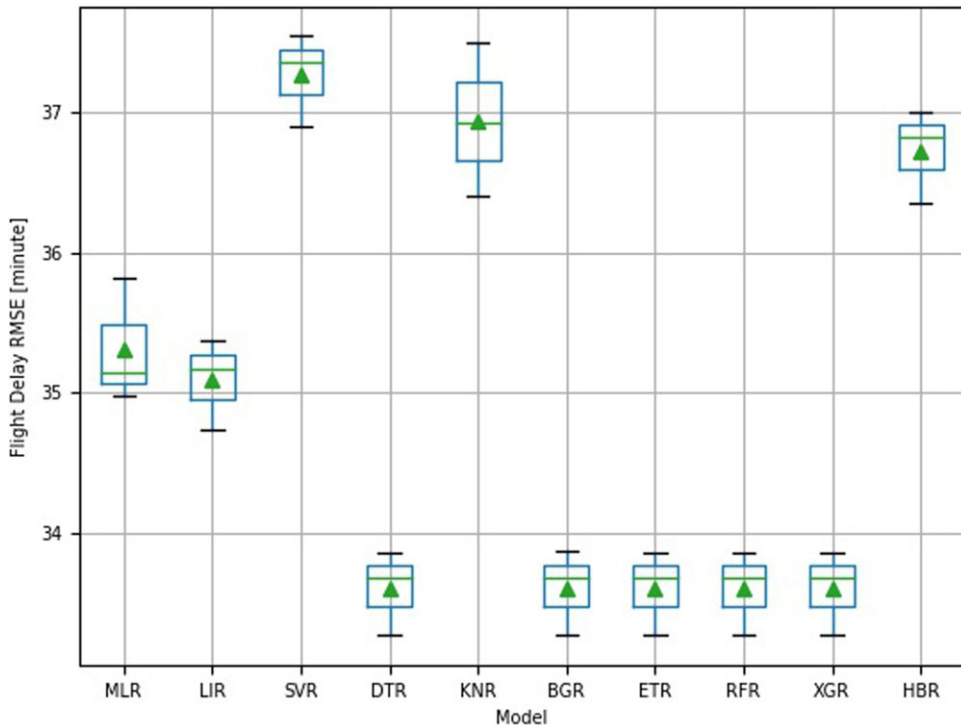
To evaluate the quality selection of the five most important or significant flight attributes (EAC, CDE, NVU, AEL, ORI), now arranged in descending order, the models are retrained with this new reduced scenario 2, considering as output the flight attribute DEM, to compare the learning models performance between the two scenarios. The results of the RMSE obtained in this reduced scenario 2 are shown in Fig. 8, where it can be seen they roughly get the same results when using the nine input flight attributes of full scenario 1. In fact, even the RMSE improves slightly for reduced scenario 2, which



**Figure 9.** RMSE comparative analysis of the 10 models under a more realistic dataset scenario 3 (five inputs: TRF, AEL, ORI, DES, NVU and one output: predicting delay (DEM)), in which the best RFR model obtains an RMSE in predicting flight delay of 33.8 minutes. Source: authors.



**Figure 10.** Rank of the importance of the five input flight attributes considered in scenario 3 (TRF, DES, ORI, NVU, AEL) in increasing order. Source: authors.



**Figure 11.** RMSE comparison of models for reduced scenario 4, with three inputs (ORI, NVU, AEL) and DEM as time delay to be predicted, in which the best RFR model obtains the same RMSE of the prediction of flight delay of 33.8 minutes than on scenario 3. Source: authors.

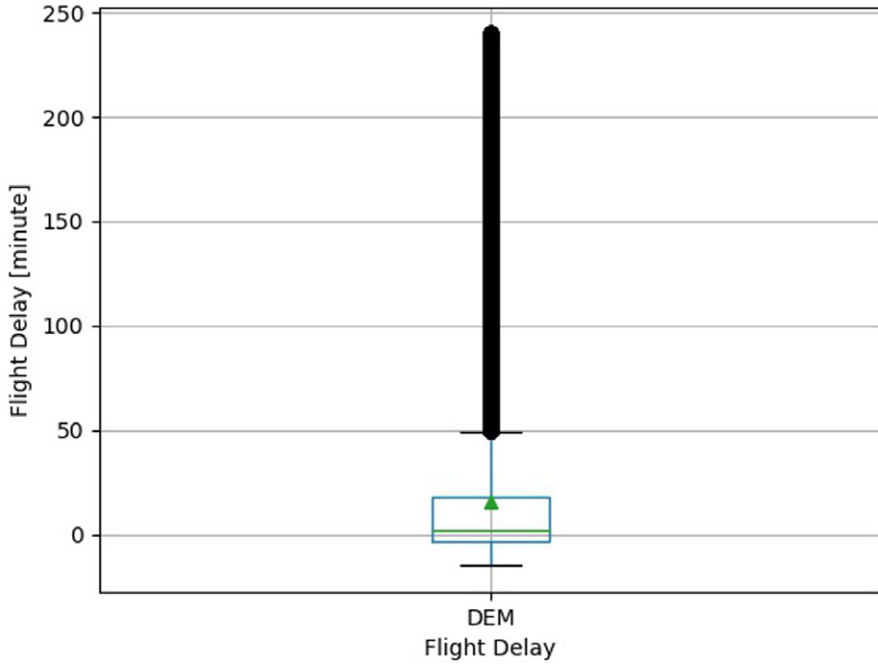
shows that the learning patterns are better grasped by the models when they are trained on this reduced scenario 2 (using only the five more significant flight attributes) than when they are trained using the full scenario 1.

