2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Prediction of Flight Arrival Delay Time using U.S. Bureau of Transportation Statistics

Jiarui LI
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
scyjl6@nottingham.edu.cn

Ran JI
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
scyrj1@nottingham.edu.cn

Cheng'ao LI
School of Aerospace Engineering
University of Nottingham Ningbo China
Ningbo, China
ssycl4@nottingham.edu.cn

Xiaoying YANG
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
scxxy1@nottingham.edu.cn

Jiayi LI
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
0009-0009-7887-9784

Yiran LI
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
yiran.li@nottingham.edu.cn

Xihan XIONG
Deparment of Computing
Imperial College London
London, UK
xihan.xiong20@imperial.ac.uk

Yutong FANG
Teaching Research Center
Ningbo Open University
Ningbo, China
fangyt@nbou.cn

Shusheng DING
Business School
Ningbo University
Ningbo, China
dingshusheng@nbu.edu.cn

Tianxiang CUI
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
tianxiang.cui@nottingham.edu.cn

***Abstract*—According to the data from the Bureau of Transportation Statistics (BTS), the number of passengers and flights has been increasing year by year. However, flight delay has become a pervasive problem in the United States in recent years due to various factors, including human factors such as security regulations, as well as natural factors such as bad weather. Flight delay not only affects the profits of airlines but also affects the satisfaction of passengers. Therefore, a model that can predict the arrival time of airplanes needs to be developed. Machine learning methods have been widely applied to prediction problems. In this paper, a variety of machine learning and computational intelligence methods, including linear regression, decision tree (DT), random forest (RF), gradient boosting (GB), gaussian regression models and genetic programming were trained on the U.S. Department of Transportation's (DOT) BTS dataset. The results show that genetic programming performs best and can be used to predict the arrival time of the U.S. flights in advance, which is beneficial for airlines and passengers to make timely decisions.**

***Index Terms*—Big data, Air flight, Airport, Delay, Machine learning, Computational intelligence, Prediction, Regression**

I. Introduction

Big data analysis has been successfully applied in many fields, including but not limited to biology [1], healthcare [2], geography [3], traffic and transportation [4]–[6], and the Internet [7]. Using big data analysis can help with discovering new information and making better-informed decisions [8], [9]. While travelling by air is becoming a more popular option, illustrated by the growth in passenger count between 2016 and 2019 in the Bureau of Transportation Statistics (BTS) data [10], the application of big data in the aviation field is still quite limited [4].

One serious problem airlines and travellers share alike with aviation is flight delays. In 2019, only 79% of flights in the United States (US) arrived on time. Delays cause crucial loss of time, money and resources for passengers and airlines [4], [10]–[12], and are estimated to cost $30 billion for all parties involved in 2019 in the US alone [13]. Other than losses directly caused by delays, there are also indirect costs such as the inevitable decrease in demand due to unreliable service, which can amount to an even larger amount of money [4].

Flight delays are generally classified as five main reasons, extreme weather, late-arriving aircraft, national aviation system, air carrier and security [14]–[17]. Therefore, the effective prediction of flight delays enables airlines and traffic management departments to make timely adjustments, which will have a positive effect on passengers and various aviation departments and reduce losses. Machine learning has been a hot-field method, a good quantum of research uses machine-learning methods to predict flights.

In this work, different models for predicting arrival delays based on flight departure information are proposed. The dataset used to train and test our model includes the flight data of the relevant airports in the states of the United States obtained from the DOT Bureau of Transportation Statistics database. It mainly contains (data column). In the prediction process, the preprocessed flight departure data are used to predict the possible arrival difference and delay information of the flight through the model. The machine learning and

