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

Applying Machine Learning to Aviation Big Data for Flight Delay Prediction

Yushan Jiang
Department of Electrical Engineering and Computer Science
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

Yongxin Liu
Department of Electrical Engineering and Computer Science
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

Dahai Liu
College of Aviation
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

Houbing Song
Department of Electrical Engineering and Computer Science
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

***Abstract*—Flight delay has been a serious and widespread problem that needs to be solved. One promising solution is the flight delay prediction. Although big data analytics and machine learning have been applied successfully in many domains, their applications in aviation are limited. This paper presents a comprehensive study of flight delay spanning data pre-processing, data visualization and data mining, in which we develop several machine learning models to predict flight arrival delays. Two data sets were used, namely Airline On-Time Performance (AOTP) Data and Quality Controlled Local Climatological Data (QCLCD). This paper aims to recognize useful patterns of the flight delay from aviation data and perform accurate delay prediction. The best result for flight delay prediction (five classes) using machine learning models is 89.07% (Multilayer Perceptron). A Convolution neural network model is also built which is enlightened by the idea of pattern recognition and success of neural network method, showing a slightly better result with 89.32% prediction accuracy.**

***Index Terms*—Flight Delay, Machine Learning, Aviation Data Analytics**

I. Introduction

Big data analytics has been successfully used in many domains to uncover information for informed decision-making [1], [2]. One example was to enable the move from internet of things to real-time control [3], [4]. However, the application of big data analytics in aviation is limited. Flight delay has been a serious and widespread problem plaguing both domestic and international air travel. In recent years, as aviation delays in the United States are increasing, the development of the civil aviation industry has been severely affected. According to the statistical data of airline delay, only 79 percent of flights in 2019 arrived on time, which results in tens of billion US dollars loss including the cost to airlines and passengers, demand loss, and other indirect costs. As one major business issue which gives rise to a significant economic and reputation loss to airlines, flight delay needs to be fully studied so as to reduce the cost by utilizing its own characteristics.

The delays of certain flights will be propagated to others and airport operation efficiency will deteriorate over time without timely monitoring and interpose. Therefore, it is preferred to capture any kind of perturbations in the schedule and operation stage and predict the flight delay before unexpected consequences occur. Many studies, which rely heavily on simulation and modeling, have been performed to simulate the delay and provide a base to establish management strategies. However, those simulation-based methods are not always able to emulate real airport behavior and provide a solution in a timely manner. Besides, there are multiple factors that may cause flight delay and need to be identified and considered at the same time, such as bad weather conditions, airport congestion in peak-hour, personnel and passenger delays, technical difficulties, and so on. It is difficult for models to handle the complex operating environment involving multiple delay factors and make appropriate assumptions. Therefore, a prediction model with both high accuracy and real-time performance is needed to establish better operation strategies to address flight delay issues efficiently. With the development of current aviation information systems, more and more flight data are available including factors that might contribute to flight delays from macro-level to micro-level. It enables aviation data analytics using machine learning techniques.

In this paper, we performed an aviation data analytic and apply machine learning techniques to realistic aviation dataset for flight arrival delay prediction. Instead of blindly testing machine learning models, we leverage the strength of data visualization to discover potential patterns of flight delay for a better understanding of the explored data and reasonable factors selection before building prediction models. Moreover, many previously conducted work only used on-time performance of flights for binary classification (delayed or not delayed) while we also take weather impacts into account and aim at a more practical five-classes delay classification task. Besides the traditional machine learning models, we also attempt to use a convolutional neural network model in deep learning field enlightening by the idea of applying pattern recognition to delay features, while a majority of

978-1-7281-6609-4/20/$31.00 ©2020 IEEE
DOI 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00114

previous works adopting deep learning models implemented their prediction models with recurrent neural networks.

The remainder of this paper is structured as follows: Section 2 presents a literature review of flight delay prediction. Section 3 provides a data exploration including manipulating two separate datasets into a tidy one and visualizing some potential delay patterns. Section 4 presents the construction and evaluation of prediction models. Section 5 provides conclusions and future work.

