

Research Article

A Multilayer Perceptron Neural Network with Selective-Data Training for Flight Arrival Delay Prediction

Hajar Alla , Lahcen Moumoun , and Youssef Balouki 

Laboratory of Mathematics, Computer and Engineering Sciences, Mathematics and Computer Science Department, Faculty of Science and Techniques, Hassan First University of Settat, Settat, Morocco

Correspondence should be addressed to Hajar Alla; allahaajar@gmail.com

Received 15 February 2021; Revised 12 April 2021; Accepted 29 May 2021; Published 15 June 2021

Academic Editor: Jianping Gou

Copyright © 2021 Hajar Alla et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Flight delay is the most common preoccupation of aviation stakeholders around the world. Airlines, which suffer from a monetary and customer loyalty loss, are the most affected. Various studies have attempted to analyze and solve flight delays using machine learning algorithms. This research aims to predict flights' arrival delay using Artificial Neural Network (ANN). We applied a MultiLayer Perceptron (MLP) to train and test our data. Two approaches have been adopted in our work. In the first one, we used historical flight data extracted from Bureau of Transportation Statistics (BTS). The second approach improves the efficiency of the model by applying selective-data training. It consists of selecting only most relevant instances from the training dataset which are delayed flights. According to BTS, a flight whose difference between scheduled and actual arrival times is 15 minutes or greater is considered delayed. Departure delays and flight distance proved to be very contributive to flight delays. An adjusted and optimized hyperparameters using grid search technique helped us choose the right architecture of the network and have a better accuracy and less error than the existing literature. The results of both traditional and selective training were compared. The efficiency and time complexity of the second method are compared against those of the traditional training procedure. The neural network MLP was able to predict flight arrival delay with a coefficient of determination R^2 of 0.9048, and the selective procedure achieved a time saving and a better R^2 score of 0.9560. To enhance the reliability of the proposed method, the performance of the MLP was compared with that of Gradient Boosting (GB) and Decision Trees (DT). The result is that the MLP outperformed all existing benchmark methods.

1. Introduction

In the last few years, air transport has experienced a high growth and demand mainly because of its comfort, speed, safety, and efficiency. The massive increase in air traffic has resulted in congestion in the airspace and airports leading to traffic delays. Flight delays are an inconvenience to airlines, airports, passengers, and aviation authorities. Delays occur due to mechanical and technical issues, slot restrictions, personnel labor strike, runway, airport or airspace lack of capacities, the hub status of the airports, or poor weather. The authors in [1] concluded in their study that thunderstorms are responsible for flight delay duration. As consequences, a delayed flight can be costly to passengers by arriving late or cancelling their personal scheduled

appointments and to airlines by a large economic and customer loyalty loss. It also affects the environment by increasing fuel consumption and gas emission [2]. Hence, a delay prediction turns out very important. International Civil Aviation Organization (ICAO) has enabled a program called Air Traffic Flow Management (ATFM) with the objective of ensuring that the traffic volume is compatible with the capacities declared by aviation authorities in order to reduce ground and en-route delays. Another application of traffic management is the Free Route Airspace (FRA) concept which consists of using the shortest routes possible in order to reduce flight time, CO₂ emissions, and fuel waste. Moreover, several other models have been developed to solve delays problem based on probability, statistics, graph and network representations, operational research studies,

and so on. Ours is based on machine learning algorithms. Recently, the use of artificial neural network (ANN) has become widely recommended in different fields: medical applications, pharmaceutical sciences, engineering, banking, social media, and so on. One of its greatest advantages is the capability to rapidly learn from its environment (data, tasks, and so on). It is also able to identify redundant and noise variables during training [3]. To predict traffic arrival delays using ANN, we chose to apply the multilayer perceptron (MLP) because of its reliability and better performance. Unlike other statistical techniques, MLP can model highly nonlinear functions and has been shown to be effective when presented with new unseen data. The MLP has been applied to a wide variety of tasks such as prediction, function approximation, or pattern classification [4].

Only few studies related to flight delay prediction using MLP were conducted, but to the best of our knowledge, no one has adopted MLP-based selective-data training. The proposed method was applied to predict the arrival delay of United States domestic flights in the year 2018. Open-data sourced from the Bureau of Transportation Statistics were used [5]. To enhance the performance of the proposed prediction model, a selective procedure which consists of keeping only the delayed flight data was employed separately and compared with the traditional procedure. Hyperparameter optimization has proven better results on the performance of machine learning algorithms over manual search which tends to be annoying and time-consuming [6]. As a search method to find the best parameters, we used grid search technique which examines the entire search space. Each dimension of a grid represents a successive or discrete variable to be optimized after several trial-and-error processes [7]. In order to enhance the reliability of the proposed method, we attempted to evaluate the model using two other regression algorithms, namely, gradient boosting (GB) and decision trees (DT).

The main contributions of the proposed work can be summarized as follows:

- (i) The proposed model can be utilized to estimate and predict flight arrival delays.
- (ii) It can be extended to other applications only by adapting or changing the data. It can be used to predict departure delays instead of arrival. Air transport can be replaced by maritime or railway fields by predicting ship or train departure or arrival delays and so on.
- (iii) We introduce a novel technique, namely, selective training to help the system in focusing on relevant data and avoid overfitting. Data were refined, transformed, and prepared for the learning using preprocessing and cleaning techniques.
- (iv) To better determine the architecture of the network, adjust, and tune the hyperparameters, grid search technique was adopted. As a consequence, the best configuration of our model was generated.
- (v) To ensure the reliability and efficacy of the proposed method, results are evaluated and compared with

some existing systems from the literature. To prove much more the effectiveness of the study, other machine learning models such as gradient boosting and decision trees were applied and compared with the multilayer perceptron performance. The complexity of the traditional training and the novel technique was calculated and compared.

The rest of the paper is structured as follows: Section 2 presents a preview of the artificial neural network. Section 3 shows a brief review of previous research studies related to flight delay prediction with machine learning algorithms. The methodology proposed in this study is described in Section 4, followed by the experimental results of the predictive model in Section 5. Section 6 defines the computational complexity of both traditional and selective training models. Section 7 provides a conclusion and suggests future works.