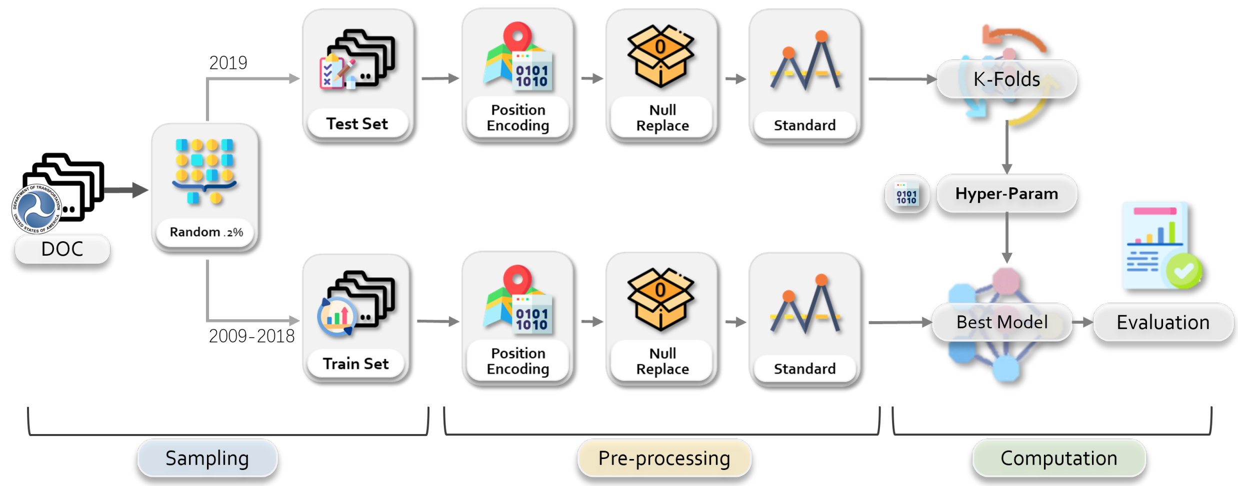

Fig. 1. The data process and predict procedure.

computational intelligence algorithms implemented in this work include Linear Regression, Decision Tree (DT), Random Forest (RF), Gradient Boost (GB), Gaussian Regression and Genetic Programming.

The rest of this work is organized as follows: Section II shows the related work of prediction in flight delay. Section III illustrates the problem statement of our work. Section IV illustrates the dataset and preprocessing step in this work. In the next part of Section IV, the procedure of data preprocessing is discussed. In the third part of Section IV, models are be described. Section V discusses the results of the model. Section VI provides the conclusion of this work.

II. Literature Review

As mentioned in the introduction, flight delays are a common problem encountered by airlines at present. When the arrival or departure time of a flight exceeds the estimated or scheduled time, the flight will be delayed.

In recent years, several studies have introduced methods for predicting urban traffic flow. Fei et al. [5] proposed a novel orthogonal spatial-temporal graph convolutional network (OSTGCN) to forecast traffic flow. Roudbari et al. [6] introduced an alternative traffic prediction model utilizing GNN-RNN cells. Xu et al. [18] proposed a method using genetic programming for traffic flow prediction. These works demonstrate the successful application of machine learning and computational intelligence in urban traffic prediction and suggest its potential for exploring airline traffic prediction.

Attempts to predict delays have been made by exploring air traffic patterns. Rebello and Balakrishnan et al. [17] created variables indicative of NAS status and used systematic dependencies between airports to predict future network-related delays. Klein et al. [19] focus on prediction from weather conditions and propose a model to predict delay of flight based on the Weather Impacted Traffic Index (WITI), which can be used to evaluate the impact and severity of weather. Belcastro et al. [20] proposed a method based on big data. A MapReduce program using parallel algorithms is executed on the cloud platform by analyzing and mining flight information and weather conditions, and it is used to predict flight delays caused by weather.

The latest research has also focused on using machine learning methods to predict delays and cancellations. All studies have concluded that there is a close relationship between arrival delays and departure delays [21]. Therefore, studies often use departure delays as inputs to predict arrival delays, although some studies predict arrival delays by using information of weather conditions as inputs. Some studies categorize flight delay prediction as a classification problem, while others categorize delay prediction as a regression problem, predicting delays in minutes.

Jiang et al. [4] used Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP) and data visualization methods to solve the two-category problem of flight delays and the classification task of five-level delays. Alonso et al. [21] proposed a unimodal model established through machine learning methods to classify the length of flight delays for arrivals at Porto Airport based on given airport departure flight information. They transformed the regression problem of delay time into a classification problem of time intervals. The model achieved a coefficient result of 0.7 for its prediction outcomes.