II. Related Work

A basic data analytics using machine learning is presented as follows. Firstly data pre-processing is needed including the data merging and cleansing, as each dataset is untidy with messy redundant records and missing values. Next, data visualization can be performed to extract and visualize the graphic representation of data clearly and efficiently, so that some useful patterns of flight delay can be found. Then feature engineering extracts and transforms features from processed data to ones that can be more suitable for representing the underlying task and give rise to improvement on predict model accuracy. Lastly, prediction models are built and trained using different machine learning methods, then evaluated. Therefore, prediction models regarding flight delay issues can differ a lot in both the choices of model construction to capture different concentrations or better results.

As mentioned in the introduction, there are many factors (features) that might cause flight delay. In [5], a general flight dataset is analyzed and some useful patterns of the relationship between feature and flight delay are shown by data visualization methods, which gives an insight into vital features influencing delay and provides a foundation for further model construction. Based on identified features, some basic machine learning models are built for the prediction task, where the data used in each paper is minorly different but from the same data source, Bureau of Transportation Statistics. In [6], both logistic regression classification and decision tree model are used for a binary flight delay classification, where departure/arrival delay is defined as a flight departure/arrival delay of 15 minutes or more. In both models, arrival delay prediction accuracy is higher by a margin than that of departure delay, with the best result 88.12%. [7] compared the performance of more supervised machine learning algorithms in prediction task, including decision trees, random forest, the AdaBoost and the K Nearest-Neighbors, while [8] mainly used gradient boosting model. Both papers emphasized the importance of data balance in real delay predicting tasks. This is because flights without delay outnumbers ones with delay and such an imbalance may cause a high prediction precision but low recall, making the classifiers invalid. [9] still aimed at binary delay classification but with three different metrics, departure delay, arrival delay and cancellation, which makes the prediction more comprehensive. Besides, the data used in this paper is from one single airport, London Heathrow Airport for 5 years, which can be used as a base model for generalizing or transferring to others.

In [10] a two-stage prediction model is proposed. Firstly a binary classification flight delay is performed as usual. Then a regression model is applied to those flight which is classified as delayed predicting the actual delay length. In the classification stage, the best result for departure delay and arrival delay is 86.48% and 94.35% respectively using gradient boosting model. In the regression stage, extra-trees regressor can predict the delay with an error of approximately 8 minutes, which is acceptable. Moreover, instead of selecting features manually, a recursive feature elimination technique is used to eliminate the weakest feature after each epoch, which can be seen as a greedy algorithm for finding the optimal feature subset.

Previous studies mainly consider features of flight and operation information with relatively fewer weather features, while [11] proposed a model that aims at analyzing the impact of weather on flight delay. It worth noting that the convective hazard records with location information provided by National Convective Weather Forecast are a set of 2D maps, where each map is discretized into pixels spanning the areas around studied airports and each pixel is assigned with a binary value indicating the existence of convective weather activities. The delays in airport of interest are predicted with status and duration, where the results show a good prediction performance regarding flight delay with weather-concentrated data.

Recently deep neural network method for flight delay prediction has gained favors with the development of machine learning. To capture the time characteristic of flight delay data, recurrent neural network models are preferred, especially Long Short-Term Memory RNN model [12]. [13] proposed a two-stage binary prediction model, where the day-to-day delay status of each airport is predicted using a LSTM model in the first stage and used as an input to predict individual flight with artificial neural networks in the second stage. [14] also proposed a LSTM model considering the air traffic flow. Unlike other studies directly using or modifying prepared dataset, this study established an ADS-B message-based aviation data platform for gathering partial necessary data and summarize air traffic flow which is defined as hourly flights per air route. Up to 4-class classification is performed for proposed LSTM model and random forest model. Although LSTM model suffers an overfitting issue due to limited training dataset, it shows the feasibility to get ADS-B data involved in delay prediction task and promising performance can be expected after gathering more years of data.

III. Data Exploration

In this section, the exploration of the final dataset used in flight delay prediction task is performed, including data pre-processing stage where data is merged and cleansed to a tidy one and data visualization stage where the graphic representation of flight delay data is extracted and visualized.

A. Data And Pre-processing

In this section, the data source and pre-processing steps including the data merging and cleansing will be introduced, as each dataset is untidy with messy redundant records and missing values. The tool used in this stage is Python and R.