2. Artificial Neural Network (ANN)

As a simulation and imitation of the brain neural network, ANN is a mathematical structure that necessitates less formal statistical training to develop. It has the advantage of being able to detect complex nonlinear relationships between independent and dependent variables and every possible interaction between predictor variables. It can be elaborated using multiple different training algorithms [8]. ANN performs better with large dataset unlike SVM and random forest which show a high accuracy and precision with smaller data set [9]. It has the power to be adjusted in order to lower its error without being sure that the error could not be lower still [10]. The authors in [11] considered ANN among the most powerful machine learning algorithms for time-series predictions. A large-scale empirical comparison between ten supervised learning methods demonstrated that neural networks are more competitive and efficient than boosting, random forests, bagging, and support vector machines [12].

ANN applications are endless. In classification, ANN is used for image and speech recognition, abnormal event detection, customer purchasing patterns, and so on. In regression, it is applied for stock market predictions, forecasting applications, real-time optimization, model-predictive control, and so on. The authors in [13] used ANN to predict the fatigue crack growth and propagation in both short and long crack regimes. The authors in [14] applied ANN to detect and classify COVID-19 disease from X-ray images using capsule networks. The authors in [15] chose a multilayer perceptron (MLP) neural network to assess the safety of earthquake hazard and identify vulnerable buildings. Other applications of ANN are highlighted further in the paper.

3. Literature Review

Researchers have carried out flight delay prediction studies from different perspectives and using several modeling methods: statistical analysis, probabilities, queuing modeling, simulation techniques, network representations, and

machine learning. Unlike these traditional methods which generally have proven to be weak, slow, and limited, machine learning algorithms become more popular in leading to higher accuracy and dealing with a huge amount of data. Several research studies have been conducted on flight delay prediction using machine learning (ML) algorithms.

3.1. Neural Network Models. The authors in [16] presented a comprehensive study of traffic delays using several machine learning models. The MLP model has proven 89.07% accuracy compared with the convolutional neural network (CNN) which showed a slightly better prediction accuracy of 89.32%. The authors in [17] applied the ANN to predict airborne delay due to air traffic control using actual operation data observed by radars. The performance of the ANN was compared with that of the queuing analysis method. ANN was able to predict the delay average value but unable to learn the propagation of the delay compared with the other method. The authors in [18] selected ANN to create a prediction model of air delays in the route linking São Paulo to Rio de Janeiro. Random search technique was utilized for hyperparameterization of the network. The results showed an accuracy superior of 90%. The researchers [19] introduced a new type of multilevel input layer ANN capable of handling nominal variables in order to predict the delay of incoming flights at JFK airport. The authors in [20] applied decision trees, random forest, multilayer perceptron, and different sampling techniques to predict flight delays. The best model was the MLP with 85% accuracy. The authors in [21] implemented a supervised machine learning model that predicts delay deviation time of new Lithuanian airports flights. They used the grid search technique with seven algorithms: probabilistic neural network, multilayer perceptron, decision trees, random forest, tree ensemble, gradient boosted trees, and support vector machines. MLP has shown an accuracy of 96.02% for departures but has problems with arrivals with 47% accuracy.

3.2. Other ML Models. The authors in [22] provided a cost-sensitive delay prediction model using supervised machine learning algorithms such as decision trees, random forest, Adaboost, and kNN. Authors [23] developed a flight delay predictive model by combining multilabel random forest classification and approximated the delay propagation model. They demonstrated that feature selection had a better performance than using all the features from the dataset. The study in [24] aimed to predict departure and arrival flight delays in an individual airport using the gradient boosted decision tree algorithm. It showed better performance as compared with other methods. To estimate airline delays, the authors in [25] utilized binary supervised and unsupervised machine learning classification algorithms. Paper [26] consists of a two-stage predictive model (classification and regression) employing supervised machine learning algorithms in order to predict on time performance of flights. The authors in [27] utilized decision tree, logistic regression, and neural networks classifiers to predict flight arrival delays for the year 2015 in the United States. An accuracy of 91%

was achieved by all the three classifiers. Authors in [28] applied multiple linear regression, decision trees, and random forest algorithms using R-studio to predict and identify the critical parameters responsible for flight delay. The authors in [29] created a Bayesian network model to analyze flight departure delay in a large hub airport. The model proved to have a high convergence and accuracy rates. The authors concluded that parameters learning can reflect departure delay. The authors in [30] used an improved SVM model, KNN, and random forest to predict flight delays with an R^2 , respectively, equal to 0.71, 0.14, and 0.09. The authors in [31] compared deep belief network combined with support vector regression (DBN-SVR) results to those of k-nearest neighbors (kNN), support vector machine (SVM), and linear regression (LR). The coefficient of determination R^2 was 0.93, 0.87, 0.87, and 0.82, respectively. Table 1 represents a summary of prior studies for flight delay prediction.

The present work stands out from the previous academic literature by applying two different approaches to predict flight arrival delay in the United States context. Despite being reliable and presenting a good performance, only few studies have employed MLP to predict flight delays unlike other machine learning methods which are very common and are applied most of the time. This lack in the existing literature encouraged us to use the MLP in the same context but treated differently. This study is, to the best of our knowledge, the first attempt to apply the selective training in the flight prediction field. Instead of blindly and manually testing MLP parameters, we adopted the grid search technique to find the best architecture of the model.

4. Proposed Methodology

4.1. Motivation of the Proposed Method. Flight delay prediction is crucial not only for passengers and airlines but for every player in the aviation and transportation systems. The proposed model is able to predict flight arrival delays in United States context. By establishing a model with a high ability of predicting air traffic delays, airlines will be able to inform their passengers of the delay in advance. Moreover, the proposed model, if utilized appropriately, can enable aviation stakeholders to analyze, study, and reduce occurring delays by finding the best course of actions to take during the decision-making process, without necessarily having to invest in airport infrastructure and installations.

A survey (survey link: <https://t.ly/VO4Q>) was established, before carrying out this study, in order to identify the importance of predicting and minimizing flight delays from the point of view of pilots, air traffic controllers, airport personnel, and passengers from different countries: Morocco, Egypt, Sweden, and United States. All target population confirmed having experienced flight delays that engendered many consequences such as frustrations, delay propagation, missing the next flight in case of a stop-over, and not being able to check-in in time for non 24 hours reception hotels at destination. They all agreed that predicting air traffic delay is important. Pilots claimed that flight delays lead to a work under pressure, stress, and

TABLE 1: Summary of flights delay prior studies.

Reference	Objective	Study case	Methods	Features	Results
[21]	Delay deviation time prediction	Lithuanian airports	Probabilistic neural network (PNN) MLP, decision trees (DT), random forest (RF), tree ensemble (TE), GB trees, SVM	Flight date, flight number, company, departure airport, destination airport, temperature, sky information, wind speed, wind angle, visibility, scheduled time, and classes	DT, RF, TE = 62% for arrival delays and 96.02% for departure. PNN, MLP = 96.02% on the departure dataset. MLP = 47% on arrival dataset. GB trees = for arrivals 88.59%; for departures 96.02%. SVM = 32.70% for arrivals and 82.88% for departures
[16]	A comprehensive study of traffic delays	US flights	MLP, CNN	—	MLP: 89.07%, CNN: 89.32%
[23]	Flight delay prediction	US main airports	RF classifier, delay propagation model	BTS dataset features, NOAA meteorological features, scheduled departures, arrivals, direction, airport	Arrival accuracy: 0.85 departure accuracy: 0.82
[30]	Flight departure delay prediction	Beijing Capital International Airport (PEK)	Improved SVM, KNN and RF	Delay time, airline, scheduled departure time, scheduled arrival airport, type of aircraft, departure direction, scheduled flight duration, order of flight in task ring, duration of ground service, median of flight's historical delay time, standard deviation of the flight's historical delay time	R^2 : improved SVM = 0.71, KNN = 0.14, RF = 0.09
[31]	Flight delay prediction for commercial air transport	Beijing Capital International Airport (PEK)	DBN-SVR, kNN, SVM, LR	Airlines, air traffic control in formation, types of aircrafts, check-in and closing time of flights, boarding and taking-off time, parking area, gate locations, runway, and fuel filling time	R^2 : DBN-SVR = 0.93, kNN = 0.87, SVM = 0.87, LR = 0.82
[18]	Prediction of air delays	The route linking São Paulo (Congonhas) to Rio de Janeiro (Santos Dumont)	ANN with random search technique	—	Accuracy superior of 90%
[20]	Flight delays prediction	Hartsfield-Jackson Atlanta International Airport	Decision trees, random forest, MLP	Flight data, weather data, airplane info, delay propagation information	MLP: 85.63% accuracy, RF: 84.26%, DT: 83.06%
[22]	A cost-sensitive delay prediction	US domestic flights	Decision trees, random forest, adaboost, and kNN	Origin, destination, quarter of year, month, day of month, day of week, scheduled departure and arrival times, arrival delay indicator, NOAA meteorological data	RF = 82.75%, ada = 83.07%. kNN = 80.6%. DT = 82.49%

TABLE 1: Continued.

Reference	Objective	Study case	Methods	Features	Results
[17]	Airborne delay predict due to air traffic control	Tokyo Airport	ANN, queuing analysis method	Departure time, estimated time of arrival, estimated time to enter sector, sector entrance point, forecasted wind, number of aircraft in sector, sector entrance time interval	ANN RMSE = 183.7 seconds, queuing RMSE = 230.4 seconds
[24]	Departure and arrival flight delays prediction	An individual airport case in US	Gradient boosted decision tree algorithm	Year, month, day of month, day of week, carrier, origin airportID, dest airportID, CRSDepTime, DepDelay, DepDel15, CRSArrTime, ArrDelay, ArrDel15, and Cancelled	Arrival: 92.31%. departure: 94.85%
[26]	Prediction of on time flights performance	Domestic flights of the USA	GB, RF, adaboost, extra trees and MLP classifiers and regressors	Airline ID, flight number, origin airport ID, destination airport ID, year, quarter of year, month, day of month, day of week, scheduled departure time, scheduled arrival time, wind direction, humidity, pressure, temperature	Between 85% and 94% depending on whether it is a classifier or regressor
[27]	Flight arrival delays prediction	United States, 2015	Decision tree, logistic regression neural networks classifiers	Month, day, day of the week, flight number, origin airport, destination airport, scheduled departure, departure delay, taxi-out, distance, scheduled arrival	91% accuracy
[28]	Flights delay prediction modeling	US domestic flights	Multiple linear regression, decision trees, random forest	Departure delay, taxi in, taxi-out, carrier delay, security delay, weather delay, late aircraft delay, distance, and national air system delay	RMSE for DT = 26.5 minutes, MLR = 21.2 minutes, RF = 12.5 minutes
[19]	Incoming flights delays prediction	John F. Kennedy Airport	Multilevel input layer ANN	Day of month, day of week, code of origin, scheduled departure time, actual departure time, departure delay, scheduled arrival time, actual arrival time, arrival delay, carrier delay, weather delay, NAS delay, security delay, late aircraft delay	RMSE = 0.1366
[29]	Flight departure delay analysis	A large hub airport case	Bayesian network	Flight terminal number, airlines, flight task, airplane type, international (I) or domestic (D) flights, flight departure time duration, departure delay time	Accuracy is around 84.01% and 89.5% depending on the algorithm

concentration loss. Furthermore, flight delay prediction improves airline's benefits, and being on time reduces operational and financial risks. Air traffic controllers mentioned that delay propagation can be the worst possible scenario generating a flight accumulation and an air/ground density. Therefore, air traffic delay prediction is very important.

It is highly recommended by several organizations that safety is a major concern in aviation context. There are three primordial slogans for Air Traffic Management (ATM). In fact, the International Civil Aviation Organization (ICAO) has named in its ATM book the three slogans in order such as safety, regularity, and efficacy [32]. The Federal Aviation Administration (FAA) services (Federal Aviation Administration—Air Traffic Services, <https://www.faa.gov>) provide a safe, secure, and efficient management for the National Airspace System and international airspace assigned to U.S. control. National Aeronautics and Space Administration (NASA), in collaboration with the Office of Safety and Mission Assurance (OSMA), created a culture and an environment that keep safety in the forefront and as a priority. Since it has been agreed that aviation is a sensitive and delicate field, we decided to assure the safety by establishing a simple, smooth, unriskey but efficient predictive system. Also, using several and excessive features may lead to an overfitting. For this reason, we chose to keep our model simple and understandable.

To enforce and achieve the predictive system, we chose the MLP regressor for being effective and having a good performance compared with other traditional statistical techniques. It is a universal approximator that presents better efficiency for function approximation in high-dimensional spaces. Unlike conventional linear regression methods which suffer from the curse of dimensionality, the error convergence rate of the MLP is independent of the input dimensionality [33]. MLP is also known for being easy to implement, providing high-quality solid models with a low training time compared with more complex methods [34], which fulfills all we need to assure a safe and secure predictive model in aviation field.

As an alternative to traditional statistical modeling techniques, MLP has been applied in many scientific disciplines categorized as either prediction, function approximation, or pattern classification. In prediction, researchers chose MLP for wind speed forecasting [35], financial predictions [36], forecasting of stock prices [37], bacteria type prediction [38], temperature forecasting [39], short-term rainfall forecasting [40], and so on. In our case, the purpose of the proposed MLP-based model is to estimate flight arrival delays. However, it can be extended to other applications only by adapting or changing the data. It can be utilized to predict departures delays instead of arrivals. Air transport can be replaced by maritime or railway fields by predicting ship or train departure or arrival delays and so on.

4.2. Description of the Proposed Method. Data are the core of any machine learning algorithm. In the first stage, we collect historical flight data from Bureau of Transportation Statistics

(BTS). In order to extract appropriate information, features selection which consists of only considering the relevant features is applied. Collected data are generally noisy, incomplete, and redundant [41]. Since ours contain undesired information, cleaning and preprocessing techniques which involve correcting, transforming, replacing, and deleting data need to be performed at this stage. Then, we apply to the modified dataset either the traditional technique of selecting all data records or the selective training which consists of only accounting the delayed traffic. Next, we split the dataset into 70% for training and 30% for testing. The test set is used for the prediction. The model is trained with the multilayer perceptron regressor. Grid search is applied to generate the best hyperparameters and an optimized model. To evaluate the performance of the proposed model, accuracy, computational complexity, and error metrics are calculated and compared for both traditional and selective trainings. Gradient boosting (GB) and decision trees (DT) are used as benchmark methods to prove the efficiency of the MLP. The choice of GB and DT for the comparison will be defended in Section 5 of the paper. The flowchart of the proposed methodology is illustrated in Figure 1.

4.3. Data Source. The dataset used for the study is extracted from Bureau of Transportation Statistics (BTS). It contains more than 760 thousand samples of historical flight data from the 1st of January to the 31st of December 2018, in the United States. BTS database has proven to be reliable and full of statistical data saved since 1987. All the flights whose difference between scheduled and actual arrival times is 15 minutes or greater are considered delayed.

4.4. Features Selection. The data selected contained several features from which we have kept only relevant ones which have a high contribution to traffic delay. The features used in this project are resumed as follows:

- (i) Date of flight: the date in which the flight was performed
- (ii) Carrier: the airlines/company code
- (iii) Origin: airport of departure
- (iv) Destination: airport of arrival
- (v) CRS DEP: the scheduled departure time
- (vi) Actual DEP: the actual departure time
- (vii) DEP delay: the departure delay in minutes
- (viii) CRS ARR: the scheduled arrival time
- (ix) Distance: distance of the flight in miles
- (x) ARR delay: the arrival delay in minutes, which is our dependent variable

Logically, arrival and departure delays are highly correlated. A traffic that experiences a delay on departure will be surely delayed on arrival. It proves, according to [42], that congestion at destination airport is to a great extent originated at the departure airport. For this reason, we considered DEP delay as a contributive feature to arrival delays.

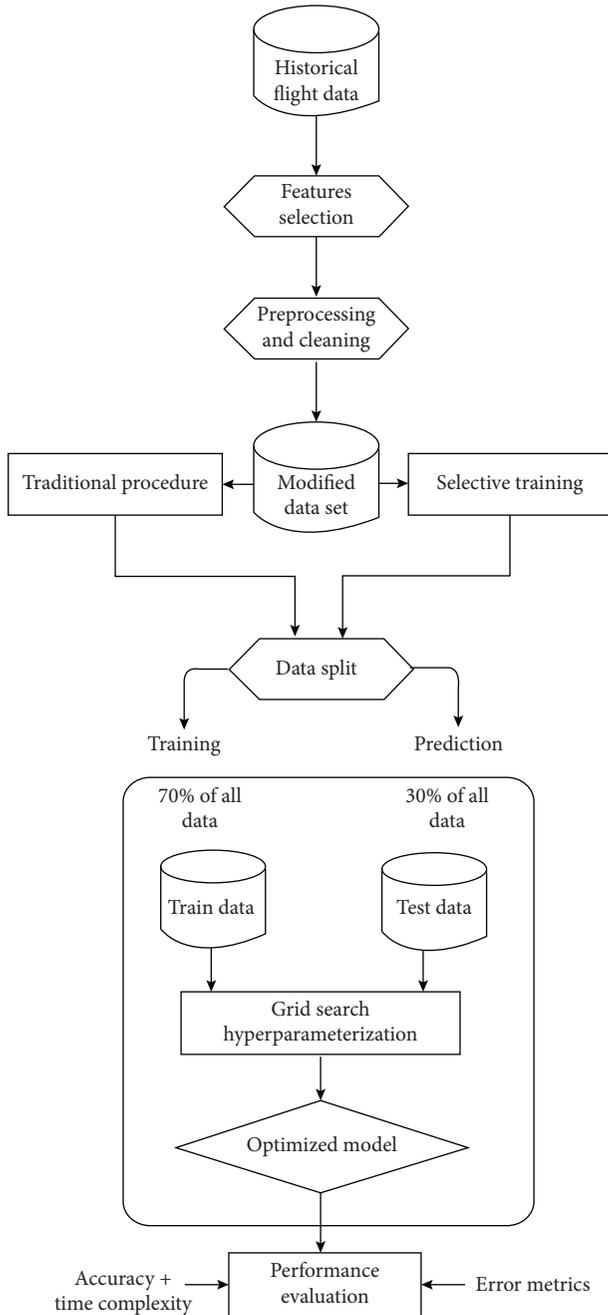


FIGURE 1: Flowchart of the proposed method.

The long-term flights are more likely to having delays due to the possible bad weather scenarios, multiple time zones travel, crew stress and fatigue, and so on. So, Distance is a relevant feature for our study.

4.5. Preprocessing and Cleaning. In the preprocessing, we eliminate unnecessary information and keep only relevant one to assure coherence. This process consists of preparing and formatting the data before training by the following:

- (i) Removing data records with missing values
- (ii) Eliminating errors and null values

- (iii) Removing duplicate data
- (iv) Converting categorical data to numeric

4.6. MLP Modeling and Hyperparameterization. Being the most common and popular network architecture in use today, MLP is a feedforward neural network composed of more than one perceptron [10]. It is applied generally in prediction, function approximation, or pattern classification fields. The authors in [15] used an optimized MLP model with three hidden layers (25, 15, 10) to predict the damage state of reinforced concrete buildings from the Duzce earthquake in Turkey. Results showed that the MLP model has a high accuracy in detecting most vulnerable buildings. Authors in [43] developed a set of multiple MLP neural networks with the backpropagation learning algorithm using an adaptive learning rate. The study was applied for thyroid disease diagnosis in the Internet of medical things in which the accuracy rate of 99% was achieved.

The MLP, compared with other traditional statistical techniques, has shown to be effective with unseen data and nonlinear systems especially in prediction applications [4]. Despite being reliable and presenting a good performance, only few studies have employed the MLP to predict flight delays. This lack in the existing literature encouraged us to use the MLP in our research with a parameters optimization. Besides, to the best of our knowledge, no one has adopted a MLP-based selective training.

Generally, most of machine learning algorithms achieve optimal results only if their parameters are tuned and adjusted properly [44]. To save search time and energy, we adopted grid search technique with the k-fold cross-validation; in our case, k was equal to three ($k = 3$). For the construction of the neural network in both traditional and selective procedures, we utilized Scikit-learn for being a well-maintained, comprehensive, and open-sourced machine learning package in Python [15]. For model training and testing, we applied the recommended 70–30 split. For the hyperparameterization, the following parameters were chosen: (a) hidden layer sizes: (50), (50, 50), (100); (b) activation function: tanh and ReLU; (c) solvers: stochastic gradient descent (Sgd) and Adam; and (e) L2 penalty (regularization term) parameter: 0.001, 0.05, and $1e - 10$. To obtain the prediction results, we followed the ANN process of multiplication, summation, and activation using the following equation:

$$y(t) = F\left(\sum_{i=0}^n w_i(t)x_i(t) + b\right), \quad (1)$$

where $x_i(t)$ is the input value (the independent variable/feature used to predict the output) in discrete time t where i goes from 0 to n inputs; $w_i(t)$ is the weight value representing the influence of input nodes on the output in discrete time t where i goes from 0 to n inputs; b is the bias that shifts the result of the activation function towards the positive or negative side; F is the transfer or activation function which decides whether a neuron should be activated or not; and $y(t)$ is the output value (the dependent variable) in discrete time t .

5. Experiments and Results

As an experimental dataset, we used more than 760 thousand samples containing US flight records from the 1st of January to the 31st of December, 2018. As we mentioned in Section 4, for model training and testing, we applied the recommended 70–30 split. We utilized Scikit-learn and Python for coding our program.

North American commercial airlines, such as American Airlines and United Airlines, have defined short haul flights as the flights where the route length or distance is shorter than 700 miles, long haul flights as being longer than 3000 miles, and medium haul flights as being in-between [45]. We classified the flights in three categories, namely, short haul flight, medium haul flight, and long haul flight depending on the route distance. In Figure 2, we plot the arrival delay against the flight category. We notice that the category long haul flight is clearly null in the figure. Since our data contain records for only domestic flights, it is obvious that the distance will be shorter than 3000 miles because the flights were performed inside the country US and not abroad. Hence, our data flights are either short or medium haul. From the results, we conclude that most of the flights delayed on arrival were classified as medium haul based on the distance. So, the longer the distance is, the higher the delay is (Figure 2).

25% from the data indicated that delayed flights on arrival had delays on departure too. Furthermore, 40% from the data demonstrated that the flights which had not been delayed on arrival had not been delayed on departure too which proves that arrival and departure delays are highly correlated. We plot the high correlation between the two features in Figure 3.

Two approaches have been adopted in our research. The best configuration of ANN for both selective training and traditional procedure was the one that presented 2 hidden layers with 50 nodes each (50, 50), ReLU as an activation function, Adam as a solver, and L2 penalty or alpha equals to $1e-10$. With the optimized configuration of the ANN, a multilayer perceptron network of 4 layers was trained in our study. The final is the output layer with a unique neuron giving the arrival delay. The final architecture of our MLP is shown in Figure 4.

The classic evaluation metric coefficient of determination R^2 , which is the measure of deviation between the regression line and the observed points [46], was applied to measure the performance of both traditional and selective procedures. We compare, in Table 2, the accuracy of our model to that of other studies from the literature review. Our model with the selective training achieved the best score of 95.60% followed by paper [18] in which a mean correct predictive capacity of 91.3% was obtained. Among the several machine learning algorithms applied in [16, 20], the best model was the MLP with an accuracy of 89.07% and 85%, respectively.

In order to verify the reliability of the proposed method, we compare the performance of the MLP to that of other benchmark methods, namely, gradient boosting

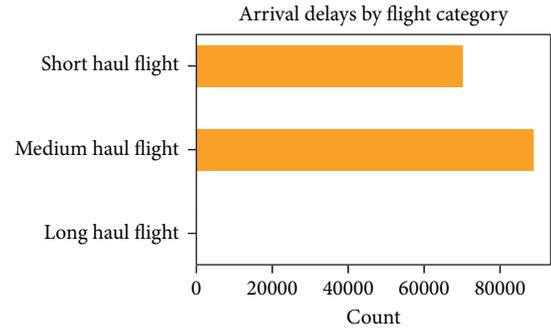


FIGURE 2: Plotting arrival delay against flight category.

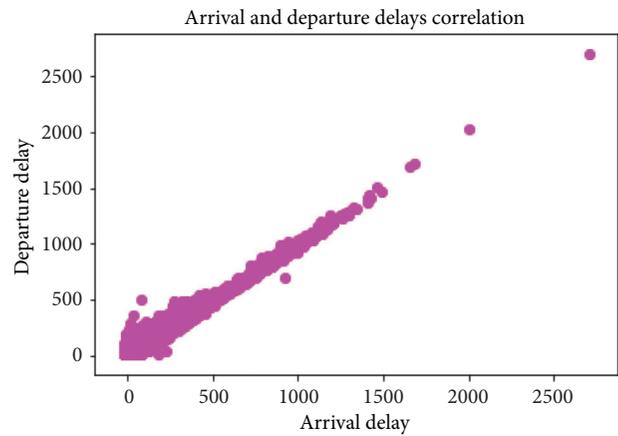


FIGURE 3: Correlation between departure and arrival delays.

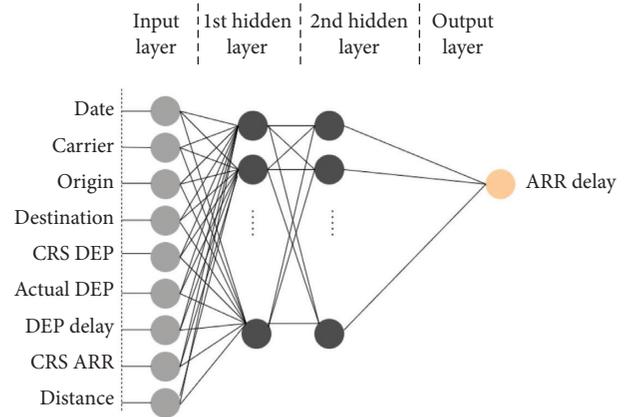


FIGURE 4: The MLP architecture of this study.

TABLE 2: Accuracy benchmark of different MLP models.

	Selective (%)	Traditional (%)	[16]	[18]	[20]
Accuracy (%)	95.60	90.48	89.07	91.3	85

(GB) and decision trees (DT). The gradient boosting model has benefited from being popular in many research areas and from the ability to handle overfitting,

missing, and noisy data. Unlike linear models, it does not require any statistical assumptions [47]. Decision tree models are quick to build and easy to interpret and understand. The predictions based on decision trees are efficient [48].

Mean absolute error (MAE), root mean squared error (RMSE), and median absolute error (MdAE) were used as evaluation metrics. MAE is the measure of how close forecast or predictions are to the eventual outcome [9]. RMSE is the square root of the mean of squares of all the errors while MdAE is the median of all absolute differences between the target and the prediction [49]. Table 3 shows the benchmark performance of the MLP proposed model with gradient boosting and decision trees. The results prove that the MLP outperformed the other algorithms in both training and testing processes. The evaluation metrics were quite similar for both training and testing steps which indicates that our model fits well. To examine the accuracy of the predictive model, we plot the high correlation between actual and predicted delay values in Figure 5.

From the results and evaluations above, the following can be easily deduced:

- (i) Arrival and departure delays are highly linked and correlated
- (ii) Distance and length of the flight are contributive to traffic delays
- (iii) An adjusted and optimized hyperparameters using grid search technique helped us choose the right architecture of the network
- (iv) The ANN-based MLP gave a high predictive arrival delay performance of 90.48%
- (v) The selective-data training with the MLP proved an increase in efficiency and a better performance of 95.60%
- (vi) The proposed model has outperformed all existing systems in Table 2
- (vii) The MLP achieved a better performance than gradient boosting and decision trees
- (viii) Evaluation metrics of testing are similar to those of training which proves that our model is good
- (ix) Actual and predicted values' high correlation indicates that our model fits well

6. Computational Complexity

Time complexity is the measure of how fast or slow an algorithm will perform depending on the input size. In order to compute the complexity of our model, the famous big O notation was applied and compared for both traditional and selective trainings.

6.1. Feedforward Pass. Following the ANN process of multiplication, summation, and activation, we have from layer m to n as follows:

TABLE 3: Evaluation metrics of the proposed MLP, GB, and DT.

Algorithm		R^2 score	RMSE	MAE	MdAE
Multilayer perceptron	Training	0.9545	18.62	12.42	8.82
	Testing	0.9560	18.48	12.40	8.87
Decision trees	Training	0.9176	24.94	16.95	12.89
	Testing	0.9101	26.25	16.97	12.89
Gradient boosting	Training	0.8914	28.63	19.27	13.04
	Testing	0.8831	29.93	19.34	13.04

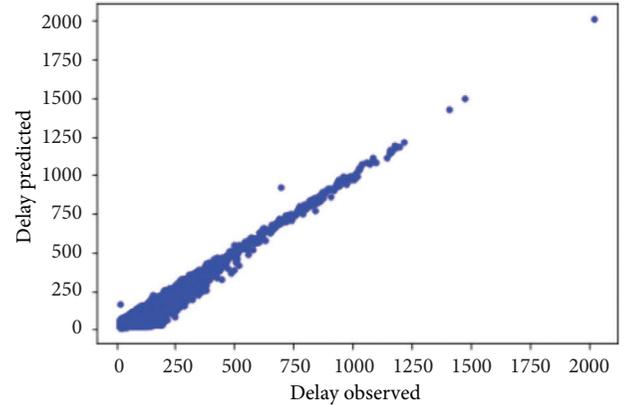


FIGURE 5: Observed and predicted delays correlation.

$$S_n = W_{nm} * X_m, \quad (2)$$

where S_n is the weighted sum of the weights W_{nm} and the inputs X_m .

If we apply the activation function, then we will have the following:

$$Y_n = F(S_n), \quad (3)$$

where Y_n is the output after the activation function F is fed to the computed value S_n .

The process will run $N-1$ times in the case we have N layers (input and output layers included). For the case of 4 layers, we will need 3 matrices to represent the weights: W_{nm} , W_{on} , and W_{lo} .

W_{nm} is a matrix with n rows and m columns containing the weights from layer m to layer n .

For t training examples, we have

$$S_{nt} = W_{nm} * X_{mt}. \quad (4)$$

Time complexity of the operation above is $\mathbf{O}(n * m * t)$.

$\mathbf{O}(n * t)$ is the time complexity of the activation function as follows:

$$Y_{nt} = F(S_{nt}). \quad (5)$$

In total, we have the following complexity: $\mathbf{O}(n * m * t + n * t) = \mathbf{O}(n * t * (m + 1)) = \mathbf{O}(n * m * t)$.

Following the same logic, the similar process will be applied when going from n to o : $\mathbf{O}(o * n * t)$, and from o to l : $\mathbf{O}(l * o * t)$.

So, in total, time complexity for feedforward propagation process is as follows: $O(n * m * t + o * n * t + l * o * t) = O(t * m * n * o * l)$.

6.2. Backpropagation Pass. Following the same process, time complexity of backpropagation is the same as feedforward pass.

For one epoch, it is equal to $O(t * m * n * o * l)$.

We multiply by number of iterations i (epochs): $O(i * t * m * n * o * l)$.

Another format of the neural network time complexity (neural network models (supervised)–MLP regressor–scikit-learn.org) which gives the same results can be used: $O(i * t * m * h^k * l)$.

6.2.1. Application. We consider t as the training samples, m as the number of features (input layer neurons), h as the number of nodes per hidden layer, k as the number of hidden layers, and l as the output layer neurons.

(1) *Traditional Training.* In this case, we have $t = 579880$, $m = 9$, $h = 50$, $k = 2$, and $l = 1$. So, the computational complexity of the traditional procedure is $O(i * 1304730 * 10^{04})$, where the max iteration number (epochs) i is equal to 200.

(2) *Selective Training.* In this case, we have $t = 127340$, $m = 9$, $h = 50$, $k = 2$, and $l = 1$. So, the computational complexity of the selective training is as follows: $O(i * 286515 * 10^{04})$ where the max iteration number (epochs) i is equal to 200.

We notice that the computational complexity of the selective training is way lower than the traditional procedure.

7. Conclusion

Reducing flight delays has been a major concern for airlines, airports, passengers, and aviation stakeholders in general. However, minimizing delay time is not such an easy topic. Hence, a traffic delay prediction turns out useful. Several researchers have tried to develop new models in order to increase the precision and accuracy of flight delays prediction. In this study, we proposed an artificial neural network (ANN) model based on supervised learning. After the hyperparameterization of the network by grid search technique was done, a multilayer perceptron (MLP) with an input layer, two hidden layers of 50 nodes each (50, 50), and an output layer was built. Departure delay and flight distance proved to be very contributive to flight delays. The experimental investigation showed that the highest score of 0.9560 has been obtained when the selective-data training was applied. The traditional training procedure has demonstrated a R^2 score of only 0.9048 compared with the other training. Time complexity of the two methods was computed by the famous big O notation and then compared. In order to boost and enhance the reliability and the efficacy of the proposed model, MLP results were compared with gradient boosting and decision trees. The MLP regressor has performed a better prediction in both training and testing

processes and must be given more attention in the next studies.

Our model treats flight arrival delays but it can be also applied to study flight departure delays. Our proposed architecture can be also used to predict train, TGV, metro, bus, ship arrival, or departure delays. In this case, the airline can be replaced by the transportation society, the airport by the station or the port, and the distance of the flight by the distance of the trip.

This work has some limitations that can be a subject for further research. Only domestic US flights information were used to predict arrival delay. International flights records were not utilized due to a lack of reliable relevant data. Another limitation would be meteorological data that were not considered and may be a challenging future research.

In the future, we will consider a flight delay prediction using real-time flight data. Furthermore, a complex but more performant deep learning model would be very interesting. Finally, all the needs and lacks that must be fulfilled in this research will be studied.

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/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed in collecting information, writing, modeling, and reviewing the article. All authors read and approved the final manuscript.

Acknowledgments

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

References

- [1] 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.
- [2] A. Sternberg, J. Soares, D. Carvalho, and E. Ogasawara, "A review on flight delay prediction," 2017, <http://arxiv.org/abs/1703.06118>.
- [3] K. Suzuki, *Artificial Neural Networks: Methodological Advances and Biomedical Applications*, BoD–Books on Demand, Norderstedt, Germany, 2011.
- [4] 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, no. 14-15, pp. 2627–2636, 1998.
- [5] B. T. S. Transtats, .Bts Transtats. <http://www.transtats.bts.gov/>. Accessed: 2020-06-01.
- [6] S. Putatunda and K. Rama, "A comparative analysis of hyperopt as against other approaches for hyper-parameter

- optimization of xgboost,” in *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, Shanghai, China, September 2018.
- [7] R. Liu, E. Liu, J. Yang, M. Li, and F. Wang, “Optimizing the hyper-parameters for svm by combining evolution strategies with a grid search,” in *Intelligent Control and Automation*, pp. 712–721, Springer, Berlin, Germany, 2006.
 - [8] J. V. Tu, “Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes,” *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996.
 - [9] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, “Supervised machine learning algorithms: classification and comparison,” *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, pp. 128–138, 2017.
 - [10] T. O. Ayodele, “Types of machine learning algorithms,” *New Advances in Machine Learning*, vol. 3, pp. 19–48, 2010.
 - [11] T. H. H. Aldhyani, M. Al-Yaari, H. Alkahtani, and M. Maashi, “Water quality prediction using artificial intelligence algorithms,” *Applied Bionics and Biomechanics*, vol. 2020, Article ID 6659314, 12 pages, 2020.
 - [12] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 161–168, Pittsburgh, PA, USA, June 2006.
 - [13] S. N. S. Mortazavi and A. Ince, “An artificial neural network modeling approach for short and long fatigue crack propagation,” *Computational Materials Science*, vol. 185, Article ID 109962, 2020.
 - [14] S. Toraman, T. B. Alakus, and I. Turkoglu, “Convolutional capsnet: a novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks,” *Chaos, Solitons & Fractals*, vol. 140, Article ID 110122, 2020.
 - [15] E. Harirchian, T. Lahmer, and S. Rasulzade, “Earthquake hazard safety assessment of existing buildings using optimized multi-layer perceptron neural network,” *Energies*, vol. 13, no. 8, p. 2060, 2020.
 - [16] Y. Jiang, Y. Liu, D. Liu, and H. Song, “Applying machine learning to aviation big data for flight delay prediction,” in *Proceedings of the 5th IEEE Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)*, pp. 665–672, Calgary, Alta, August 2020.
 - [17] N. Takeichi, R. Kaida, A. Shimomura, and T. Yamauchi, “Prediction of delay due to air traffic control by machine learning,” in *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, p. 1323, San Francisco, CA, USA, August 2017.
 - [18] D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori, and C. J. P. Alves, “Supervised neural network with multilevel input layers for predicting of air traffic delays,” in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, Rio de Janeiro, Brazil, July 2018.
 - [19] S. Khanmohammadi, S. Tutun, and Y. Kucuk, “A new multilevel input layer artificial neural network for predicting flight delays at jfk airport,” *Procedia Computer Science*, vol. 95, pp. 237–244, 2016.
 - [20] R. Henriques and I. Feiteira, “Predictive modelling: flight delays and associated factors, hartsfield-Jackson atlanta international airport,” *Procedia Computer Science*, vol. 138, pp. 638–645, 2018.
 - [21] P. Stefanovič, R. Štrimitis, and O. Kurasova, “Prediction of flight time deviation for lithuanian airports using supervised machine learning model,” *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 8878681, 10 pages, 2020.
 - [22] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, “Cost-sensitive prediction of airline delays using machine learning,” in *Proceedings of the 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, pp. 1–8, Petersburg, FL, USA, September 2017.
 - [23] J. Chen and M. Li, “Chained predictions of flight delay using machine learning,” in *Proceedings of the AIAA Scitech 2019 Forum*, p. 1661, San Diego, CA, USA, January 2019.
 - [24] S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, and S. Barman, “A statistical approach to predict flight delay using gradient boosted decision tree,” in *Proceedings of the 2017 International Conference on Computational Intelligence in Data Science (ICCIDIS)*, pp. 1–5, Chennai, India, June 2017.
 - [25] A. Dand, K. A. Saeed, and M. B. Yildirim, “Prediction of airline delays based on machine learning algorithms,” *Association for Information Systems*, vol. 11, 2019.
 - [26] B. Thiagarajan, L. Srinivasan, A. V. Sharma, D. Sreekanthan, and V. Vijayaraghavan, “A machine learning approach for prediction of on-time performance of flights,” in *Proceedings of the 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, pp. 1–6, St. Petersburg, FL, USA, September 2017.
 - [27] Kuhn, N., Jamadagni, N., 2017. Application of Machine Learning Algorithms to Predict Flight Arrival Delays. CS229.
 - [28] A. M. Kalliguddi and A. K. Leboulluc, “Predictive modeling of aircraft flight delay,” *Universal Journal of Management*, vol. 5, no. 10, pp. 485–491, 2017.
 - [29] W. Cao and X. Fang, “Airport flight departure delay model on improved bn structure learning,” *Physics Procedia*, vol. 33, pp. 597–603, 2012.
 - [30] W. Wu, K. Cai, Y. Yan, and Y. Li, “An improved svm model for flight delay prediction,” in *Proceedings of the 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, pp. 1–6, San Diego, California, USA, September 2019.
 - [31] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, “Flight delay prediction for commercial air transport: a deep learning approach,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, 2019.
 - [32] International Civil Aviation Organization ICAO, D/enleadertwodots , 2005. Global Air Traffic Management Operational Concept, Doc 9854 AN/458. ICAO.
 - [33] K.-L. Du and M. N. S. Swamy, “Multilayer perceptrons: architecture and error backpropagation,” in *Neural Networks and Statistical Learning*, pp. 83–126, Springer, Berlin, Germany, 2014.
 - [34] Z. Car, S. Baressi Šegota, N. Andeli, I. Lorencin, and V. Mrzljak, “Modeling the spread of covid-19 infection using a multilayer perceptron,” *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 5714714, 10 pages, 2020.
 - [35] M. Madhiarasan and S. N. Deepa, “Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting,” *Artificial Intelligence Review*, vol. 48, no. 4, pp. 449–471, 2017.
 - [36] V. Ravi, D. Pradeepkumar, and K. Deb, “Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms,” *Swarm and Evolutionary Computation*, vol. 36, pp. 136–149, 2017.
 - [37] A. V. Devadoss and T. A. A. Ligorì, “Forecasting of stock prices using multi layer perceptron,” *International Journal of Computing Algorithm*, vol. 2, pp. 440–449, 2013.

- [38] J. W. Gardner, M. Craven, C. Dow, and E. L. Hines, "The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network," *Measurement Science and Technology*, vol. 9, no. 1, pp. 120–127, 1998.
- [39] M. Hayati and Z. Mohebi, "Developing an intelligent forex rolling forecasting and trading decision support system II: an empirical and comprehensive assessment," *International Series in Operations Research & Management Science*, vol. 28, pp. 275–289, 2007.
- [40] P. Zhang, Y. Jia, J. Gao, W. Song, and H. Leung, "Short-term rainfall forecasting using multi-layer perceptron," *IEEE Transactions on Big Data*, vol. 6, pp. 93–106, 2018.
- [41] W. Shao, A. Prabowo, S. Zhao et al., "Flight delay prediction using airport situational awareness map," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 432–435, Chicago, IL, USA, November 2019.
- [42] 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, p. 11765, 2015.
- [43] M. Hosseinzadeh, O. H. Ahmed, M. Y. Ghafour et al., "A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things," *The Journal of Supercomputing*, vol. 77, pp. 1–22, 2020.
- [44] I. Syarif, A. Prugel-Bennett, and G. Wills, "Svm parameter optimization using grid search and genetic algorithm to improve classification performance," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, no. 4, p. 1502, 2016.
- [45] Wikipedia, Flight length,. Flight length. [http://en.wikipedia.org/wiki/Flight length](http://en.wikipedia.org/wiki/Flight_length). Accessed: 2021-03-22.
- [46] C. E. Depuydt, J. Jonckheere, M. Berth, G. M. Salembier, A. J. Vereecken, and J. J. Bogers, "Serial type-specific human papillomavirus (hvp) load measurement allows differentiation between regressing cervical lesions and serial virion productive transient infections," *Cancer Medicine*, vol. 4, no. 8, pp. 1294–1302, 2015.
- [47] P. Lu, Z. Zheng, Y. Ren et al., "A gradient boosting crash prediction approach for highway-rail grade crossing crash analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 6751728, 10 pages, 2020.
- [48] 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*, vol. 2010, Article ID 901095, 15 pages, 2010.
- [49] Scikit-Learn, . Metrics and Scoring Quantifying the Quality of Predictions. http://scikit-learn.org/stable/modules/model_evaluation.html. Accessed: 2020-07-13.