In the research of [15], a two-stage model is proposed, which firstly solves the classification problem of delay, and then predicts the actual delay time by regression methods. Xu et al. [22] used Bayesian network to analyze flight delays in two lines of the National Airspace System according to weather conditions. Choi et al. [14] also used supervised machine learning methods such as DT, RF, and AdaBoost to predict flight arrival delays based on weather conditions. In the studies of [23], [24], neural networks are used to predict flight delays for John F. Kennedy (JFK) and Esenboga International Airport respectively. Chakrabarty et al. [25] trained a Gradient Boosting model using flight details covering the top five busiest airports in the United States and the Gradient Boosting Tree (GBT) model is applied to predict flight arrival delays. In reference [26], researchers used logistic regression and random forest models to predict flight delay status and duration based on flight data recorded by the BTS. In addition, they considered some non-flight attributes such as weather and peak tourism data. The study results showed that the model's prediction accuracy was between 80% and 85%.

III. Problem Statement

Flight delays can result in significant financial losses for airlines, as well as inconvenience for passengers and issues with air traffic control systems. As a consequence, it is necessary to predict flight delay information in advance, in order to allow other passengers and air traffic control departments to be informed and take action proactively. Within the United States, machine learning regression models are utilized to predict flight delay information based on flight departure data. This technology can serve as a decision-making tool for air traffic control departments and airports.

IV. Methodology

A. Data Collection

The information regarding domestic delayed flights in the United States between 2009 and 2019 is sourced from the U.S. DOT BTS, which is responsible for monitoring the punctuality of domestic flights that are operated by major air carriers. It mainly includes delay data for domestic routes in the United States, which also provides other related information, such as airline date, original airport, destination airport, scheduled and actual departure time, taxi-out time, taxi-in time, wheels-off time, wheels-on time, scheduled and actual arrival time, flight time and delay time by various reasons, etc.

There are 68,979,001 samples in this total dataset. The randomly chosen flight data from 2009 to 2018 is used as the training set, and the randomly chosen data from 2019 is used as the test set. The train and test data set contains 121,513 and 14,861 flights, including flight data of 358 airports. All

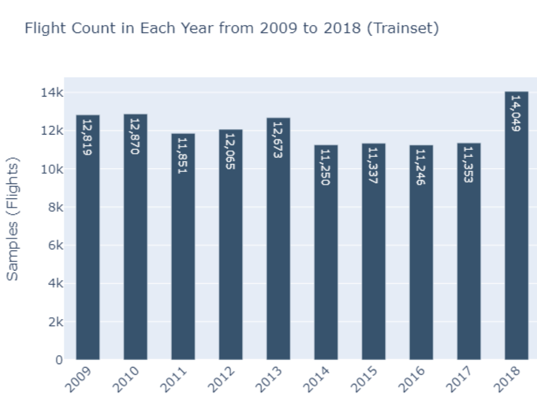

Fig. 2. Flight count of each year in training set

Fig. 3. Airline in training set demonstrates in the map.

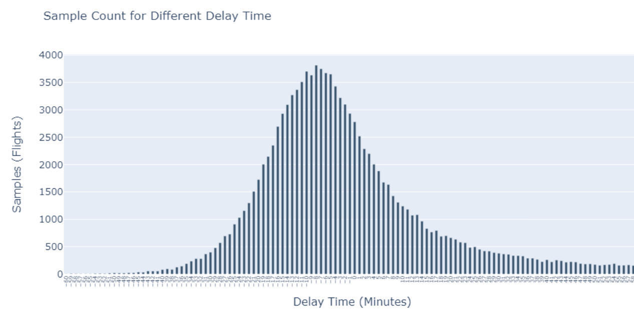

Fig. 4. Flight count of different delay level.

the attributes and descriptions in the dataset were written in the I

B. Data Pre-processing

To enable machine learning models to recognize the data, four data pre-processing methods are applied. Firstly, instead of encoding airport identifiers as numbers as in previous research [3], the airports are transformed to their longitude and latitude information to enable the models recognize the location of the airport better. Secondly, the operation carrier identifier is encoded into unique integer identifiers using the string indexer embedded by Spark. Next, null values in features such as Weather Delay Time, Carrier Delay Time, NAS Delay Time, Security Delay Time, and Late Aircraft Delay Time are replaced with 0 to identify flights that did not experience delays under those conditions. Finally, all characteristics are normalized by subtracting the average and adjusting to have a uniform variance.