The data used in this paper consists of two parts, Airline On-Time Performance Data (AOTP) and Quality Controlled Local Climatological Data (QCLCD) in the year 2016. AOTP data is provided by Bureau of Transportation Statistics, which contains the basic on-time arrival and departure information for each non-stop domestic flight, including the schedule and actual departure and arrival data, the carrier, origin and destination with airtime and non-stop distance. Besides the flight information from AOTP data, QCLCD contains basic hourly airport weather data including temperature extremes, visibility, air pressure, humidity and wind, which is provided by National Centers for Environmental Information and measured by major airport weather stations.

1) Step1: Form one complete QCLCD dataset: QCLCD contains two kinds of information, pure weather data that we need and the geographical station data including the airport callsign, which are in two separate subsets. WBAN(Weather-Bureau-Army-Navy) identifier is used as a merge key so as to attach weather information to each domestic airport. A cleansed QCLCD dataset contains the following kinds of features: observing airport, observation date with local standard time, temperature, humidity, wind speed and wind direction, air pressure and visibility.

2) Step2: Merge QCLCD with AOTP data: Airport callsign, schedule flight date, schedule flight time in ATOP data are used as the common keys which correspond to the weather station callsign, observation date and time in QCLCD, to perform the merge of QCLCD and ATOP data in an hourly basis. Due to the fact that some cases contain more than one record in an hour creating redundant cases, a cleansing is performed to accurately attach weather data to each flight. Moreover, the diverted and canceled flights are filtered out as we only concern the delay characteristics.

3) Step3: Further processing on merged dataset: Since the flight delay is actually impacted by the previous flight of the same airplane, the delay time of a previous flight is also extracted as a feature for exploration. Moreover, the average delay time of the identical flights with the same origin and destination before each flight is also calculated so that the past records can be fully used. After that, the data cleansing is performed again regarding missing values and other redundant records. In this paper, we convert the flight delay prediction to a 5-class classification problem, where the partition of delay adopted by [15] is shown as Table I. So far, the final tidy dataset has been obtained which can be used for following visualization and prediction.

B. Data Visualization

In this section, we aim at extracting and visualizing the graphic representation of data clearly and efficiently, so that some useful patterns of flight delay can be found. This step is necessary as it deepens the people's understanding of this data and helps with further potential exploration and prediction. To be specific, a general distribution of flight delay is introduced and some prominent features are described in a detailed manner with the effect of variation on flight delay. The implementing tool for this visualization section is R.

TABLE I
DELAY CLASS PARTITION

Flight Delay T in Minutes	***Delay Class***
T ≤ 15	0 (Non-delay)
$15 \leq T \leq 60$	1 (Slight delay)
$60 \leq T \leq 120$	2 (Medium delay)
$120 \leq T \leq 240$	3 (High delay)
T≥ 240	4 (Severe delay)

1) General Distribution: There are 4253524 flights in this dataset with the range of flight delay from -87 to 1660 minutes, and 4251121 flights have arrival delay less than 500 minutes. Therefore, the distribution of the flight delay is mostly distributed in the range of fewer than 500 minutes as shown in Fig.1. It is clear that the majority of cases belong to non-delay class and the overall flight delay fits log-normal distribution.