Since both scenarios full 1 and reduced 2 involve a series of flight attributes that are measured *a posteriori* and, therefore, are not known in advance, such as the EAC (flight status according to airport authority standard), CDE (delay code according to Standard IATA Delay Codes-AHM 730), MDE (reason for delay according to Standard IATA Delay Codes-AHM 730) and OBS (observation of flight from the airport authority), in this paper, the four aforementioned flight attributes from the dataset have been removed to make the predictions of flight delay departures more realistic (according to real data known before the flight departure). Particularly, a new realistic scenario 3 is now defined with only six total flight attributes, five of which are inputs (TRF, AEL, ORI, DES, NVU) and one output (DEM).

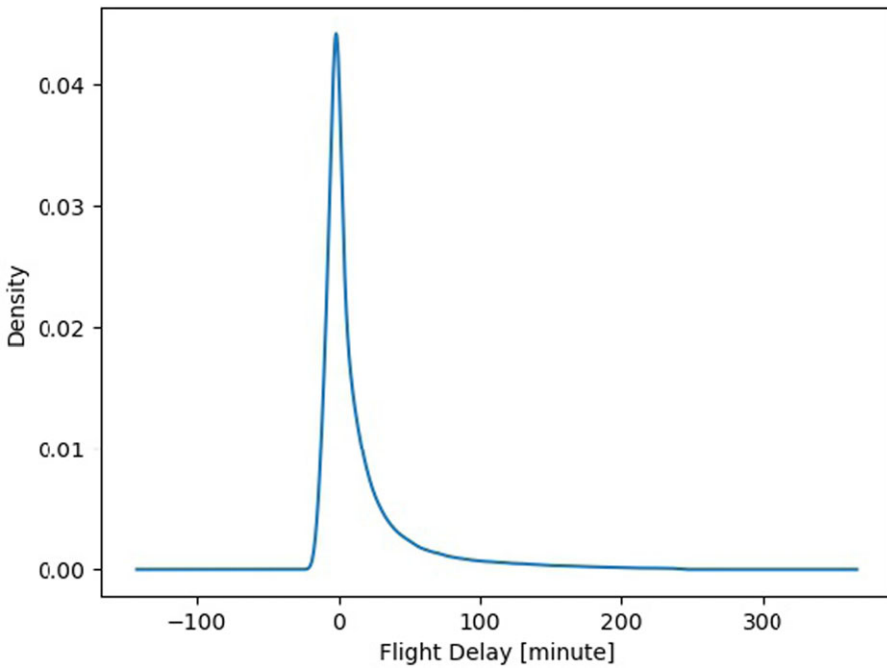
After models have been conveniently trained, an RMSE of 33.8 minutes is obtained by the best RFR model (approximately double that when the four eliminated flight attributes were retained), results are shown in Fig. 9.

The reason for this loss of precision when eliminating these four keys flight attributes (EAC, CDE, MDE, OBS) is that other variables that are not registered in the dataset may influence the level of delay, such as the type of aircraft (mainly its size), the level of congestion at the airport, weather conditions, ATM conditions, etc. When PI method is applied to scenario 3 of six flight attributes (5 inputs +1 output) to get the importance coefficients of the flight attributes in determining the flight delay time, the significance ranking is shown in Fig. 10 (flight attributes selected by the best RFR model).

Reduced scenario 4 (similar process to get the reduce scenario 2) is generated by applying reduction to scenario 3, to keep only the best three flight attributes that auto select the best (RFR) models when applying the PI method.



**Figure 12.** Time distribution of the flight departure delay. Source: authors.



**Figure 13.** Unimodal distribution (density) function of a flight delay (at departure). Source: authors.

Figure 11 shows the comparison of the RMSE for the 10 ML/DL models of scenario 4 consisting of four flight attributes (three inputs: ORI, NVU, AEL, and one output: DEM). The results show that, despite the reduction of features (to only the three most important), the same results are obtained as in scenario 3 when five inputs (flight attributes) are used. This shows that the feature auto-reductions conducted by the PI method have been carried out correctly and, most importantly, the dataset can significantly be reduced or simplified, thereby saving model training time and dataset memory, without any loss in the accuracy of the time delay predictability (DEM).

Figure 12 shows the distribution of the flight delay time (DEM) dataset flight attribute in terms of minimum (-15 minutes) and maximum (240 minutes), with the quantiles 25% and 75% and mean (15 minutes).

Finally, Fig. 13 shows that the continuous distribution function (designated as Density in the figure) of the flight delay time is unimodal, corresponding to the 15 minutes mean of the distribution.

In summary, it can be stated that with the specific dataset flight attributes provided to the models for training purposes, the predictions of the delay time made by the ML/DL models are acceptable. Depending on the scenario considered, such models include error ranges between 16 and 33 minutes, which implies acceptable range values that are similar or even better than those obtained in the related research literature [70, 71, 106, 107].

## 7.0 Conclusions

As the periodic reports of the main international civil aviation organisations (IATA, ACI, ICAO, etc.) show, the demand for air transport has evidenced continuous and relevant growth for almost three decades. Airport authorities around the world are trying to respond to this demand by expanding air infrastructure, building new airports, optimising the use of existing capacity, improving ATM procedures, etc. Even so, many airports (mainly large and international ones) continue to present congestion (often severe in certain periods of the day), a situation that inevitably leads to delays, which can be aggravated by adverse weather. Therefore, the most accurate prediction methods have become indispensable, as they contribute to formulating strategies (efficient use of capacity, better management of operations, optimisation of ATM procedures, etc.) to mitigate the proliferation of delays and prevent their spread into other airports in the network.

Therefore, the present research proposes a comparative analysis of the performance (measured in RMSE) of flight time delay prediction (at the departure of a flight) of ten ML/DL models under different scenarios configured by the known flight attributes registered on the dataset for each flight. The ML models are in their most advanced variants (combinatorial models), such as ensemble techniques in their different methodologies such as voting, bagging algorithms (e.g., segmenting the dataset to obtain the advantages of training each model with different subsets), stacking (using outputs from some models to train others) and boosting (specifically penalising unwanted errors). All these strategies serve to compensate or balance the errors of each model, so that the ensemble ML models focus on generalising better the learning process. The DL/MLP model has also been included in the comparison (although it was not the winning model because it was not adjusted in its design parameters and hyperparameters). The comparative RMSE results for the transformed time delay (using standardised output data) are of the order of 0.45 and 0.96 (for the winning RFR model) depending on the flight attributes, or scenarios considered in the training process, which translates into (acceptable) prediction range errors between 16 and 33 minutes in time delay flights departures.

On the other hand, in this work the PI method is implemented, which allows the input flight attribute to be auto ranked by their importance in determining the time delay by each model. Said flight attribute selection is re-evaluated on the corresponding reduced dataset scenarios, verifying that said auto reductions do not affect the precision (RMSE) of the time delay forecast and, in addition, allow reducing the size of the dataset.

Finally, although for the study case implemented here there is an extensive database of flight departures for a whole year associated with the entire national airport network and with many flight attributes

(linked to each flight), there are constraints that prevent better ML performance. That is, certain important flight attributes are missing (e.g., weather conditions, aircraft size, airport congestion level, ATM condition, runway system capacity, aircraft turnaround time, etc.), that if included in the dataset could significantly improve the time delay ML prediction. Another possible line of research would be to compare the behaviour of the prediction models developed here for the current case study (Colombian air transport system and using the same flight attributes) with another case study (air transport system of another country) with the same or similar characteristics.

## References

- [1] ICAO. *Presentation of 2019 Air Transport Statistical Results*, ICAO, 2019, Montreal.
- [2] IATA. *World Air Transport Statistics*, IATA, 2021, Geneva.
- [3] Graham, A. and Morrell, P. *Airport Finance and Investment in the Global Economy*, Routledge, 2017, London.
- [4] de Neufville, R. and Odoni, A. *Airport Systems, Planning, Design, and Management*, McGraw-Hill, 2013, New York.
- [5] Horonjeff, R., McKelvey, F., Sproule, W. and Young, S. *Planning and Design of Airports*, McGraw-Hill, 2010, New York.
- [6] Ashford, N., Mumayiz, S. and Wright, P. *Airport Engineering*, John Wiley, 2011, New Jersey.
- [7] ACRP. *Defining and Measuring Aircraft Delay and Airport Capacity Thresholds. ACRP Report 104*, Transportation Research Board, 2014, Washington, D.C.
- [8] Gelhausen, M., Berster, P. and Wilken, D. *Airport Capacity Constraints and Strategies for Mitigation*, Academic Press, 2020, London.
- [9] Janic, M. *Airport Analysis, Planning and Design: Demand, Capacity and Congestion*, Nova Science Publishers, 2009, New York.
- [10] Borsky, S. and Unterberger, C. Bad weather and flight delays: The impact of sudden and slow onset weather events, *Econ. Transp.*, 2019, **18**, pp 10–26. DOI: [10.1016/j.ecotra.2019.02.002](https://doi.org/10.1016/j.ecotra.2019.02.002)
- [11] University of Westminster. *European Airline Delay Cost Reference Values*, University of Westminster, 2015, London.
- [12] Peterson, E., Neels, K., Barczy, N. and Graham, T. The economic cost of airline flight delay, *J. Transport Econ. Policy*, 2013, **47**, (1), pp 107–121.
- [13] NEXTOR. *Total Delay Impact Study*, National Center of Excellence for Aviation Operations Research (Federal Aviation Administration), Washington D.C., 2010.
- [14] Cook, A., Tanner, G., Jovanović, R. and Lawes, A. The cost of delay to air transport in Europe - Quantification and management, 13th Air Transport Research Society World Conference, June 27–30, Abu Dhabi, 2009.
- [15] BTS. Bureau of Transportation Statistic (U.S. Department of Transportation), 2021. <https://www.transtats.bts.gov/HomeDrillChart.asp>
- [16] EUROCONTROL. *CODA Digest. All-causes delay and cancellations to air transport in Europe*, EUROCONTROL, Brussels, 2020.
- [17] Aerocivil. Aeronautical statistics of Colombia, 2023. <https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas>
- [18] FAA. *Cost of Delay Estimates. FAA APO-100*, Federal Aviation Administration, 2020, Washington D.C.
- [19] IATA. *Inefficiency in European Airspace*, IATA, 2013, Geneva.
- [20] Altmann, A., Tolosi, L., Sander, O. and Lengauer, T. Permutation importance: A corrected feature importance measure, *Bioinformatics*, 2010, **26**, (10), pp 1340–1347. DOI: [10.1093/bioinformatics/btq134](https://doi.org/10.1093/bioinformatics/btq134)
- [21] Breiman, L. Random forests, *Mach. Learn.*, 2001, **45**, pp 5–32.
- [22] Tan, X., Jia, X., Yan, J., Wang, K. and Bian, L. An exploratory analysis of flight delay propagation in China, *J. Air Transport Manag.*, 2021, **92**, 102025. DOI: [10.1016/j.jairtraman.2021.102025](https://doi.org/10.1016/j.jairtraman.2021.102025)
- [23] Kim, M. and Park, S. Airport and route classification by modelling flight delay propagation, *J. Air Transport Manag.*, 2021, **93**, 102045. DOI: [10.1016/j.jairtraman.2021.102045](https://doi.org/10.1016/j.jairtraman.2021.102045)
- [24] Kafle, N. and Zou, B. Modeling flight delay propagation: A new analytical-econometric approach, *Transp. Res. Part B*, 2016, **93**, pp 520–542. DOI: [10.1016/j.trb.2016.08.012](https://doi.org/10.1016/j.trb.2016.08.012)
- [25] Fleurquin, P., Ramasco, J. and Eguiluz, V. Systemic delay propagation in the US airport network, *Sci. Rep.*, 2013, **3**, pp 1–6. DOI: [10.1038/srep01159](https://doi.org/10.1038/srep01159)
- [26] Cai, Q., Alam, S. and Doung, V. A spatial-temporal network perspective for the propagation dynamics of air traffic delays, *Engineering*, 2021, **7**, pp 452–464. DOI: [10.1016/j.eng.2020.05.027](https://doi.org/10.1016/j.eng.2020.05.027)
- [27] Zhang, M., Zhou, X., Zhang, Y., Sun, L., Dun, M., Du, W. and Cao, X. Propagation index on airport delays, *Transp. Res. Record*, 2019, **2673**, (8), pp 536–543. DOI: [10.1177/0361198119844240](https://doi.org/10.1177/0361198119844240)
- [28] Wang, P., Schaefer, L. and Wojcik, L. Flight connections and their impacts on delay propagation, Digital Avionics Systems Conference, 12–16 October 2003, Indianapolis (USA).
- [29] Markovic, D., Hauf, T., Roehner, P. and Spehr, U. A statistical study of the weather impact on punctuality at Frankfurt airport, *Meteorol. Appl.*, 2008, **15**, (2), pp 293–303.
- [30] Xu, N., Sherry, L. and Laskey, K. Multifactor model for predicting delays at U.S. airports, *Transp. Res. Record*, 2008, **2052**, pp 62–71. DOI: [10.3141/2052-08](https://doi.org/10.3141/2052-08)



- [31] Zhang, J., Xu, X. and Wang, F. Airport delay performance evaluation based on fuzzy linear regression model, *J. Traffic Transp. Eng.*, 2010, **10**, (4), pp 109–114.
- [32] Hao, L., Hansen, M., Zhang, Y. and Post, J. New York, New York: Two ways of estimating the delay impact of New York airports, *Transp. Res. Part E*, 2014, **70**, pp 245–260. DOI: [10.1016/j.tre.2014.07.004](https://doi.org/10.1016/j.tre.2014.07.004)
- [33] Boswell, S. and Evans, J. Analysis of downstream impacts of air traffic delay. Project Report, Lincoln Laboratory, Massachusetts Institute of Technology, 1997.
- [34] Wong, J. and Tsai, S. A survival model for flight delay propagation, *J. Air Transport Manag.*, 2012, **23**, pp 5–11. DOI: [10.1016/j.jairtraman.2012.01.016](https://doi.org/10.1016/j.jairtraman.2012.01.016)
- [35] Wu, C. Inherent delays and operational reliability of airline schedules, *J. Air Transport Manag.*, 2005, **11**, (4), pp 273–282. DOI: [10.1016/j.jairtraman.2005.01.005](https://doi.org/10.1016/j.jairtraman.2005.01.005)
- [36] Baspinar, B. and Koyuncu, E. A data-driven air transportation delay propagation model using epidemic process models, *Int. J. Aerospace Eng.*, 2016, Article ID 4836260. DOI: [10.1155/2016/4836260](https://doi.org/10.1155/2016/4836260).
- [37] Pyrgiotis, N., Malone, K. and Odoni, A. Modelling delay propagation within an airport network, *Transp. Res. Part C*, 2013, **27**, pp 60–75. DOI: [10.1016/j.trc.2011.05.017](https://doi.org/10.1016/j.trc.2011.05.017)
- [38] Hansen, M. Micro-level analysis of airport delay externalities using deterministic queuing models: A case study, *J. Air Transport Manag.*, 2002, **8**, (2), pp 73–87. DOI: [10.1016/S0969-6997\(01\)00045-X](https://doi.org/10.1016/S0969-6997(01)00045-X)
- [39] Morrison, S. and Winston, C. The effect of FAA expenditures on air travel delays, *J. Urban Econ.*, 2008, **63**, (2), pp 669–678.
- [40] Zou, B. and Hansen, M. Flight delays, capacity investment and social welfare under air transport supply-demand equilibrium, *Transp. Res. Part A*, 2012, **46**, (6), pp 965–980. DOI: [10.1016/j.tra.2012.02.015](https://doi.org/10.1016/j.tra.2012.02.015)
- [41] Qin, Q. and Yu, H. A statistical analysis on the periodicity of flight delay rate of the airports in the US, *Adv. Transp. Stud.*, 2014, **3**, pp 93–104.
- [42] Pathomsiri, S., Haghani, A., Dresner, M. and Windle, R. Impact of undesirable outputs on the productivity of US airports, *Transp. Res. Part E*, 2008, **44**, (2), pp 235–259. DOI: [10.1016/j.tre.2007.07.002](https://doi.org/10.1016/j.tre.2007.07.002)
- [43] Kotegawa, T., De Laurentis, D., Noonan, K. and Post, J. Impact of commercial airline network evolution on the U.S. air transportation system, Proceedings of the 9th USA/Europe Air Traffic Management Research and Development Seminar, 2011, pp 572–580.
- [44] Pfeil, D. and Balakrishnan, H. Identification of robust terminal-area routes in convective weather, *Transp. Sci.*, 2012, **46**, (1), pp 56–73.
- [45] Ahmadbeygi, S., Cohn, A., Guan, Y. and Belobaba, P. Analysis of the potential for delay propagation in passenger airline networks, *J. Air Transport Manag.*, 2008, **14**, (5), pp 221–236. DOI: [10.1016/j.jairtraman.2008.04.010](https://doi.org/10.1016/j.jairtraman.2008.04.010)
- [46] Soomer, M. and Franx, G. Scheduling aircraft landings using airlines' preferences, *Eur. J. Operat. Res.*, 2008, **190**, (1), pp 277–291. DOI: [10.1016/j.ejor.2007.06.017](https://doi.org/10.1016/j.ejor.2007.06.017)
- [47] Mofokeng, T. and Marnewick, A. Factors contributing to delays regarding aircraft during A-check maintenance, IEEE Technology and Engineering Management Society Conference, TEMSCON 2017. DOI: [10.1109/TEMSCON.2017.7998375](https://doi.org/10.1109/TEMSCON.2017.7998375)
- [48] Zou, B. and Hansen, M. Flight delay impact on air fare and flight frequency: A comprehensive assessment, *Transp. Res. Part E*, 2014, **69**, pp 54–74. DOI: [10.1016/j.tre.2014.05.016](https://doi.org/10.1016/j.tre.2014.05.016)
- [49] Abdel-Aty, M., Lee, C., Bai, Y., Li, X. and Michalak, M. Detecting periodic patterns of arrival delay, *J. Air Transport Manag.*, 2007, **13**, (6), pp 355–361. DOI: [10.1016/j.jairtraman.2007.06.002](https://doi.org/10.1016/j.jairtraman.2007.06.002)
- [50] Zhong, Z., Varun, D. and Lin, Y. Studies for air traffic management R&D in the ASEAN-region context, *J. Air Transport Manag.*, 2017, **64**, pp 15–20. DOI: [10.1016/j.jairtraman.2017.06.020](https://doi.org/10.1016/j.jairtraman.2017.06.020)
- [51] Ferrer, J., Oliveira, P. and Parasuraman, A. The behavioral consequences of repeated flight delays, *J. Air Transport Manag.*, 2012, **20**, pp 35–38. DOI: [10.1016/j.jairtraman.2011.11.001](https://doi.org/10.1016/j.jairtraman.2011.11.001)
- [52] Lubbe, B. and Victor, C. Flight delays: Towards measuring the cost to corporations, *J. Air Transport Manag.*, 2012, **19**, pp 9–12. DOI: [10.1016/j.jairtraman.2011.11.004](https://doi.org/10.1016/j.jairtraman.2011.11.004)
- [53] Li, Q. and Jing, R. Characterization of delay propagation in the air traffic network, *J. Air Transport Manag.*, 2021, **94**, 102075. DOI: [10.1016/j.jairtraman.2021.102075](https://doi.org/10.1016/j.jairtraman.2021.102075)
- [54] Kim, M. and Bae, J. Modeling the flight departure delay using survival analysis in South Korea, *J. Air Transport Manag.*, 2021, **91**, 101996. DOI: [10.1016/j.jairtraman.2020.101996](https://doi.org/10.1016/j.jairtraman.2020.101996)
- [55] Mohammadian, I., Abbasi, B., Abareshi, A. and Goh, M. Antecedents of flight delays in the Australian domestic aviation market, *Transp. Res. Interdiscip. Perspect.*, 2019, **1**, 100007. DOI: [10.1016/j.trip.2019.100007](https://doi.org/10.1016/j.trip.2019.100007)
- [56] Du, W., Zhang, M., Zhang, Y., Cao, X. and Zhang, J. Delay causality network in air transport systems, *Transp. Res. Part E*, 2018, **118**, pp 466–476.
- [57] Lambelho, M., Mitici, M., Pickup, S. and Marsden, A. Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions, *J. Air Transport Manag.*, 2020, **82**, 101737. DOI: [10.1016/j.jairtraman.2019.101737](https://doi.org/10.1016/j.jairtraman.2019.101737)
- [58] Chen, Y. and Lin, J. Determinants of flight delays at East Asian airports from an airport, route and network perspective, *J. Air Transport Manag.*, 2021, **94**, 102064. <https://doi.org/10.1016/j.jairtraman.2021.102064>
- [59] Tu, Y., Ball, M. and Jank, W. Estimating flight departure delay distributions — A statistical approach with long-term trend and short-term pattern, *J. Am. Stat. Assoc.*, 2008, **103**, (481), pp 112–125. DOI: [10.1198/016214507000000257](https://doi.org/10.1198/016214507000000257)
- [60] Mueller, E. and Chatterji, G. Analysis of aircraft arrival and departure delay characteristics, AIAA's Aircraft Technology, Integration and Operations Conference, 1–3 October 2002, Los Angeles (USA).

- [61] Pamplona, D. and Alves, C.J. An overview of air delay: A case study of the Brazilian scenario, *Transp. Res. Interdiscip. Perspect.*, 2020, **7**, 100189. DOI: [10.1016/j.trip.2020.100189](https://doi.org/10.1016/j.trip.2020.100189)
- [62] Fernandes, N., Moro, S., Costa, C. and Aparicio, M. Factors influencing charter flight departure delay, *Res. Transp. Business Manag.*, 2020, **34**, 100413.
- [63] Hunter, G., Boisvert, B. and Ramamoorthy, K. Advanced national airspace traffic flow management simulation experiments and validation, Proceedings of the 39th Conference on Winter Simulation, 2007, pp 1261–1267.
- [64] Bakhshandeh, R., Shahgholian, K. and Shahraki, A. Model for reduce flights delays using System Dynamics (Case Study in Iranian Airports Company), *Interdiscip. J. Contemp. Res. Bus.*, 2013, **4**, (9), pp 746–757.
- [65] Santos, G. and Robin, M. Determinants of delays at European airports, *Transp. Res. Part B*, 2010, **44**, pp 392–403. DOI: [10.1016/j.trb.2009.10.007](https://doi.org/10.1016/j.trb.2009.10.007)
- [66] Wang, C. and Wang, X. Airport congestion delays and airline networks, *Transp. Res. Part E*, 2019, **122**, pp 328–349. DOI: [10.1016/j.tre.2018.12.008](https://doi.org/10.1016/j.tre.2018.12.008)
- [67] Wang, Y., Zhu, Y., Zhu, C., Wu, F., Yang, H., Yan, Y. and Hu, C. Indicator of serious flight delays with the approach of time-delay stability, *Physica A*, 2019, **518**, pp 363–373. DOI: [10.1016/j.physa.2018.11.038](https://doi.org/10.1016/j.physa.2018.11.038)
- [68] Zhang, M., Chen, S., Sun, L., Du, W. and Cao, X. Characterizing flight delay profiles with a tensor factorization framework, *Engineering*, 2021, **7**, pp 465–472. DOI: [10.1016/j.eng.2020.08.024](https://doi.org/10.1016/j.eng.2020.08.024)
- [69] Carvalho, L., Sternberg, A., Goncalves, L., Cruz, A., Soares, J., Brandão, D., Carvalho, G. and Ogasawara, E. On the relevance of data science for flight delay research: A systematic review, *Transport Rev.*, 2021, **41**, (4), pp 499–528. DOI: [10.1080/01441647.2020.1861123](https://doi.org/10.1080/01441647.2020.1861123)
- [70] Truong, D. Using causal machine learning for predicting the risk of flight delays in air transportation, *J. Air Transport Manag.*, 2021, **91**, 101993. DOI: [10.1016/j.jairtraman.2020.101993](https://doi.org/10.1016/j.jairtraman.2020.101993)
- [71] Guo, Z., Yu, B., Hao, M., Wang, W., Jiang, Y. and Zong, F. A novel hybrid method for flight departure delay prediction using random forest regression and maximal information coefficient, *Aerospace Sci. Technol.* 2021. DOI: [10.1016/j.ast.2021.106822](https://doi.org/10.1016/j.ast.2021.106822)
- [72] IATA. *Airport Handling Manual*, IATA, 2021, Geneva.
- [73] Guss, W. and Salakhutdinov, R. On universal approximation by neural networks with uniform guarantees on approximation of infinite dimensional maps, 2019, arXiv:1910.01545v1.
- [74] Vang-Mata, R. *Multilayer Perceptrons: Theory and Applications*, Nova Science Publication, 2020, Hauppauge (NY).
- [75] Agarap, A. Deep Learning using Rectified Linear Units (ReLU), 2019, arXiv:1803.08375v2.
- [76] Arora, R., Basu, A., Mianjy, P. and Mukherjee, A. Understanding deep neural networks with rectified linear units, 2018, arXiv:1611.01491v6.
- [77] Hara, K., Saito, D. and Shouno, H. Analysis of function of rectified linear unit used in deep learning, 2015 International Joint Conference on Neural Networks (IJCNN). DOI: [10.1109/IJCNN.2015.7280578](https://doi.org/10.1109/IJCNN.2015.7280578)
- [78] Arora, S. and Barak, B. *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009, Cambridge.
- [79] Bao, Y., Xiong, T. and Hu, Z. Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method, *Discrete Dyn. Nature Soc.*, 2012, ID 431512, pp 1–12. DOI: [10.1155/2012/431512](https://doi.org/10.1155/2012/431512)
- [80] Schonlau, M. and Zou, R. The random forest algorithm for statistical learning, *Stata J.*, 2020, **20**, (1), pp 3–29. DOI: [10.1177/1536867X20909688](https://doi.org/10.1177/1536867X20909688)
- [81] Montgomery, D., Peck, E. and Vining, G. *Introduction to Linear Regression Analysis*. Wiley, 2012, Hoboken (NJ).
- [82] Rokach, L. *Ensemble Learning: Pattern Classification Using Ensemble Methods*, World Scientific Publishing, 2019, Singapore.
- [83] Ganjisaffar, Y., Caruana, R. and Lopes, C. Bagging gradient-boosted trees for high precision, low variance ranking models, SIGIR'11, July 24–28, 2011, Beijing (China).
- [84] Friedman, J. Greedy function approximation: A gradient boosting machine, *Ann. Stat.*, 2000, **29**, pp 1189–1232.
- [85] Hanif, I. Implementing Extreme Gradient Boosting (XGBoost) classifier to improve customer churn prediction, ICSA 2019, August 02–03, Bogor (Indonesia). DOI: [10.4108/eai.2-8-2019.2290338](https://doi.org/10.4108/eai.2-8-2019.2290338)
- [86] Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system, KDD'16, August 13–17, 2016, San Francisco (USA). DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)
- [87] Wang, L., Zheng, C., Zhou, W. and Zhou, W. A new principle for tuning-free Huber Regression, *Statistica Sinica*, 2022, **32**, pp 1–25. DOI: [10.5705/ss.202019.0045](https://doi.org/10.5705/ss.202019.0045)
- [88] Sun, Q., Zhou, W. and Fan, J. Adaptive Huber Regression, 2018, arXiv:1706.06991v2 [math.ST].
- [89] Camana, M., Ahmed, S., Garcia, C. and Koo, I. Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks, *IEEE*, 2020, **20**, pp 19921–19933. DOI: [10.1109/ACCESS.2020.2968934](https://doi.org/10.1109/ACCESS.2020.2968934)
- [90] Geurts, P., Ernst, D. and Wehenkel, L. Extremely randomized trees, *Mach. Learn.*, 2006, **63**, pp 3–42. DOI: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1)
- [91] Izza, Y., Ignatiev, A. and Marques-Silva, J. On explaining Decision Trees, 2020, arXiv:2010.11034v1 [cs.LG].
- [92] Azadkia, M. Optimal choice of  $k$  for  $k$ -nearest neighbor regression, 2020, arXiv:1909.05495v4 [math.ST].
- [93] Goodfellow, I., Bengio, Y. and Courville, A. *Deep Learning*, MIT Press, 2016, Cambridge (MA).
- [94] Chen, S., Kuo, S., Chang, K. and Wang, Y. Improving the forecasting accuracy of air passenger and air cargo demand: The application of back-propagation neural networks, *Transp. Plann. Technol.*, 2012, **35**, (3), pp 373–392.
- [95] Kuo, S. and Chen, S. Air passenger and air cargo demand forecasting: Applying artificial neural networks to evaluating input variables, 12th WCTR, July 11–15, 2010, Lisbon (Portugal).
- [96] Liu, Y., Gao, Y. and Yin, W. An improved analysis of stochastic gradient descent with momentum, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver (Canada).

- [97] Gower, R., Loizou, N., Qian, X., Sailanbayev, A., Shulgin, E. and Richtarik, P. SGD: General Analysis and Improved Rates, 2019, arXiv:1901.09401v4 [cs. LG].
- [98] Dingari, M., Reddy, M. and Sumalatha, V. Air traffic forecasting using artificial neural networks, *Int. J. Sci. Technol. Res.*, 2019, **8**, (10), pp 556–559.
- [99] Hastie, T., Tibshirani, R. and Friedman, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Springer, 2013, New York.
- [100] Kingma, D. and Ba, J. Adam: A Method for Stochastic Optimization, 2014, arXiv:1412.6980.
- [101] Aggarwal, C. *Neural Networks and Deep Learning*, Springer International, 2018, Cham.
- [102] Hassoun, M. *Fundamentals of Artificial Neural Networks*, MIT Press, 1995, Cambridge (MA).
- [103] Pedrycz, W. and Chen, S. *Deep Learning: Concepts and Architectures*, Springer, 2020, Cham.
- [104] Chang, C. and Lin, C. LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, (3), pp 1–27.
- [105] Rousseeuw, P. and van Driessen, K. Elliptic envelope. A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, 1999, **41**, (3), pp 212–223.
- [106] Henriques, R. and Feiteira, I. Predictive modelling: Flight delays and associated factors, Hartsfield–Jackson Atlanta International Airport, *Procedia Comput. Sci.*, 2018, **138**, pp 638–645.
- [107] Belcastro, L., Marozzo, F., Talia, M. and Trunfio, P. Using scalable data mining for predicting flight delays, *ACM Trans. Intell. Syst. Technol.*, 2016, **8**, (1), Article 5. DOI: [10.1145/2888402](https://doi.org/10.1145/2888402)
- [108] ACI. *Annual World Airport Traffic Report*, ACI (Airports Council International), 2021, Montreal.
- [109] IATA. *World Air Transport Statistics 2021*, IATA, 2021, Geneva.
- [110] Díaz Olariaga, O. The role of regional airports in connectivity and regional development, *Periodica Polytechnica Transp. Eng.*, 2021, **49**, (4), pp 1–13. DOI: [10.3311/PPtr.16557](https://doi.org/10.3311/PPtr.16557)
- [111] Díaz Olariaga, O. and Pulido, L. Measurement of airport efficiency. The case of Colombia, *Transport Telecommun.*, 2019, **20**, (1), pp 40–51. DOI: [10.2478/tjt-2019-0004](https://doi.org/10.2478/tjt-2019-0004)
- [112] Díaz Olariaga, O. and Zea, J.F. Influence of the liberalization of the air transport industry on configuration of the traffic in the airport network, *Transp. Res. Procedia*, 2018, **33**, pp 43–50.
- [113] Díaz Olariaga, O. and Carvajal, A. Perspectiva geográfica del desarrollo de la conectividad aérea en Colombia, *Boletín Geográfico*, 2020, **42**, (2), pp 145–168.
- [114] Kearns, S. *Fundamentals of International Aviation*, Routledge, 2021, London.
- [115] Budd, L. and Ison, S. *Air Transport Management*, Routledge, 2017, London.
- [116] Borsky, S. and Unterberger, C. Bad weather and flight delays: The impact of sudden and slow onset weather events, *Econ. Transp.*, 2019, **18**, pp 10–26. DOI: [10.1016/j.ecotra.2019.02.002](https://doi.org/10.1016/j.ecotra.2019.02.002)
- [117] Choi, S., Kim, Y., Briceno, S. and Mavris, D. Prediction of weather-induced airline delays based on machine learning algorithms, IEEE/AIAA 35th Digital Avionics Systems Conference Proceeding, 2016. DOI: [10.1109/DASC.2016.7777956](https://doi.org/10.1109/DASC.2016.7777956)
- [118] ACI. *Guide to Airport Performance Measures*, ACI World, 2012, Montreal.
- [119] FAA - Federal Aviation Administration. Definitions of variables Available, 2022, [https://aspmhelp.faa.gov/index.php/APM:\\_Analysis:\\_Definitions\\_of\\_Variables](https://aspmhelp.faa.gov/index.php/APM:_Analysis:_Definitions_of_Variables)
- [120] OAG. *OAG Punctuality League*, OAG Aviation Worldwide, 2017, Chicago.
- [121] ICAO. *Manual on Collaborative Air Traffic Flow Management (ATFM). Doc 9971*, ICAO, 2018, Montreal.
- [122] European Observatory on Airport Capacity & Quality. *Delays to Air Transport in Europe: Methods of Measuring, Reporting and Analyzing*, European Commission, 2015, Brussels.
- [123] EUROCONTROL. *A Matter of Time: Air Traffic Delay in Europe*, EUROCONTROL, 2007, Brussels.