After feature selection, the factors listed in TABLE II are considered to be effective information affecting flight delays. These data were extracted from the dataset and were used in training.

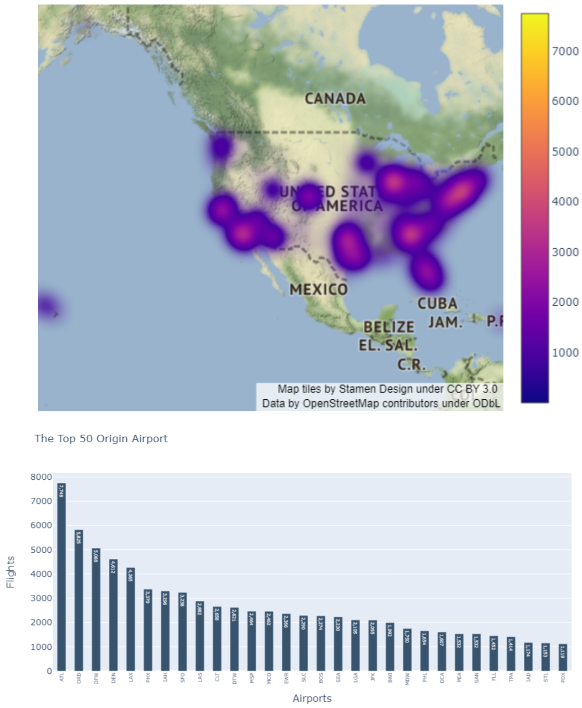

Fig. 5. Illustration of the amount of original airport distribution.

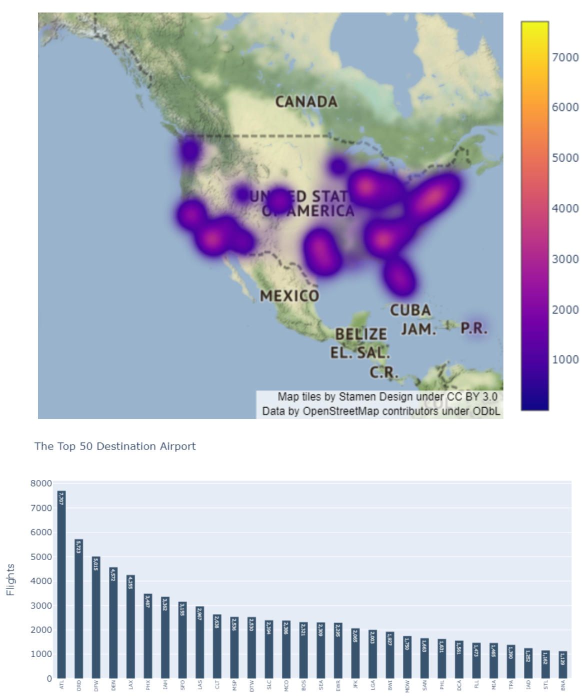

Fig. 6. Illustration of the amount of destination airport distribution.

TABLE I
ATTRIBUTES AND DESCRIPTIONS OF DATASET

Attributes	**Comment**
Flight Date	The date of the flight.
Carrier	Unique Carrier Code of different carriers.
Carrier Number	An identification number assigned by US DOT to identify a unique airline.
Original Location	The identifier of the original airport.
Destination Location	The identifier of the arrival airport.
Scheduled Departure Time	Time of the flight scheduled departure time.
Actual Departure Time	Time of the flight actual departure time.
Taxi-out Time	Time of the flight taxing-out time.
Taxi-in Time	Time of the flight taxing-in time.
Wheels-off Time	Time of the plane retracting its wheels.
Wheels-on Time	Time of the plane touching down with its wheels.
Delay Departure Time	Delay time of the flight departure.
Scheduled Arrival Time	Time of the flight scheduled arrival time.
Actual Arrival Time	Time of the flight actual arrival time.
Carrier Delay Time	Time of the flight delay due to carrier factors. Carrier Delay is within the control of the air carrier, such as aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, baggage, bird strike, etc.
Cancelled	Whether the flight was cancelled or not.
Cancelled Code	The reason for cancelling the flight.
Weather Delay Time	Time of the flight delay due to weather factors.
NAS Delay Time	Time of the flight delay due to national air system factors, such as non-extreme weather conditions, airport operations, heavy traffic volume, etc.
Security Delay Time	Time of the flight delay due to security factors, such as re-boarding of aircraft because of security breach, evacuation of a terminal or concourse, etc.
Late Aircraft Delay	Time of the flight delay due to the late arrival of the previous flight that utilized the same plane that will be departing.
Distance	Distance covered by the airline.

C. Prediction Models

To solve the predicting flight arrival time and judging delay information problems, as mentioned in the literature review, the following popular algorithms are used for training and tests:

- Linear Regression [27]
- Decision Tree (DT) [10], [28]
- Random Forest (RF) [11], [29]
- Gradient Boosting (GB) [30]
- Gaussian Regression [26], [31]
- Genetic Programming [18]

To improve results, the following methods are applied:

1) *K-Fold Cross-validation* [32]: Cross-validation is a measure for evaluating model performance. K-fold cross-validation is a technique that randomly divides the original sample into K sub-samples. In this work, the data was split into 3 folds. Then, a single sub-sample is used as validation data for testing the model, and the remaining 2 sub-samples are used as training data. These processes are repeated 3 times. As the original dataset is large, as mentioned in the previous section, random selection was performed to handle it. Furthermore, the k-

TABLE II
SELECTED ATTRIBUTES AND THEIR TYPES

Attributes	Type	Comment
Flight Date(mmdd)	Integer	The flight date of this flight.
Carrier ID	Integer	An identification number assigned by US DOT to identify a unique airline.
Original Location	(Float, Float)	The coordinator of the original airport including latitude and longitude.
Destination Location	(Float, Float)	The coordinator of the destination airport including latitude and longitude
Scheduled Departure Time (min)	Float	Time (min) of the flight scheduled departure time.
Departure Delay Time (min)	Float	Delay time (hhmm) of the flight departure. Type in Integer.
Distance	Float	Distance of the airline.
Carrier Delay Time (min)	Integer	Time (min) of the flight delay by carrier factors. Carrier Delay is within the control of the air carrier, such as aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, baggage, bird strike, etc.
Weather Delay Time (min)	Integer	Time (min) of the flight delay by weather factors.
NAS Delay Time (min)	Integer	Time (min) of the flight delay by national air system factors, such as non-extreme weather conditions, airport operations, heavy traffic volume, etc.
Security Delay Time (min)	Integer	Time (min) of the flight delay by security factors, Time of the flight delay due to security factors, such as re-boarding of aircraft because of security breach, evacuation of a terminal or concourse, etc.
Late Aircraft Delay Time (min)	Integer	Time (min) of the flight delay by the late arrival of the previous flight that utilized the time of the same plane that will be departing.

fold cross-validation method is employed to counteract potential bias resulting from the random data selection, as mentioned in [33].

2) *Hyper-Parameter Tuning* [34]: To ensure a good fit of the machine learning model to the data, hyperparameters must be carefully configured prior to training. In our approach, several sets of hyperparameters are tested using K-Fold validation, and the set of parameters that yields the best performance is selected as the final hyperparameter configuration. For the Decision Tree and Random Forest models, different values were tested for the maximum depth of the tree, including 5, 6, 7, and 8. For the Linear Regression and Gaussian Regression models, different convergence tolerances were tested, including 1e-4, 1e-5, 1e-6, and 1e-7.

V. RESULTS AND DISCUSSION

A. Prediction Results

To analyze the performance of the models, root mean squared error (RMSE) and R^2 score are used for evaluation regression results. As mentioned in the previous section, training data is from 2009-2018, 10 years, and test data is in 2019. To evaluate the models, K-Fold Cross Validation (K=3) are performed, the regression results of which are shown in Table III and Table IV.

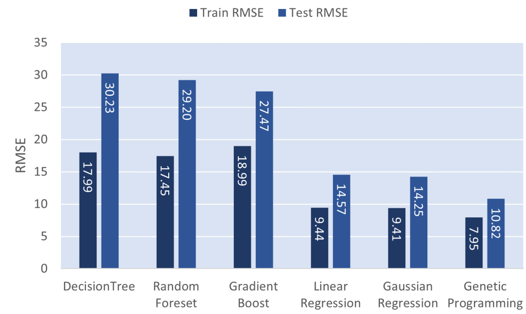

Fig. 7. Comparisons of different methods on RMSE

The RMSE and R^2 scores results of Decision Tree [10], Random Forest [11], Linear Regression [27], Gaussian Regression [26], and Genetic Programming are shown in Fig 7, Table III and Table IV. Gaussian Regression and Genetic programming are shown to outperform than other methods.

B. Findings

The Decision Tree models yield unsatisfactory results due to their discrete nature. This is evident when comparing their performance with Random Forest and Gradient Boost models. With more trees adapted, the model can fit the data more smoothly and achieve better results. In contrast, linear regression somewhat predicts airplane arrival time differences by smoothly conforming to data patterns. Furthermore, flight departure details like time and distance serve as inputs for delay predictions. While literature suggests using weather data like temperature and humidity for predictions, our dataset lacks this information. This remains a potential area for future

TABLE III
RESULTS ON TRAIN SET OF DIFFERENT MODELS IN THIS WORK

Regressor	RMSE	R^2 Score
Decision Tree	17.9888	0.7935
Random Forest	17.4475	0.8058
Gradient Boost	18.9884	0.7700
Linear Regression	9.4382	0.9432
Gaussian Regression	9.4127	0.9480
Genetic Programming	**7.9542**	**0.9527**

TABLE IV
RESULTS ON TEST SET OF DIFFERENT MODELS IN THIS WORK

Regressor	RMSE	R^2 Score
Decision Tree	30.2331	0.6314
Random Forest	29.1984	0.6562
Gradient Boost	27.4730	0.6956
Linear Regression	14.5720	0.9144
Gaussian Regression	14.2470	0.9218
Genetic Programming	**10.8194**	**0.9499**

exploration. As highlighted, flight delay is a pervasive issue. Our proposed predictive model aids airlines and passengers in prompt decision-making, thus mitigating the impact of delays to some extent.

VI. Conclusion

In this work, we propose a method using machine learning and genetic programming to predict flight delay time based on the departure information of flights. To efficiently handle large-scale datasets. Experimental results reveal that the Gaussian regression machine learning and genetic programming approaches outperform other evaluated methods. The findings indicate that the proposed method can predict flight delay time, mitigating economic losses for airlines and passengers.

Aknowledgement

This Project is Supported by Ningbo Natural Science Foundation (Project ID 2023J194).

References

[1] V. Marx, "The big challenges of big data." in *Nature.*, no. 498, 2013, p. 255–260.

[2] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, pp. 1–10, 2014.

[3] R. Kitchin, "The real-time city? big data and smart urbanism," *GeoJournal*, vol. 79, pp. 1–14, 2014.

[4] Y. Jiang, Y. Liu, D. Liu, and H. Song, "Applying machine learning to aviation big data for flight delay prediction," in *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2020, pp. 665–672.

[5] Y. Fei, M. Hu, X. Wei, and M. Chen, "Orthogonal spatial-temporal graph convolutional networks for traffic flow forecasting," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2022, pp. 71–76.

[6] N. S. Roudbari, Z. Patterson, U. Eicker, and C. Poullis, "Simpler is better: Multilevel abstraction with graph convolutional recurrent neural network cells for traffic prediction," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2022, pp. 01–10.

[7] C. Snijders, U. Matzat, and U.-D. Reips, ""big data": big gaps of knowledge in the field of internet science," *International journal of internet science*, vol. 7, no. 1, pp. 1–5, 2012.

[8] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-generation big data analytics: State of the art, challenges, and future research topics," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1891–1899, 2017.

[9] G. Dartmann, H. Song, and A. Schmeink, *Big data analytics for cyber-physical systems: machine learning for the internet of things*. Elsevier, 2019.

[10] R. Boggavarapu, P. Agarwal, and R. K. D.H, "Aviation delay estimation using deep learning," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, 2019, pp. 689–693.

[11] K. Sharma, R. L. Eliganti, B. S. K. Meghana, and G. Gayatri, "Error calculation of flight delay prediction using various machine learning approaches," in *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*, 2022, pp. 1–5.

[12] 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 *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, 2017, pp. 1–6.

[13] F. A. Administration, "Cost of delay estimates 2019," https://www.faa.gov/data_research/aviation_data_statistics/media/cost_delay_estimates.pdf, 07 2020.

[14] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 2016, pp. 1–6.

[15] 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 *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 2017, pp. 1–6.

[16] S. Addu, P. R. Ambati, S. R. Kondakalla, H. Kunchakuri, and M. Thottempudi, "Predicting delay in flights using machine learning," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2022, pp. 374–379.

[17] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation research part C: Emerging technologies*, vol. 44, pp. 231–241, 2014.

[18] C. Xu, Z. Li, and W. Wang, "Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming," *Transport*, vol. 31, no. 3, pp. 343–358, 2016. [Online]. Available: https://doi.org/10.3846/16484142.2016.1212734

[19] A. Klein, C. Craun, and R. S. Lee, "Airport delay prediction using weather-impacted traffic index (witi) model," in *29th Digital Avionics Systems Conference*. IEEE, 2010, pp. 2–B.

[20] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, "Using scalable data mining for predicting flight delays," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, pp. 1–20, 2016.

[21] H. Alonso and A. Loureiro, "Predicting flight departure delay at porto airport: A preliminary study," in *2015 7th International Joint Conference on Computational Intelligence (IJCCI)*, vol. 3, 2015, pp. 93–98.

[22] N. Xu, G. Donohue, K. B. Laskey, and C.-H. Chen, "Estimation of delay propagation in the national aviation system using bayesian networks," in *6th USA/Europe Air Traffic Management Research and Development Seminar*. FAA and Eurocontrol Baltimore, 2005.

[23] 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.

[24] E. Demir and V. B. Demir, "Predicting flight delays with artificial neural networks: Case study of an airport," in *2017 25th Signal Processing and Communications Applications Conference (SIU)*, 2017, pp. 1–4.

[25] N. Chakrabarty, "A data mining approach to flight arrival delay prediction for american airlines," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. IEEE, 2019, pp. 102–107.

[26] V. Natarajan, S. Meenakshisundaram, G. Balasubramanian, and S. Sinha, "A novel approach: Airline delay prediction using machine learning," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 1081–1086.

[27] Y. Ding, "Predicting flight delay based on multiple linear regression," in *IOP conference series: Earth and environmental science*, vol. 81, no. 1. IOP Publishing, 2017, p. 012198.

[28] I. H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with java implementations," *Acm Sigmod Record*, vol. 31, no. 1, pp. 76–77, 2002.

[29] K. Sharma and T. S. Kiranmai, "Prediction of cardiovascular diseases using genetic algorithm and deep learning techniques," *INTERNATIONAL JOURNAL OF EMERGING TRENDS IN ENGINEERING AND DEVELOPMENT*, vol. 3, no. 10.26808, 2021.

[30] 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 *2017 International conference on computational intelligence in data science (ICCIDS)*. IEEE, 2017, pp. 1–5.

[31] Z. Chen, D. Guo, and Y. Lin, "A deep gaussian process-based flight trajectory prediction approach and its application on conflict detection," *Algorithms*, vol. 13, no. 11, p. 293, 2020.

[32] K. Ito and R. Nakano, "Optimizing support vector regression hyperparameters based on cross-validation," in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, vol. 3, 2003, pp. 2077–2082 vol.3.

[33] H. Moayedi, A. Osouli, H. Nguyen, and A. S. A. Rashid, "A novel harris hawks' optimization and k-fold cross-validation predicting slope stability," *Engineering with Computers*, vol. 37, pp. 369–379, 2021.

[34] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1934–1965, 2019.