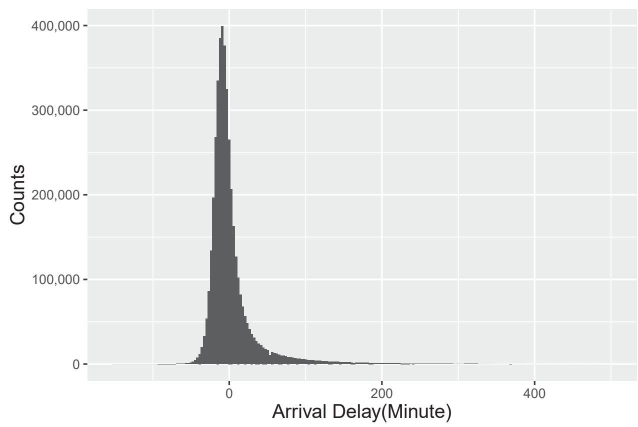

Fig. 1. The distribution of flight delay

2) Airline: Airline is one feature that would be used to predict the flight delay as the operating mechanism of different airlines may cause issues like slower airplane fueling and maintaining, slower baggage loading which takes up more time. The dataset contains 12 different airline companies, we compare the average delay with a standard error of the mean as shown in Fig.2. This figure clearly shows that Spirit Airlines, JetBlue Airways, Frontier Airlines and Virgin Atlantic Airways are more likely to have delay issues with the longer average delay, while the corresponding deviations are also relatively larger at the same time which indicates a bit more unstable delay situation. Besides, Alaska Airlines, Delta Airlines and Hawaiian Airlines have obviously less average delays and are likely to have earlier arrivals than scheduled ones. Therefore, the percentage of delay is likely to vary with the airline company.

3) Month and Day of Week: The date related features can also play important roles in flight delay as there would be a lot

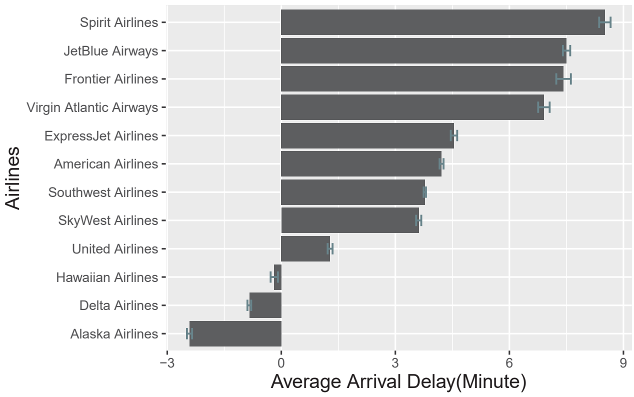

Fig. 2. Flight delays in different airlines

more people traveling in a certain period of time like holidays, which causes delays by challenging the resilience of airport congestion. Therefore, the trend of flight delay with respect to date such as Month and day of week can be studied to help analyze the general delay situation in a week or a year. Fig.3 shows the delay pattern regarding both features. It is obvious that June, July, August and December have more average delay time compared with others, which indicates that flight-heavy months are likely to be in summer and late winter. A similar phenomenon is observed in the scale of week, namely the delays at beginning of week like Monday and middle of week like Thursday and Friday are more than others, while delays on Saturday are likely to be the least.

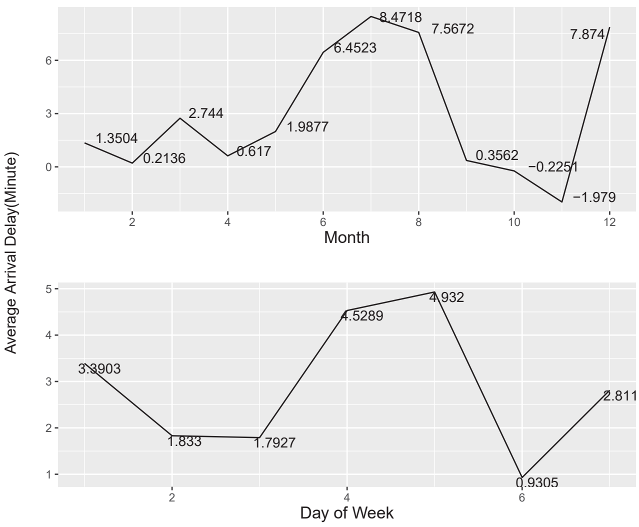

Fig. 3. Month/day of week and flight delay

4) Schedule Elapsed Time: The duration of flight is also an interesting feature that may impact the delay of aircraft. In this case, Schedule Elapsed Time(SET) is considered to be visualized and explored. To be specific, each flight is labeled according to TABLE I and grouped by SET, after which the percentages of each class with the same SET are calculated. Fig.4 shows how flight delay varies with the variation of SET. It is clear that the percentage of non-delay class decrease as the SET increases, with obvious deviation for SET $>$ 400 minutes, while the percentages of classes from slightly delay to extreme delay increase as SET goes up, which means generally flights scheduled to fly longer are more likely to delay.

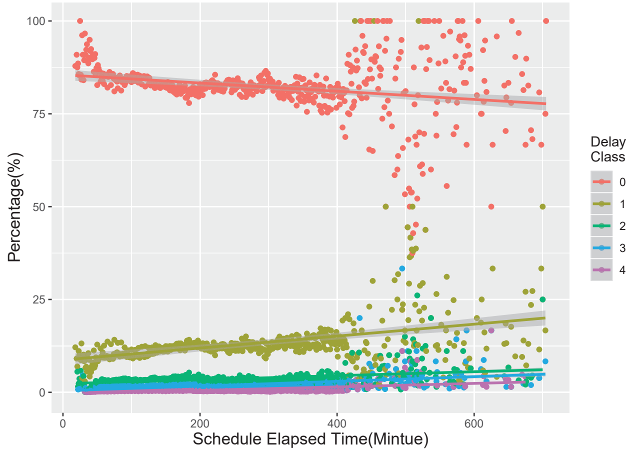

Fig. 4. Schedule elapsed time and flight delay

5) State: The state of departure airport is another feature that needs to be studied for flight delay as shown in Fig.5. The tree-map is used where the area of each rectangle stands for the number of flights and the color indicates the average arrival delay in minutes. It is clear that states with the top five largest number of flights are more likely to have longer delays such as Florida, Texas, California and Illinois, while the flights from Georgia still perform better in delays even with a large number of flights. Besides, states like Alaska, Hawaii and Montana have obviously less delay time with relatively fewer flights. These patterns prove that the flight delay does vary with the state of departure airport.

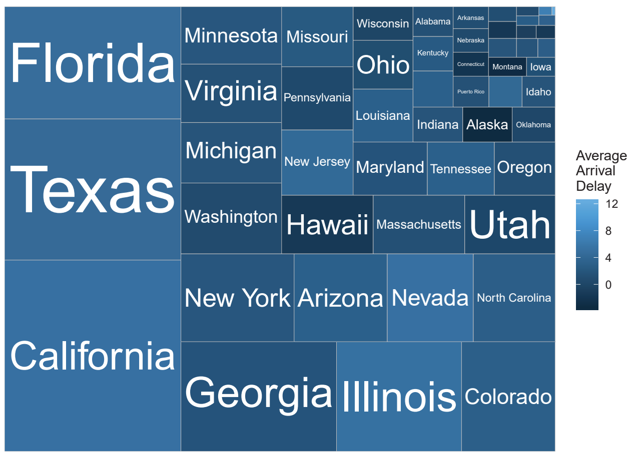

Fig. 5. Flight delay and number of flights in different states

6) Airport Congestion: The airport congestion is an indicator showing the level of busyness in each airport, which may have a direct influence on flight delay. To study this, the first step is to divide all flights to different dates, different departure hours and different airports. Then the number of divided cases is calculated and presented as flights per hour in departure airports while the arrival delay is averaged and presented as the average flight delay from an airport in different hours. Here a bunch of records are obtained with the same number of flights per hour which belongs to different airports. To observe the general situation, these records get averaged again by the number of airports sharing the same congestion and the result is plotted in Fig.6. It can be found that generally the average delay increases as the more airplanes take off per hour. However, there is a decrease in average delay in between, before it goes up again, which indicates certain capability and capacity of some airports dealing with congestion issue.

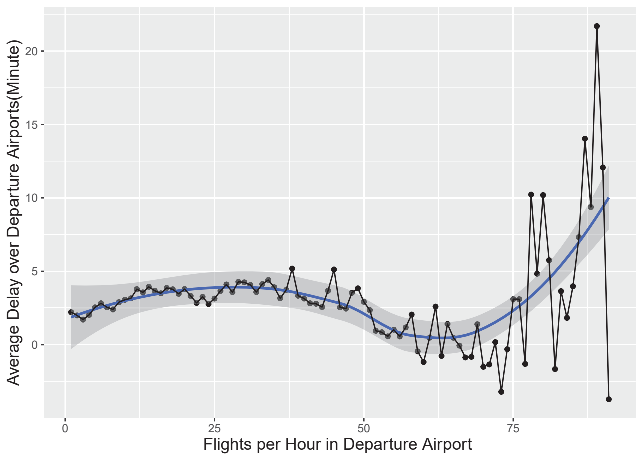

Fig. 6. Airport congestion and flight delay

7) Weather: Besides operation airline, flight date, flight time, location and airport congestion which were previously visualized and explored, the last but not least feature that needs to be studied is the weather, as it is known that weather conditions can significantly impact the flights. In this case we mainly study the visibility and wind speed as two weather-related features out of eight. In this dataset, visibility refers to prevailing surface visibility whose average value is 3 miles. 5-class labeling is applied again to present the trend of data with the existence of large samples. According to the result shown in Fig.7, an obvious increase of non-delay class, as well as decreases of other classes can be observed as the prevailing surface visibility gets better, which indicates the direct influence of visibility on flight delay.

For wind speed, its calculation is based on a 2-minute average just prior to observation time. in this case, we only consider flights in an extremely windy situation. According to the airplane specifications, a commonly-used Boeing 737-800 has a maximum allowable crosswind component of approximately 33 knots on a dry runway. Therefore, we pick 33 as the general threshold to see if there is any pattern for wind

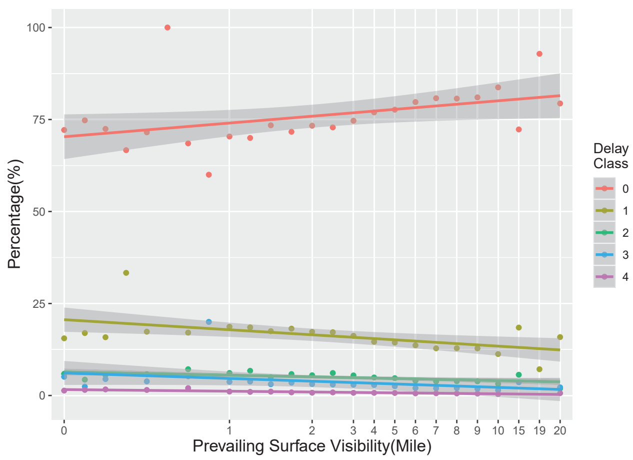

Fig. 7. Visibility and flight delay

speed and flight delay in this data set. The result is shown in Fig.8, and apparently, the average delay goes up as the wind speed increases in a severely windy situation.

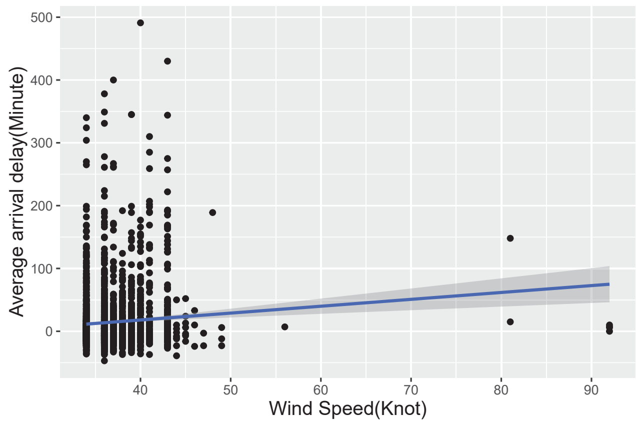

Fig. 8. Flight delay in a blustery situation

IV. Prediction Model Construction and Evaluation

Since several important features have been identified and studied in previous sections, we start to explore the dataset to a deeper degree and build several prediction models based on different machine learning techniques. This section is organized as follows: Firstly feature engineering will be performed including feature selecting, encoding and normalizing. Secondly, the processed data will be applied to machine learning models for prediction and evaluation. Lastly, a model using deep learning neural network is built for prediction and comparison as well.

A. Feature Engineering

Feature engineering is the process of using data mining techniques to extract and transform features from raw data to ones that can be more suitable for representing the underlying task and give rise to improvement on predict model accuracy. In this subsection, feature selection, feature encoding and feature normalization will be introduced.

1) Feature Selection: First of all, it is necessary to specify the prediction task. As mentioned in data pre-processing, all arrival delay period is labeled to 5 classes according to the TABLE I, where the partition of delay is also applied in other works. Therefore, the prediction is converted to a classification task. Secondly, all features related to arrival information are deleted such as the time of aircraft's wheels on the ground, the time of arrival, the actual elapsed time, airtime and so on. Only delay class label is kept for model training and testing. Moreover, data visualization is performed as a reference to remove redundant features in the underlying task.

2) Feature Encoding: There are two types of features in the dataset, namely the continuous one and discrete one. Feature encoding is only for discrete features. This step aims at transforming a discrete feature into a continuous numerical value and so that it can be used in the model. Discrete features are divided into two classes, the ordinal one and the categorical one. For an ordinal feature whose orders have extra information, such as month, day of month, day of week, schedule departure hours, time zone and so on, the encoding is simply mapping each value to numerical number (1,2,3...) according to orders within the feature.

As for categorical features, there are six ones in this dataset, namely departure and arrival airport, departure and arrival state, wind direction and operating airline. The choices of encoding categorical features are more than that of ordinal features. In this paper, two categorical encoding methods are mainly used, namely LabelCount encoding, Binary encoding. Besides, three alternatives are provided, including the above two encoding and hybrid encoding using different methods for different features.

LabelCount encoding sorts the categories according to the frequency of each category within the feature, in the training set. Compared with standard frequency encoding which gives the same encoding to different categories having the same frequency, LabelCount has a specific advantage that it is not sensitive to outliers. Binary encoding converts categories to binary bit strings, which is similar to One-hot encoding but showing strong robustness when the number of categories is large, without the issues of sparsity. The choice of hybrid encoding is this paper is that using LabelCount encoding for departure and arrival airport which have relatively more categories, and Binary encoding for the rest. After encoding, the next step is feature scaling.

3) Feature Scaling: Feature scaling is important for some machine learning models where Euclidean distance is used for calculating objective functions. That is because some features with a lot boarder range of values are likely to dominate the objective functions so that the model can not work properly. Moreover, it also facilitates the converge of models using gradient descent for updating weights. In this paper three normalization method is considered, i.e. Min-Max Scaler, Standard Scaler, and Normalizer. Min-Max Scaler is the simplest scaling method which rescales the original feature range to [0,1]. The formula is shown below.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Standard Scaler makes the values of each feature in the data have zero-mean and unit-variance by subtracting the mean and being divided by standard deviation as shown in the below formula.

$$x' = \frac{x - \bar{x}}{\sigma} \tag{2}$$

Normalizer scales each component of a vector to unit length by dividing individuals by the Euclidean length of the vector as shown in the formula below. It is worth noting that in this case, this normalizer directly performs to each record instead of each feature.

$$x' = \frac{x}{\|x\|} \tag{3}$$

B. Model Prediction and Evaluation

So far, discrete features have been encoded and the whole dataset has been scaled. The next step is to apply the machine learning models to the dataset with different choices of feature engineering and to evaluate the results. In this section, firstly the dataset gets separate and sampled. After that, the sampled training set is sent to the machine learning model. Once the model is trained, the test set will be used to evaluate the model and output the test accuracy as prediction accuracy. Besides, the best choice of each model and the best model will be discussed. Moreover, a small convolutional neural network is also built for evaluating the data performance on deep neural networks.

1) Training data and test data in a skewed dataset: Firstly the training data and test data are separated with the ratio of 7:3. It can be observed that this dataset is skewed since most records belong to non-delay class and the ratio of each delay class from non-delay to severe delay is near 202:28:8:3:1. The skewness of this dataset is extreme, where the precision of classification may be very high but the recall will be pretty low making the classifier invalid. In order to alleviate this class imbalance problem, we under-sample the majority classes in training data giving rise to a different ratio as 5:5:2:2:1.

2) Model Training and Prediction: Four common-used machine learning models are used, namely Support Vector Machine, Decision Tree, Random Forest and Multilayer Perceptron. In the training stage, K-Fold Cross Validation(K=5) is used, where the training data is divided into K folds, then each fold is treated as a validation once while the k-1 remaining folds form the training set. As for hyper-parameters tuning, a grid search method is applied, combined with above K-Fold Cross Validation, so that a set of relatively good hyper parameters can be selected after validating, and the best model with its parameters can be evaluated by the test set.

3) Evaluation of Feature Scaling: After the training stage, the prediction accuracy is presented as below table regarding the choice of feature scaling. There are three scaling methods in this paper. Min-Max Scaler is firstly tested. However, the prediction accuracies using this scaling method are low in general with the best result that is no more than 63%. Therefore, the other two used scalers are mainly evaluated as shown in TABLE II. It is clear that Normalizer actually has better performance on Support Vector Machine, Decision Tree and Random Forest models by a large amount increase of accuracies comparing with Standard Scaler. However, the Standard Scaler works better on Multilayer Perceptron based on Neural Network method and the increase of accuracy is marginal. In general, Multilayer Perceptron model has better adaptability on both scalers.

TABLE II
PREDICTION ACCURACY AND FEATURE SCALING

Model	Standard Scaler	Normalizer
Support Vector Machine	73.98%	**80.14%**
Decision Tree	71.24%	**76.42%**
Random Forest	73.95%	**80.21%**
Multilayer Perceptron	**89.07%**	87.83%

4) Evaluation of Feature Encoding: The feature encoding methods are also evaluated as shown in TABLE III. It is clear that the models using all of three chosen encoding methods have good accuracies on prediction with minor differences. In general, LabelCount encoder has a better performance on tested models than that of others because of the highest prediction accuracies on Support Vector Machine, Random Forest and Multilayer Perceptron, while Hybrid Encoder suits Decision Tree best. It is worthy-noting that Binary encoder shows an increase of feature dimension without deteriorating much on the accuracy, which might owe to its characteristics separating categories well on a high dimensional feature space. According to above evaluation, the conclusion can be drawn that Mulitlayer Perceptron shows the best results among all choices of feature engineering comparing with other machine learning models, where the highest prediction accuracy reaches 89.05%.

TABLE III
PREDICTION ACCURACY AND FEATURE ENCODING

Model	LabelCount	Binary	Hybrid
Support Vector Machine	**80.14%**	78.36%	79.15%
Decision Tree	76.35%	76.38%	**76.42%**
Random Forest	**80.21%**	78.87%	79.33%
Multilayer Perceptron	**89.07%**	88.55%	88.82%

The confusion matrix is also presented in Fig.9, from which we can evaluate the performance of the built model with the existence of skew data. It is obvious that the prediction model works well on classifying non-delay class, as well as high delay and severe delay classes with at least 92% accuracy. For the slight delay class where a delay is less than 60 minutes' arrival, the performance is also good enough with 87% prediction accuracy. However, for medium delay class, the prediction model shows relatively weak ability distinguishing it from the slight delay class with only 78% accuracy. In general, the model is most likely to classify the delay class to the true one or adjacent ones, which suggests good robustness on prediction tasks.

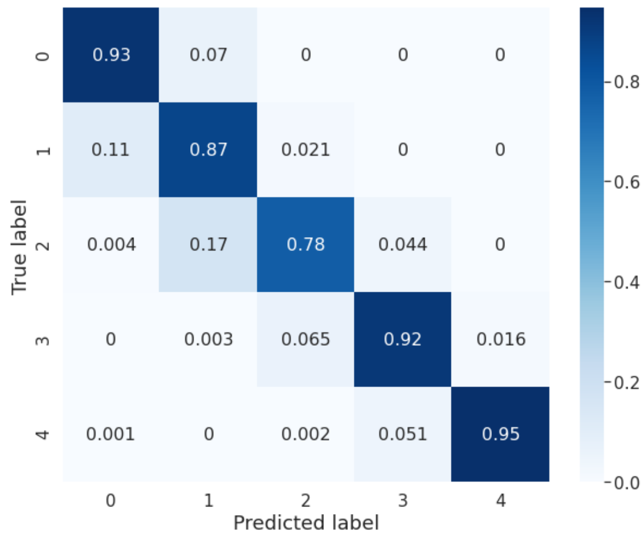

Fig. 9. Confusion Matrix of flight delay prediction

C. An Attempt Using Convolutional Neural Network.

From the above results and evaluations, it can be easily found that neural network method has a significant advantage on this dataset since its best accuracy is 8.86% larger than that of second-best one. It also suggests a potential to use deep neural networks for this task. Convolutional neural network is one of deep learning method which is most commonly used in visual images tasks with satisfactory results. This method uses many filters to extract the spatial information and construct feature maps for the next layer of network so that it can capture high dimensional features, which is suitable for complex tasks. