Towards Flight Delays Reduction: The Effect of Aircraft Type and Part of Day on Arrival Delays Prediction

Hajar Alla *, Lahcen Moumoun, and Youssef Balouki

Laboratory of Mathematics, Computer and Engineering Sciences, Mathematics and Computer Science Department, Faculty of Sciences and Techniques, Hassan First University of Settat Settat, 26000, Morocco;
Email: lahcen.moumoun@uhp.ac.ma (L.M.), youssef.balouki@uhp.ac.ma (Y.B.)
*Correspondence: h.alla@uhp.ac.ma (H.A.)

***Abstract*—The basic objective of this study is to develop a model that analyzes and predicts the occurrence of flight arrival delays in the United States. Macroscopic and microscopic delay factors are discussed. In this research, we proposed new features that, to the best of our knowledge, were never used in previous studies, namely departure Part and Arrival Part of the day (Mornings, Afternoons, Evenings, Nights) and type of aircraft. U.S. domestic flight data for the year 2018, extracted from the Bureau of Transportation Statistics (BTS), were adopted in order to train the predictive model. We used efficient Machine Learning classifiers such as Decision Trees, K-Nearest Neighbors, Random Forest and Multilayer Perceptron. To overcome the issue of imbalanced data, sampling techniques were performed. We chose Grid Search technique for best parameters selection. The performance of each classifier was compared in terms of evaluation metrics, parameters tuning, data sampling and features selection. The experimental results showed that tuning and sampling techniques have successfully generated the best classifier which is Multilayer Perceptron (MLP) with an accuracy of 98.72% and a higher number of correctly classified flights.**

***Keywords*—machine learning classification, flight delay prediction, multilayer perceptron, random forest, decision trees, k-nearest neighbors**

I. INTRODUCTION

Flight delays affect passengers, airlines and airport managers. Indeed, it is a major preoccupation in air transportation systems. According to the Bureau of Transportation Statistics (BTS) [1] in 2017, 19.07% of all domestic flights in the United States were delayed. Many factors are responsible for traffic delays, including bad weather conditions, technical problems, late passengers, airport crowdedness, runway queues, a lack of airport infrastructure, aircraft turnover delays (fueling/refueling, loading/unloading, boarding/disembarking, etc.), and late arriving aircraft, which impacts the following flight by creating a delay propagation. All these factors result in airline economic losses and penalties, flight cancellations, passengers complaints and dissatisfaction, delay propagation on next flights, fuel consumption, and gas emissions. To solve flight delay issues, studies and research were conducted to find solutions for traffic delays. The objective of this study is to predict flight arrival delays using efficient Machine Learning classification algorithms such as Multilayer Perceptron, Random Forest, Decision Trees and K-Nearest Neighbors (K-NN). In comparison with other conventional statistical approaches, the Multilayer Perceptron has proven to be successful with nonlinear systems, unknown and unseen data, particularly in prediction applications [2]. Random forest is known for being a sophisticated and powerful prediction tool that makes it easier to analyze the data and, more importantly, to identify influential elements for the problem of interest [3]. Decision tree models are famous for being quick and simple to create, interpret, and learn. Their forecasts are mostly effective [4]. Simple to train and utilized in a large variety of applications, k-NN is one of the most commonly used methods in classification. It is an uncomplicated algorithm with generally good performance [5]. Traffic information for U.S. domestic airline flights for the year 2018 was extracted from the BTS database [1]. The part of day is a contributing factor to delays. In fact, depending on the season, airlines or passengers may choose the same part time of the day to perform their flights, either at night, in the evening, in the afternoon, or in the morning, which may lead to a peak period causing density and traffic delays. Types of aircraft do not perform the same. Some are sensitive to bad weather conditions; others have a low rate of climb or descent or a low velocity. An aircraft with more seats is more subject to flight delays since it transports more passengers who can be late for boarding, which causes delays. For this purpose and to enhance the performance of the proposed prediction model, we created new features that, to the best of our knowledge, were not used in previous studies, namely, Aircraft Type, Departure part and Arrival part of the day. Through our work, we will demonstrate that airports of origin and air carriers are also responsible for flight delays. The problem of imbalanced

Manuscript received January 5, 2023; revised February 1, 2023; accepted March 22, 2023; published September 25, 2023.

data in classification has become more challenging recently. To handle this issue, under-sampling and over-sampling techniques are often employed. Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links are used for over-sampling and under-sampling imbalanced data, respectively. To improve the efficacy of our method, we utilized a combination of both techniques and applied SMOTE-Tomek to balance our dataset. Hyper-parameters tuning is implemented in machine learning to get optimal values of the parameters and reach a higher level of accuracy. In this research, we used Grid Search technique to find the best parameters for Random Forest, Decision Trees and K-NN. We examined and compared the performance of the proposed method based on classification metrics, parameters tuning, data sampling, and feature selection.

The rest of the paper is organized as follows: Section II provides a review of the existing literature on flight delay prediction. Section III describes in detail the methodology proposed to develop the predictive model. In Section IV, we compare the performance of all classifiers in terms of evaluation metrics, parameters tuning, data balancing, and feature selection. Section V concludes the research and recommends perspectives and further work.

II. Literature Review

Nowadays, more interest has been given to flight delays because of the massive increase in air traffic. Existing studies in this area concentrated the most on statistical methods based on approximations and probabilities, leading to unsure conclusions and results. Idris and Clarke *et al.* [6] focused on the prediction of taxi-out delays for aircraft based on a queuing model. Tu and Ball *et al.* [7] described and studied major factors that influence flight departure delays by estimating the entire delay distribution. Alla and Moumoun *et al.* [8] conceived a predictive model to predict flight delay using machine learning algorithms such as random forest, decision trees, and multiple linear regression. As a result, Random Forest proved to be the best model. Zhong and Varun *et al.* [9] performed an investigation about flight delays based on Monte Carlo simulations considering runway occupancy time, separation between aircraft, and other operational scenarios at the airport. Nogueira and Aguiar *et al.* [10] proposed solutions in order to optimize taxi-scheduling operations for aircraft with Ant Colony Optimization. The model shows a performance of 32% in taxiing time minimization, aircraft collision conflicts included. The objective of the study conducted by Jiang and Xu *et al.* [11] is to reduce both the total aircraft taxiing and waiting times using a genetic algorithm. Mueller and Chatterji [12] studied departure, en route, and arrival flight delays using historical data from United States airports based on density function modeling. Unlike traditional statistical methods, which have generally proven to be weak, slow, and limited, machine learning algorithms have become more popular recently because they lead to a higher accuracy and deal with a huge amount of data. Several research studies have been conducted using Machine Learning (ML) algorithms for flight delay prediction. Achenbach and Spinler [13] focused on predicting flight arrival times when aircraft are in the block. They considered features such as weather forecasts, time and airport congestion data using gradient boosting and linear regression algorithms. Kalliguddi and Leboulluec *et al.* [14] conceived a predictive model to predict flight delay using machine learning algorithms such as random forest, decision trees, and multiple linear regression. As a result, Random Forest proved to be the best model. Bhuvaneshwari and Elakiya *et al.* [15] applied machine learning techniques and statistical models in order to forecast flight delays in the United States. Supervised machine learning algorithms were selected by Choi and Kim *et al.* [16] to produce a flight delay predictive system. They applied a sampling technique called costing, which enhanced the performance of the cost-sensitive classifiers. Thiagarajan and Srinivasan *et al.* [17] have implemented a machine learning approach to predict whether the flight is delayed or not using Gradient Boosting Classifier for the classification and Extra-Trees Regressor for the regression. In the research [18], flight and weather data have been utilized to predict on-time arrival traffic using machine learning methods. The proposed system revealed 77% of the accuracy for Random Forest. Wu and Cai *et al.* [19] adopted Support Vector Machine (SVM) to estimate traffic departure delays using historical data from Beijing Capital International Airport. Compared to other algorithms, SVM was the best, with an R^2 of 0.71 and a lower time of execution. Chakrabarty and Kundu *et al.* [20] developed a system in which they estimate flight arrival delays operated by American Airlines with the use of four supervised machine learning algorithms. Kuhn and Jamadagni [21] picked features such as departure delay, origin airport, destination airport, and distance between airports in order to predict flight arrival delays. The proposed model achieved an accuracy of 91% for the prediction. In order to solve the flight-to-gate assignment problem, Zoutendijk and Mitici [22] applied Random Forest for flight delay prediction. Truong [23] performed causal data mining with machine learning algorithms. The proposed system achieved a high prediction accuracy of 91.97% for flight delays. Alla and Moumoun *et al.* [24] developed a multilayer perceptron neural network to forecast traffic arrival delays based on selective training. The model with selective training proved to be the best compared with the traditional training one with a higher accuracy and a lower computational time. In the existing literature, authors have worked differently on the problem of flight delays. In this study, we proposed a system that predicts air traffic delays employing supervised machine learning algorithms, where classification is performed to estimate the occurrence of the delay. New features that contribute to flight delays are considered for the first time in this study. Parameter optimization and data sampling were adopted in order to improve the accuracy of the proposed system.

III. Materials and Methods

A. Problematic of the Research

According to the ICAO's document no. 9859 [25], safety has always been the number one priority in all aviation activities. The future viability of the air transportation sector may be dependent on its ability to maintain the public's perception of safety while traveling. Safety management is thus a must for a long-term aviation business.

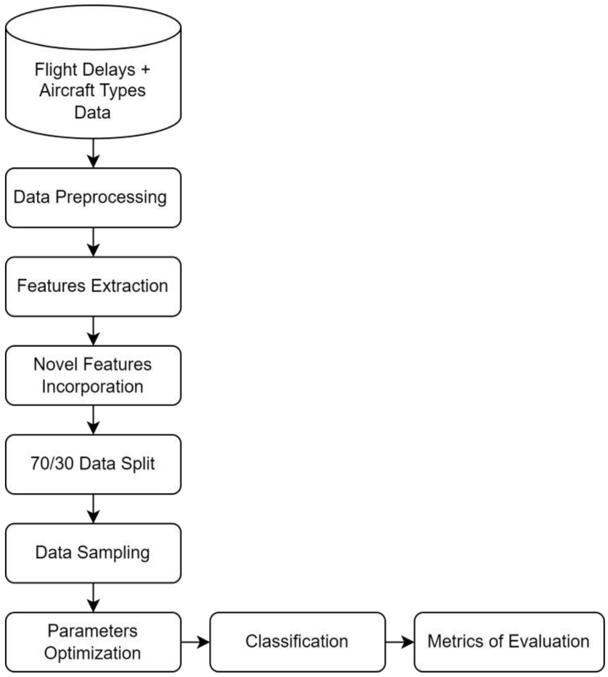

Figure 1. Flowchart of the proposed method.

Delay is regarded as the second-most important element in the air transportation system, behind safety. It is detrimental to passengers, airlines, and airport staff. Numerous initiatives have been made to alleviate this problem. It is necessary to have a system that forecasts flight arrival delays and notifies airport staff, airlines, and travelers. Flight delays can be caused by a number of macro-level reasons, including poor weather, technical difficulties, late and unruly passengers, airport crowding, runway lines, a lack of airport infrastructure, delays with aircraft maintenance, delays with flight checklists, etc. In this study, our goal is to find and use new microlevel elements that contribute to flight delays in order to build a predictive model based on effective machine learning classifiers. Fig. 1 outlines the steps used to produce and generate the proposed model. Flight on-time data were taken from the BTS [1] database. Types of aircraft, constructors, and registration numbers were extracted from Github. Techniques for preprocessing and cleaning data were used to eliminate information that is duplicated, repetitive, noisy, and useless. We selected as features only necessary attributes that are pertinent and responsible for flight delays. Other microlevel elements of delays were explored in order to increase the effectiveness of the suggested system. We developed three novel features that, as far as we know, were the first attempt in this study: type of aircraft, arrival and departure parts of the day. We will prove later that the proposed new features have enhanced the accuracy of the suggested system. We split the data into 70% for training and 30% for testing. The dataset was imbalanced since the percentage of non delayed flights was higher than that of delayed flights. Therefore, sampling techniques using a combination of Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links were performed. For hyperparameter optimization, we chose Grid Search method. Finally, we evaluated the efficacy of the suggested model based on classification metrics, parameter tuning, data sampling, and feature selection.

IV. Data Collection

Firstly, data concerning on-time and delayed flights inside the U.S. for all the year 2018 were taken from the BTS database [1]. More than 100,000 flight recordings were available in our dataset. We chose this database due to its open access archives, trustworthiness, integrity, and neutrality [26]. Following the rule established by the Federal Aviation Administration [27], we classified as delayed all the flights whose actual arrival times are 15 mins or more bigger than their scheduled arrival times, and the other flights as on-time. Second, we obtained more than 400,000 records of aircraft constructors, types and registrations from the International Civil Aviation Organization (ICAO) database [28].

V. Data Preprocessing

A. Data Cleaning

Database data often contained mistakes, omissions, misunderstandings, and incompleteness. It could also be obtrusive and unreliable [29]. Therefore, data that can be understood and analyzed by computers and machine learning algorithms should be cleaned. We eliminated redundant information, empty fields, and missing values. We converted categorical data into numbers.

B. Data Balancing

Data after transformation contained 93,769 flights. The length of non delayed flights was 73,297 samples and that of delayed flights was 20,472 samples. The training data were imbalanced since the percentage of non delayed flights was 78% and that of delayed flights was only 22%, which made the data not equally presented. To solve the problem of class balancing, combined undersampling and oversampling techniques were used. The combination was chosen mainly because undersampling and oversampling, if utilized separately, have many problems. With a large sample of dataset, undersampling might be the reason for information loss. However, cases of overfitting and overgeneralization have been noticed when using oversampling methods [30]. Hence, the combination of both techniques was suggested to overcome this kind of issue. The sampling algorithm used in this study is SMOTE-Tomek. The Synthetic Minority Oversampling Technique (SMOTE) is an oversampling method that generates synthetic minority class examples in the dataset.

In order to over-sample the minority class, synthetic examples are introduced from a minor-class tuple and its nearest neighboring k-tuples [31]. SMOTE has been proven to be efficient at improving the accuracy of classifiers. According to Garcia and Luengo *et al.* [32], it is one of the most influential data preprocessing and sampling algorithms in machine learning and data mining. The Tomek Technique is the ability to delete samples of data from the majority class that are close to the minority class data by Tomek links. The SMOTE-Tomek technique is a combination of oversampling by SMOTE and undersampling by Tomek links. Jonathan and Putra *et al.* [33] used SMOTE-Tomek to improve the precision-recall of the selling category on a beauty platform. Experimental results for a common disease prediction conducted by Zeng and. Zou *et al.* [34] showed that combining SMOTE with the Tomek links technique is effective. Evaluation metrics and performance are evidently more improved and efficient than those with only SMOTE.

VI. Features Description

A. *Features Extraction*

Numerous variables and data fields were present in the extracted data. We removed non-essential features and kept just the ones that affect flight delays.

Airlines delay: According to BTS [1], airlines delays are brought on by technical problems with aircraft, pushback, icing/de-icing, fueling/refueling, charging/unloading, catering, or cleaning operations. Besides, delay propagation, which can be generated from an earlier flight delay, is also considered an important factor. Crew members, airport personnel stress, fatigue [26] or protestions [24], are also responsible for airline flight delays or cancellations.

Flight Distance: The travel time is influenced by the distance between the airports of Departure and Arrival. If it is greater, there is a larger chance that the flight may be delayed. According to Alla and Moumoun *et al.* [24], the delay increases with distance. Departure Delay: Delays in departure lead to delays in arrival. The majority of delayed traffic at the destination airport, according to Oza and Sharma *et al.* [35], derives from the outgoing airport. A delay at the origin will be propagated to the destination, leading to flight arrival delays [24].

B. *Novel Features Incorporation*

For the choice of our proposed features, we were based on international organizations, agencies, and associations judgments that work to achieve smooth and safe air traffic management and which govern and administrate civil and commercial aviation across Europe and the United States.

Type of Aircraft: In the FAA pilot handbook [27], performance refers to an aircraft's ability to land and take off in a short distance, carry large loads, fly at high altitudes at high speeds, and/or traverse long distances. Specific aircraft models may be particularly vulnerable to certain sorts of weather conditions, resulting in cancellations or delays. Older aircraft models may be more susceptible to mechanical problems that cause delays. Furthermore, some aircraft types may have more complicated maintenance requirements, causing repairs to take longer.

Based on their performance, certain aircraft may be more adaptable in terms of being on time [36]. Flight delays are triggered by disparities in aircraft types, in accordance with [37]. The capacity of airports and the sky is generally defined by light aircraft such as Cessna, Piper, Beechcraft, etc. When a light aircraft leads the route, all the following flights will be delayed, even if they are heavy or medium aircraft. That is because the velocity of the preceding aircraft was smaller. The following high-performance flights are not allowed then to increase speed as long as the low-performance aircraft is preceding. This is how flight delays are generated. The document no. 9854 of ICAO [38] defines the longitudinal separation between two aircraft as follows:

10 mins, if both aircraft have the same performance.

5 mins if the preceding aircraft is maintaining a true airspeed of 37 km/h (20 kt) higher than the succeeding aircraft or more.

3 mins if the preceding aircraft is maintaining a true airspeed of 74 km/h (40 kt) higher than the succeeding aircraft or more.

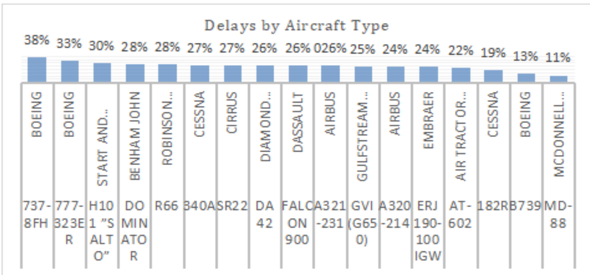

Figure 2. Data set arrival delays by type of aircraft. Source: own calculation.

We deduce that the difference in performance determines the longitudinal separation between aircraft. If the preceding traffic is faster than the succeeding one, the separation is lower (since the second cannot catch the first one), and the risk of delay will be lower too. In Fig. 2, we plot all the types of aircraft and their constructor used in our dataset organized by percentage of arrival delays. We notice, for example, that 27% of the flights performed by Cessna were delayed but only 13% by Boeing 739 were delayed too.

Part of the Day: Everyone is aware that the part of day is a major element to examine when traveling. Practically every traveler has a preferred boarding time. Some people like to travel early in the morning, while others prefer to take the flight late at night, and so on. There are also times of the day when there are a lot of traffic jams, which can cause traffic delays or cancellations.

The British government has placed restrictions on the flights performed during the night period from 23:30 to 06:00 Local Time at Heathrow airport in order to reduce the density of air traffic and avoid delays. It was not a perfect solution since, according to the report published by the International Air Transport Association (IATA) [39], there are a modest but commercially significant number of flights during the larger "night period" from 23:00 to 06:00. Late-night departures facilitate business connections in emerging economies as well as essential overnight freight transport. Furthermore, early morning operations are vital for passengers' business connections. According to the Federal Aviation Administration (FAA) handbook [40], the pilot is influenced by several factors of night operations that must be considered during night flight shifts, such as night visions, night illusions, the lighting systems, stress and fatigue, jet-lag, etc., which affect the flight's safety and its possibility of not arriving on time. The National Aeronautics and Space Administration (NASA), through a paper written by Elizabeth and Gregory *et al.* [41], identified early morning starts as a major fatigue factor for pilots and crew members. A system that implements rules on flight, duty time limitations and rest requirements for commercial air transport with aircraft was proposed by the European Aviation Safety Agency (EASA) [42] to manage the stress and fatigue of air navigation personnel. According to a study conducted by Nate Silver [43] in 2014 named *Fly Early, Arrive On-Time*, the period between 6 a.m. and 7 a.m. is the optimal time to fly. From Fig. 3, flights scheduled to depart within that time part were just 8.6 mins late on average. Flights departing before 6 a.m. or between 7 a.m. and 8 a.m. are almost excellent. Throughout the rest of the morning and afternoon, for every hour later the passenger departs, an extra minute of delay should be expected. Delays peak is 20.7 mins, more nearly doubling that of early-morning flights between 6 p.m. and 7 p.m. They will stay at 20 mins or more till 9 p.m.

Due to visibility conditions, we expect more aircraft in the morning and afternoon than in the evening. To this end, we decided to add two novel aspects to our work: the departure and arrival time parts of the day, which represent the part-moment of the day when the flight departs and the part-moment when the plane arrives, either in the early morning, late morning, afternoon, evening or, night. To correlate flight delays in our model with the part of the day in which they were performed, we utilized data about sunrise and sunset times for the year 2018 in the United States from the Worlds´ date, time and time zones database [44].

To provide airlines, airport managers and passengers with a system that predicts in advance the occurrence of flight delays, the features previously discussed should be explored in this study.

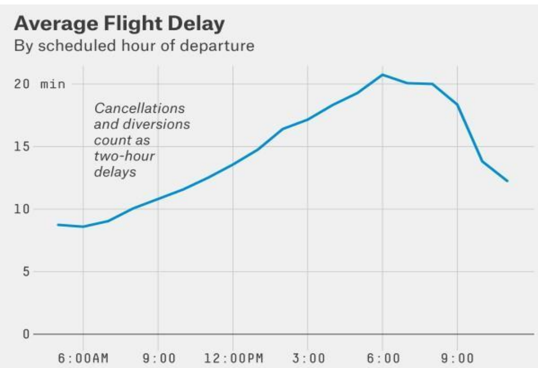

Figure 3. Flight delays by scheduled departure time. Source: Bureau of Transportations Statistics (BTS) [1].

VII. Machine Learning Classifiers

For the classification of a flight in two categories: Delayed or On-time (non delayed), we employed machine learning algorithms, e.g., K-Nearest-Neighbors, Decision Trees, Random Forest, and Multilayer Perceptron.

A. K-Nearest-Neighbors

The K-NN algorithm is part of an Instance-Based Learning model in which a new instance is classified by making comparisons against the most similar and close instance in the training dataset. K is the number of similar instances and closest neighbors being compared with the new instance [45]. Using two populations designated by A and B, C is affected to the population A, if at least 1/2 k of the k values neighbors to C originate from A.

B. Decision Trees

Decision trees are a popular and widely-used machine learning technique generally utilized to address prediction situations. Specifically, determining a discrete class labels (classification) or predicting a continuous value (regression) from a series of predictors [46]. It consists of nodes that constitute a rooted tree with no incoming edges. Every other node has only one incoming edge. An internal or test node is one that has outgoing edges. The other nodes are known as leaves (sometimes called terminal or decision nodes). Each internal node in a decision tree divides the instance space into two or more sub-spaces based on a discrete function of the input feature values. Each leaf is allocated to a class that represents the ideal target value. Additionally, the leaf might contain a probability vector showing the likelihood of the target attribute having a specific value. Based on the results of the tests along the path, instances are identified by routing them down from the tree's root to a leaf [47].

C. Random Forest

Random Forests are an ensemble of classification and regression decision trees that constitute basic models that employ binary splits on predictors to produce forecasts [48]. They were first proposed in 2001 by Breiman [49]. Segments of data are sampled, a randomized tree predictor is generated on each single piece, and then the predictors are aggregated together [50]. Input features from the training dataset are used to randomly build the trees. Random forests generally outperform other classification models in terms of prediction accuracy. Random Forest is used in the area of Bioinformatics [51], in route pavement maintenance [52], in remote sensing [53], in network intrusion detection [54], in modeling surface water salinity [55], and so on.

D. Multilayer Perceptron

A Multilayer Perceptron (MLP) neural network has been considered an alternative to traditional statistical techniques [24] due to its ability to learn in a better way complex relationships between input and output patterns [56]. It is composed of an input layer, one or many hidden layers, and an output layer. Each input has a connected weight, and each output has a transfer or activation function [57]. Eq. (1) is used to calculate the output value of each neuron.

$$y_j = F\left(b + \sum_{i=1}^{n} x_i w_i\right) \quad (1)$$

where x_i is the value of the input i, w_i is the weight, b is the bias, F is the activation function, and y is the output value of the neuron j.

VIII. Parameters Tuning

To find optimal parameters for our classifiers, improve accuracy, save time and energy, we adopted Grid Search method for parameters optimization using cross validation (cv=10).

For the parameters optimization, we used the class GridSearchCV available in Scikit Learn [58], which is a package in Python. In order to tune the parameters, we followed the idea of [59], in which all possible combinations for Random Forest were considered to find optimal parameters. Table I provides the results of tuning parameters for Random Forest, Decision Trees and KNearest Neighbors. In Random Forest, 7 combinations of parameters were tested in order to extract the best one. "max features" represents the number of features to consider when looking for the best split and "n estimators" is the number of trees in the forest. In Decision Trees, "criterion" contains the function used to measure the quality of a split, "max depth" is the maximum depth of the tree and "max leaf nodes" represents the total number of terminal nodes in a tree. In K-Nearest Neighbors, the number of neighbors to use is interpreted by the parameter "n_neighbors", the distance metric to use for the tree is performed by the parameter "metric". The power parameter for the Minkowski metric is "*p*". The parameters of each classifier are fully explained in [58]. All potential combinations are considered and optimal ones are chosen. The best of Random Forest was number 5 with 95.27% of accuracy, "max features: sqrt", and "n estimators: 700" as the optimal parameters. Among the values selected for Decision Trees, the best were, "max leaf nodes": 450, "max depth": 30, "criterion": "entropy", and 93.18% of accuracy. The best parameters in case of K-NN represents, "n_neighbors":5, "*p*": 1, "metric": "minkowski" with 90.50% as the best accuracy. To find optimal parameters of the MLP classifier, we focused on the analysis of prior studies. According to Ref. [61–68], it is sufficient to use one single hidden layer to learn N arbitrary samples in a feed-forward neural network with N hidden neurons when using almost any transfer function Sigmoid, Logistic, etc. But the network would then become very large with the augmentation of input samples. To reduce the number of hidden nodes, Tamura and Tateishi [69] proposed the use of a second hidden layer. Huang [70] demonstrated that in a two hidden layer network, the number of hidden nodes able to learn N samples with a small trivial error can be represented by Eqs. (2) and (3).

$$L_1 = \sqrt{(m+2)N} + \frac{2\sqrt{n}}{\sqrt{m+2}} \quad (2)$$

$$L_2 = \frac{m\sqrt{n}}{\sqrt{m+2}} \quad (3)$$

While L_1 and L_2 are respectively the numbers of nodes in the first and second hidden layers and m the output neurons.

Since L_1 and L_2 provide a very big number of nodes, the network will be more complex and may overfit the training data. According to Stathakis [71], an accurate topology will contain fewer nodes than that suggested by both Eqs. (2) and (3). Since no established methodology and exact solution was found to estimate an optimally or quasi-optimally ANN architecture, each researcher apply a different method to come up with satisfying results. Several methods are utilized, such as trial and error, heuristic search, exhaustive search, Grid Search, Random Search, Bayesian, pruning, and constructive algorithms. In our case, we applied Grid Search technique to find the best parameters. In the first selection of tuples of hidden layers and neurons, we were inspired by the study of Sonawane *et al.* [72] by using one single hidden layer and hidden neurons from 5 to 20 with a step of 5 ((5), (10), (15), (20)). The authors got the highest accuracy of 98.58% for 20 neurons for Heart Disease Prediction. As a second selection, we adopted the structure used by Stefanovic and Štrimaitis *et al.* [73] to estimate Flight Time Deviation for Lithuanian Airports which consists of a single layer with 10 to 100 hidden neurons by a step of 10 ((10), (20), (30), (40), (50), (60), (70), (80), (90), (100)). As a third run, we proposed a combination of single and two hidden layers ((50), (100), (50, 50), (100, 50), (100, 100)). For each run, we fixed the selection of activation function, solver, learning rate, number of iterations and regularization term (alpha), as represented in Table I. The best parameters of our MLP classifier were: "activation": "tanh", "alpha": 0.01, "hidden layer sizes": (20), "learning rate": "adaptive", "max iteration": 1000, "solver": "adam" with a best accuracy of 97.44% with the hyper-parameters tuning.

IX. Experimental Results and Discussion

The model used data for 63 U.S. airports and 4 carriers (Jet Blue Alaska Airlines American Airlines and Delta Airlines) operating domestic flights in the United States, and extracted from the BTS [1] database. Based on the registration (Tail number) of each flight, we were able to obtain from ICAO [28] database the constructor and the type of the aircraft used in each flight. More than 100,000 flight records and 400,000 aircraft types were collected before the study was carried out. Preprocessing techniques were performed in order to prepare the data for the training. A percentage of 70% was dedicated to the training set and 30% to the testing. To investigate new aspects leading to flight delays and produce a good forecast, novel features were proposed in this paper. Efficient Machine learning classifiers were chosen in order to build a good predictive model. Parameters optimization and data balancing with respectively Grid Search and SMOTE-Tomek techniques were implemented.

A. Performance of Features

Part of the Day Effect: We discovered, once data based on novel features were studied and analyzed, that departing aircraft were mostly behind schedule on nights, afternoons, and late mornings. To explain more, because it is less subject to congestion, the morning is the most requested part of the day from passengers to travel in. Also, it has typically a moderate temperature and weather. Traffic jams, bad visibility, crosswinds, airport and airspace density, all cause delays especially in the afternoons and nights. Delays on afternoons and evenings can also propagate to following flights, which might extend the delay and cause late arrivals at night. Lastly, from Fig. 4 and Fig. 5, we observe that flights operated around midday were less influenced by delays. The figures are a representation of plotting the results of delayed flights and the part of the day based on our own calculation. It means taking all the delayed flights values and searching whether they occurred in the early morning (Actual Arrival Time is from midnight to 06:05 U.S. Local Time), late morning (Actual Arrival Time is bigger than 06:05 and less than 12:00 U.S. Local Time), noon (Actual Arrival Time is equal to 12:00 U.S. Local Time), afternoon (Actual Arrival Time is bigger than 12:00 and less than 18:00 U.S. Local Time), evening (Actual Arrival Time is higher than 18:00 but less than 21:07 U.S. Local Time), or night (Actual Arrival Time is from 21:07 to 23:59 U.S. Local Time) based on the world's top ranking website [44] for date, time and time zones. After classifying all the flights into (Early Morning, Late Morning, Noon, Afternoon, Evening and Night), we plotted the results. In Fig. 4, the label "count" represents the sum of all the flights that were delayed at departure based on the part of the day. For example, 400 flights were delayed on departure in the early morning. In Fig. 5, the label "count" announces the sum of all the flights that were delayed at arrival based on the part of the day. For example, 3,500 flights were delayed on arrival in the afternoon).

Type of Aircraft Effect: In Fig. 6, we illustrated the results of the five most delayed types of aircraft. Arrival delays were particularly noticeable on Airbus and Embraer constructors. One suspected cause of the delay was the low performance compared, for example, with Boeing. The figure represents how many delays a specific type of aircraft experienced. For example, the Embraer 190 was the most delayed, with a percentage of delay of 30% compared with the other types of aircraft. Data of aircraft types, models, and registrations were selected from ICAO's [28] database.

Airlines Effect: Fig. 7 shows disparities within the delay that occurred due to circumstances within the airline's control for each company used in our dataset. We can explain that by the fact that each airline has its own policies about how to handle and avoid flight delays by defining Standard Operating Procedures (SOPs) or briefings in order to promote a culture of minimum delays. Any shortcomings in these procedures might contribute to delays. In Fig. 7, the label "Delays" means how many delays a specific airline experienced. For example, the carrier Delta Airlines was delayed 390 times.

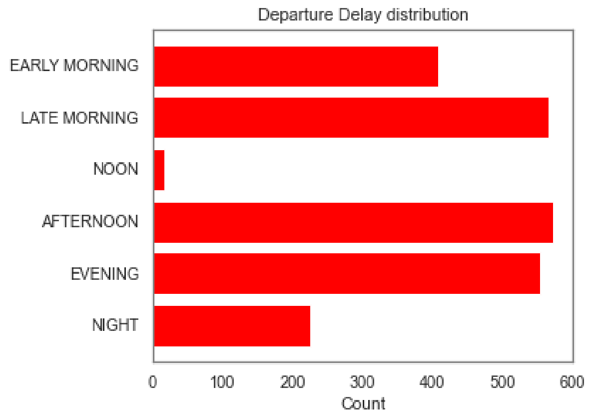

Figure 4. Departure delay by part of the day. Source: own calculation.

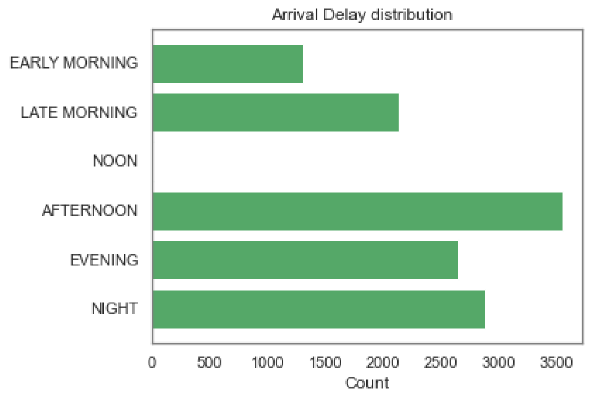

Figure 5. Arrival delay by part of the day. Source: own calculation.

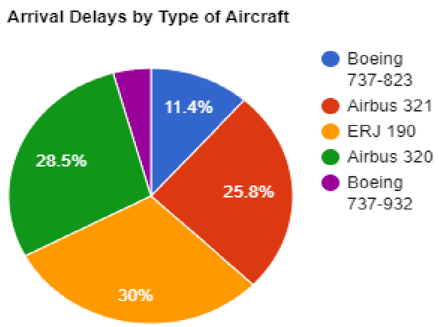

Figure 6. Arrival delay by type of aircraft. Source: own calculation.

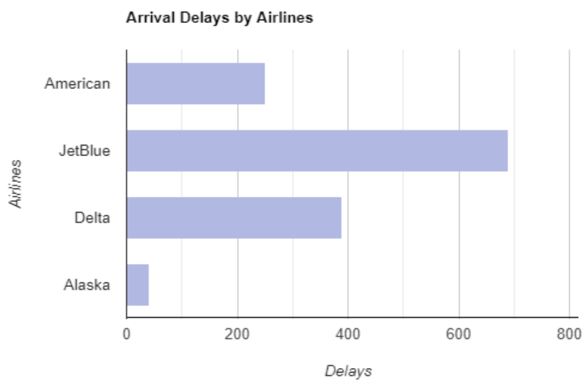

Figure 7. Arrival delay by Carrier. Source: own calculation.

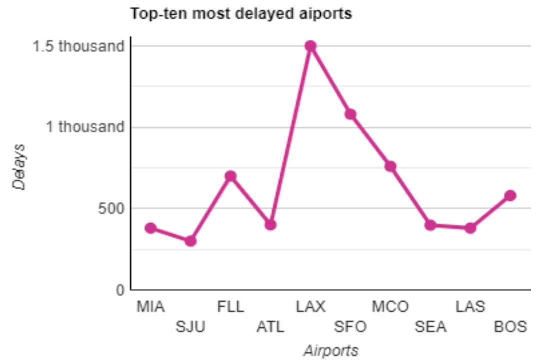

Figure 8. The 10-most delayed airports of departure. Source: own calculation.

Airport of Departure Effect: We decided to display in Fig. 8 only the top-ten most delay-affected departure airports from our dataset since the representation of 62 airports could be unreadable in a figure. In this distribution, Los Angeles was the airport with the highest percentage of delays, and San Juan was the airport with the lowest one. The difference in airport delay classification can be caused by many parameters, such as airport infrastructure, capacity, location, employees, operational management, etc. The label "Delays" demonstrates how many delays (the sum of delays per airport) a specific airport faced.

TABLE I. BEST PARAMETERS USING GRID SEARCH TECHNIQUE

Algorithm	No	Tuning Parameters	Score (10-fold)	Best Parameters
Random Forest	1	'max_features': ['sqrt', 'log2'], 'n_estimators': [290, 300, 310],	95.24%	'max_features':'sqrt', 'n_estimators': 300
	2	'max_features': ['sqrt', 'log2'], 'n_estimators': [250, 300, 350, 400]	95.19%	'max_features':'sqrt', 'n_estimators': 250
	3	'max_features': ['sqrt', 'log2'], 'n_estimators': [50, 100, 150, 200]	95.20%	'max_features':'sqrt', 'n_estimators': 200
	4	'max_features': ['sqrt', 'log2'], 'n_estimators': [275,300,325]	95.25%	'max_features':'sqrt', 'n_estimators': 325
	5	'max_features': ['sqrt', 'log2'], 'n_estimators':[200,300, 500,700]	95.27%	'max_features':'sqrt', 'n_estimators': 700
	6	'max_features': ['sqrt', 'log2'], 'n_estimators': [500,1000,1500]	95.21%	'max_features':'sqrt', 'n_estimators': 500
	7	'max_features': ['sqrt', 'log2'], 'n_estimators': [10,100,1000, 2000],	95.23%	'max_features':'log2', 'n_estimators': 2000
Decision Trees	1	'criterion': ['gini', 'entropy'], 'max_depth': [5, 10, 15, 20, 25, 30, 35, 40, 45, 50], 'max_leaf_nodes': [50, 100, 150,200, 250, 300, 350, 400, 450, 500]	93.18%	'criterion':'entropy', 'max_depth':30, 'max_leaf_nodes':450
K-Nearest Neighbors	1	'n_neighbors': [5,10,15,20,25], 'p': [2,1], 'metric': ['str','callable', 'minkowski']	90.50%	'metric':'minkowski', 'n_neighbors':5, 'p': 1
Multilayer Perceptron	1	'hidden_layer_sizes': [(5),(10), (15),(20)], 'activation' : ['logistic', 'tanh', 'relu'], 'solver' : ['lbfgs','sgd', 'adam'], 'alpha' : [0.01,1e-6, 1e-2], 'learning_rate' : ['constant', 'invscaling','adaptive'], 'max_iter' : [100, 500, 1000,2000]	98.37%	'activation': 'tanh', 'alpha': 0.01, 'hidden_layer_sizes': (15,), 'learning_rate': 'constant', 'max_iter': 1000, 'solver': 'adam'
	2	'hidden_layer_sizes': [(10), (20),(30), (40),(50),(60),(70),(80), (90),(100)], Same other parameters above	97.44%	'activation': 'tanh', 'alpha': 0.01, 'hidden_layer_sizes': 20, 'learning_rate': 'adaptive', 'max_iter': 1000, 'solver': 'adam'
	3	'hidden_layer_sizes': [(50), (100),(50,50), (100,50), (100,100)], Same other parameters above	98.21%	'activation': 'tanh', 'alpha': 0.01, 'hidden_layer_sizes': 50, 'learning_rate': 'adaptive', 'max_iter': 1000, 'solver': 'adam'

B. Performance of Parameters Tuning

This work used the Grid Search approach to tune the parameters of Multilayer Perceptron, Random Forest, Decision Trees, and KNearest Neighbors. It has the objective of increasing the quality of the classification by identifying the best parameters. For that, it is primordial to study the evaluation results with and without hyperparameter optimization. Based on Table II and Table III findings, we witness that all metrics of the four algorithms have been properly improved after the tuning operation in terms of accuracy, precision, recall, F1 score, and Kappa metric.

C. Performance of Data Sampling

Once parameter tuning by Grid Search was done, we applied the sampling technique using SMOTE-Tomek. From Table IV, the performance of all classifiers has increased when data balancing has been performed compared to that without sampling in Table III. We obtained satisfying outcomes in terms of accuracy, precision, recall, F1 score and Kappa metric, which means that the model was able to better predict the positive values. The classification task was then handled properly.

TABLE II. CLASSIFICATION PERFORMANCE WITHOUT PARAMETERS TUNING

Algorithm	Accuracy	Precision	Recall	F1 Score	Kappa
MLP	96.06%	95.89%	94.41%	96.07%	93.34%
RF	93.89%	92.38%	77.76%	93.54%	80.58%
DT	89.72%	76.01%	75.91%	89.72%	69.89%
KNN	87.59%	86.49%	50.21%	86.32%	56.85%

TABLE III. CLASSIFICATION PERFORMANCE WITH PARAMETERS TUNING AND WITHOUT DATA SAMPLING

Algorithm	Accuracy	Precision	Recall	F1 Score	Kappa
MLP	97.44%	97.12%	97.66%	97.38%	96.07%
RF	96.02%	92.47%	95.75%	93.65%	81.01%
DT	92.34%	89.60%	89.46%	92.24%	76.35%
KNN	91.88%	89.11%	90.75%	86.44%	57.07%

D. Confusion Matrix

We determined confusion matrix measures for each classifier with sampling and parameter tuning. In Tables V–VIII, TP indicates the number of delayed flights that were correctly classified, and FP shows the ones that were wrongly assigned. Similarly, TN indicates the number of non-delayed (on-time) flights that were correctly classified, and FN shows the ones that were wrongly assigned. The results show that parameter tuning and data sampling have successfully generated the best classifier, which was the Multilayer Perceptron with an accuracy of 98.72% and a higher number of correctly classified flights.

TABLE IV. CLASSIFICATION PERFORMANCE WITH DATA SAMPLING

Algorithm	Accuracy	Precision	Recall	F1 Score	Kappa
MLP	98.72%	98.50%	98.69%	98.10%	98.01%
RF	96.80%	95.86%	96.07%	95.84%	91.38%
DT	93.01%	95.15%	91.20%	93.07%	86.90%
KNN	92.44%	90.95%	92.71%	91.61%	83.15%

TABLE V. CONFUSION MATRIX OF KNN

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8254	FP= 829
Actual YES (delayed)	FN= 649	TP= 8627

TABLE VI. CONFUSION MATRIX OF RANDOM FOREST

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8737	FP= 394
Actual YES (delayed)	FN= 384	TP= 8847

TABLE VII. CONFUSION MATRIX OF DECISION TREES

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8686	FP= 382
Actual YES (delayed)	FN= 860	TP= 8435

TABLE VIII. CONFUSION MATRIX OF MLP

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8968	FP= 172
Actual YES (delayed)	FN= 60	TP= 9170

E. Comparison of Performance

In Table IX, we compared the results of our proposed best model with other studies from the literature review. As we noticed, our proposed model was the best in terms of accuracy compared with the other researches. Existing studies from the literature review adopted only data extracted from the databases they explored. In our study, besides the attributes selected from BTS [1] and ICAO [28], we proposed and added new features that were never used in previous studies, to the best of our knowledge, namely Part of the day in departure and arrival and Type of aircraft. After that, the data were cleaned, transformed, balanced, and prepared for the training and testing processes.

The hyperparameter optimization gave a good performance and enhanced the accuracy of the proposed methods compared with those of existing studies in the literature.

TABLE IX. ACCURACY COMPARISON OF OUR PROPOSED MODEL WITH THAT OF LITERATURE STUDIES

References	Accuracy
Proposed Model	98.72%
[17]	94.35%
[14]	94.00%
[74]	91.31%
[73]	88.59%
[16]	83.07%
[18]	77.00%

X. CONCLUSION

As air demand develops year after year, flight delay has become an essential study topic. For that, experts and scientists examined aircraft delays from many viewpoints. This research aimed to predict flight arrival delays. We designed and incorporated three additional novel features to improve the effectiveness of the proposed model:

Departure Part of the day, Arrival Part of the day, and type of aircraft. So as to achieve optimal results, data sampling and parameters optimization were explored using SMOTE-Tomek and Grid Search tools. With the Multilayer Perceptron, the model attained the greatest accuracy of 98.72%.

Airport authorities, companies, and customers can utilize the suggested model as a way to estimate aircraft arrival delays. It also has the potential to be used by air traffic control service providers in making decisions. In reality, if they are notified of flight arrival schedules and potential delays ahead of time, they will be advised of traffic rush hours in order to prepare for the flight approach sequence in advance or call for reinforcement teams if it is obligatory.

The dataset only provides on-time statistics for non-stop national flights. In a future study, we plan to expand the dataset by incorporating stopovers and foreign flights. Finally, it could be useful to integrate other attributes into the research in order to improve the classification of the model.

Data Availability

The data used to support this study can be found in Transportation statistics, United States Department of Transportation, http://www.transtats.bts.gov/.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Hajar Alla wrote the paper, conceived and designed the analysis; Lahcen Moumoun collected and analyzed the data; Youssef Balouki directed and supervised the research; All authors read and approved the final manuscript.

Acknowledgment

This work was supported by Laboratory of Mathematics, Computer and Engineering sciences, Settat, Morocco.

References

[1] Bureau of Transportation Statistics. (June 1, 2022). [Online]. Available: http://www.transtats.bts.gov/

[2] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, pp. 2627–2636, 1998.

[3] R. Genuer and J. M. Poggi, *Random Forests*, Springer International Publishing, 2020.

[4] H. Nefeslioglu, E. Sezer, C. Gokceoglu, A. Bozkir, and T. Duman, "Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey," *Mathematical Problems in Engineering*, 901095, 2010.

[5] M. M. Kumbure, P. Luukka, and M. Collan, "A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean," *Pattern Recognition Letters*, vol. 140, pp. 172–178, 2020.

[6] H. Idris, J. P. Clarke, R. Bhuva, and L. Kang. (2001). Queuing model for taxi-out time estimation. [Online]. Available: http://dspace.mit.edu/bitstream/handle/1721.1/37322/TaxiOutModel.pdf

[7] Y. Tu, M. O. Ball, and W. S. Jank. "Estimating flight departure delay distributions—A statistical approach with long-term trend and short-term pattern," *Journal of the American Statistical Association*, vol. 103, no. 481, pp. 112–125, 2008.

[8] H. Alla, L. Moumoun, and Y. Balouki, "A multilayer perceptron neural network with selective-data training for flight arrival delay prediction," *Scientific Programming*, pp. 1–12, 2021

[9] Z. W. Zhong, D Varun, and Y. J. Lin, "Studies for air traffic management R&D in the ASEAN-region context," *Journal of Air Transport Management*, vol. 64, pp. 15–20, 2017.

[10] K. B. Nogueira, P. H. Aguiar, and L. Weigang, "Using ant algorithm to arrange taxiway sequencing in airport," *International Journal of Computer Theory & Engineering*, vol. 6, 857361, 2014.

[11] Y. Jiang, X. Xu, H. Zhang, and Y. Luo, "Taxiing route scheduling between taxiway and runway in hub airport," *Mathematical Problems in Engineering*, 2015.

[12] E. Mueller and G. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Proc. AIAA's Aircraft Technology, Integration, and Operations (ATIO) Technical Forum*, 2002, p. 5866.

[13] A. Achenbach and S. Spinler, "Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights," *Operations Research Perspectives*, vol. 5, pp. 265–279, 2018.

[14] A. M. Kalliguddi and A. K. Leboulluec, "Predictive modeling of aircraft flight delay," *Universal Journal of Management*, vol. 5, pp. 485–491, 2017.

[15] R. Bhuvaneshwari, S. Elakiya, K. Hemavathi, S. Manisha, and S. Delhirani, "Flight delay prediction using supervised machine learning," *Computer Science*, 2019.

[16] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Cost-sensitive prediction of airline delays using machine learning," in *Proc. 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, IEEE, 2017, pp. 1–8.

[17] B. Thiagarajan, L. Srinivasan, A. V. Sharma, D. Sreekanthan, and V. Vija-yaraghavan, "A machine learning approach for prediction of on-time performance of flights," in *Proc. 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, IEEE, 2017, pp. 1–6.

[18] N. Etani, "Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data," *Journal of Big Data*, vol. 6, pp. 1–17, 2019.

[19] W. Wu, K. Cai, Y. Yan, and Y. Li, "An improved SVM model for flight delay prediction," in *Proc. 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, IEEE, 2019, pp. 1–6.

[20] N. Chakrabarty, T. Kundu, S. Dandapat, A. Sarkar, and D. K. Kole, "Flight arrival delay prediction using gradient boosting classifier," in *Proc. Emerging Technologies in Data Mining and Information Security*, Springer, 2019, pp. 651–659.

[21] N. Kuhn and N. Jamadagni, "Application of machine learning algorithms to predict flight arrival delays," in *Proc. AUTUMN 2017*, 2017.

[22] M. Zoutendijk and M. Mitici, "Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem," *Aerospace*, vol. 8, 152, 2021.

[23] D. Truong, "Using causal machine learning for predicting the risk of flight delays in air transportation," *Journal of Air Transport Management*, vol. 91, 101993, 2021.

[24] H. Alla, L. Moumoun, and Y. Balouki, "A multilayer perceptron neural network with selective-data training for flight arrival delay prediction," *Scientific Programming*, 5558918, 2021.

[25] International Civil Aviation Organization (ICAO), *Safety Management Manual (SMM)*, 1st ed. 2006.

[26] H. Alla, L. Moumoun, and Y. Balouki, "Flight arrival delay prediction using supervised machine learning algorithms", *Intelligent Systems in Big Data, Semantic Web and Machine Learning*, pp. 231–246, 2021.

[27] Federal Aviation Administration, "Aircraft performance," in *Pilot Hand Book*, 2022, ch. 11.

[28] The International Civil Aviation Organization (ICAO). [Online]. Available: https://www.icao.int

[29] L. Moreira, C. Dantas, L. Oliveira, J. de A. Soares, and E. S. Ogasawara, "On evaluating data preprocessing methods for machine learning models for flight delays," in *Proc. 2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[30] H. Shamsudin, U. K. Yusof, A. Ayalakshmi, and M. N. A. Khalid, "Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card

fraudulent transaction dataset," in *Proc. 2020 IEEE 16th International Conference on Control & Automation (ICCA)*, IEEE, 2020, pp. 803–808.
[31] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *Proc. 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, IEEE, 2016, pp. 1–6.
[32] S. Garcia, J. Luengo, and F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," *Knowledge Based Systems*, vol. 98, 129, 2016.
[33] B. Jonathan, P. H. Putra. and Y. Ruldeviyani, "Observation imbalanced data text to predict users selling products on female daily with smote, tomek, and smotetomek," in *Proc. 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, IEEE, 2020, pp. 81–85.
[34] M. Zeng, B. Zou, F. Wei, X. Liu, and L. Wang, "Effective prediction of three common diseases by combining smote with tomek links technique for imbalanced medical data," in *Proc. 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, IEEE, 2016, pp. 225–228.
[35] S. Oza, S. Sharma, H. Sangoi, R. Raut, and V. Kotak, "Flight delay prediction system using weighted multiple linear regression," *International journal of Engineering* and *Computer Science*, vol. 4, pp. 11668–11677, 2015.
[36] M. Tielrooij, C. Borst, M. M. V. Paassen, and M. Mulder, "Predicting arrival time uncertainty from actual flight information," in *Proc. the 11th USA/Europe Air Traffic Management Research and Development Seminar*, 2015, pp. 577–586.
[37] A. Cavcar and M. Cavcar, "Impact of aircraft performance characteristics on air traffic delays," *Turkish Journal of Engineering and Environmental Sciences*, vol. 28, pp. 23–24, 2004.
[38] International Civil Aviation Organization (ICAO), *Global Air Traffic Management Operational Concept*, 1st ed. 2005.
[39] The International Air Transport Association (IATA). The importance of night flights at Heathrow. [Online]. Available: https://www.iata.org/en/iata-repository/publications/economicreports/theimportanceofnightflightsat-heathrow/
[40] Federal Aviation Administration (FAA), *Airplane Flying Handbook (FAAH80833B)*, 2022, ch. 10.
[41] E. Co, K. B. Gregory, J. M. Johnson, and M. R. Rosekind, *Crew Factors in Flight Operations*, no. NASA/TM-1999-208799, 1999.
[42] European Aviation Safety Agency (EASA). Notice of proposed amendment (NPA) NO 201014A. [Online]. Available: https://www.easa.europa.eu/
[43] N. Silver. *Fly Early, Arrive on Time*. [Online]. Available: https://fivethirtyeight.com/features/flyearlyarriveontime/.2014
[44] *World Date, Time and Time Zones*. [Online]. Available: https://www.timeanddate.com
[45] D. Mladenovic, S. Jankovic, S. Zdravkovic, S. Mladenovic, and A. Uzelac, "Night traffic flow prediction using k-nearest neighbors algorithm," *Operational Research in Engineering Sciences: Theory and Applications*, vol. 5, pp. 152–168, 2022.
[46] S. Talebi, J. Waczak, B. A. Fernando, A. Sridhar, and D. J. Lary, "Data-driven EEG band discovery with decision trees," *Sensors*, vol. 22, 3048, 2022.
[47] L. Rokach and O. Maimon, "Decision trees," in *Data Mining and Knowledge Discovery Handbook*, Springer, 2005, pp. 165–192.
[48] J. L. Speiser, M. E. Miller, J. Tooze, and I. P. Edward, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, pp. 93–101, 2019.
[49] L. Breiman, "Random forests," in *Machine Learning*, 2001, vol. 45, pp. 5–32.
[50] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2001.
[51] Y. Qi, "Random forest for bioinformatics," in *Ensemble Machine Learning*, Springer, 2012, pp. 307–323.
[52] C. Han, T. Ma, G. Xu, S. Chen, and R. Huang, "Intelligent decision model of road maintenance based on improved weight random forest algorithm," *International Journal of Pavement Engineering*, vol. 23, pp. 985–997, 2022.
[53] M. Belgiu and L. Dragut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.
[54] N. Farnaaz and M. Jabbar, "Random forest modeling for network intrusion detection system," *Procedia Computer Science*, vol. 89, pp. 213–217, 2016.
[55] M. A. Khan, M. I. Shah, M. F. Javed, M. I. Khan, S. Rasheed, M. El-Shorbagy, E. R. El-Zahar, and M. Malik, "Application of random forest for modelling of surface water salinity," *Ain Shams Engineering Journal*, vol. 13, 101635, 2022.
[56] R. M. Ehsan, S. P. Simon, and P. Venkateswaran, "Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron," *Neural Computing and Applications*, vol. 28, pp. 3981–3992, 2017.
[57] F. A. A. Soares, E. L. Flores, C. D. Cabacinha, G. A. Carrijo, and A. C. P. Veiga, "Recursive diameter prediction for calculating merchantable volume of eucalyptus clones without previous knowledge of total tree height using artificial neural networks," *Applied Soft Computing*, vol. 12, pp. 2030–2039, 2012.
[58] ScikitLearn. Metrics and scoring quantifying the quality of predictions. [Online]. Available: http:scikitlearn.org
[59] M. M. Ramadhan, I. S. Sitanggang, F. R. Nasution, and A. Ghifari, "Parameter tuning in random forest based on grid search method for gender classification based on voice frequency," *DEStech Transactions on Computer Science and Engineering*, vol. 10, 2017.
[60] D. Stathakis, "How many hidden layers and nodes?" *International Journal of Remote Sensing*, vol. 30, pp. 2133–2147, 2009.
[61] P. Coulibaly, F. Anctil, and B. Bobée, "Hydrological forecast by artificial neural networks: State of the art," *Canadian Journal of Civil Engineering*, vol. 26, 293–304, 1999. (in French)
[62] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.
[63] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feed forward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
[64] O. Kisi, "Daily river flow forecasting using artificial neural networks and auto-regressive models," *Turkish Journal of Engineering and Environmental Sciences*, vol. 29, pp. 9–20, 2005.
[65] M. R. Zadeh, S. Amin, D. Khalili, and V. P. Singh, "Daily outflow prediction by multi-layer perceptron with logistic sigmoid and tangent sigmoid activation functions," *Water Resources Management*, vol. 24, pp. 2673–2688, 2010.
[66] M. Rezaeian-Zadeh, H. Tabari, and H. Abghari, "Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions," *Arabian Journal of Geosciences*, vol. 6, pp. 2529–2537, 2013.
[67] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35–62, 1998.
[68] P. H. Talaee, "Multilayer perceptron with different training algorithms for streamflow forecasting," *Neural Computing and Applications*, vol. 24, pp. 695–703, 2014.
[69] S. Tamura and M. Tateishi, "Capabilities of a four-layered feedforward neural network: Four layers versus three," *IEEE Transactions on Neural Networks*, vol. 8, pp. 251–255, 1997.
[70] G. B. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Transactions on Neural Networks*, vol. 14, pp. 274–281, 2003.
[71] D. Stathakis, "How many hidden layers and nodes?" *International Journal of Remote Sensing*, vol. 30, pp. 2133–2147, 2009.
[72] J. S. Sonawane and D. Patil, "Prediction of heart disease using multilayer perceptron neural network," in *Proc. International Conference on Information Communication and Embedded Systems (ICICES2014)*, IEEE, 2014, pp. 1–6.
[73] P. Stefanovič, R. Štrimaitis, and O. Kurasova, "Prediction of flight time deviation for Lithuanian airports using supervised machine learning model," *Computational Intelligence and Neuroscience*, 2020.
[74] J. Qu, S. Wu, and J. Zhang. "Flight delay propagation prediction based on deep learning," *Mathematics*, vol. 11, no. 3, 494, 2023.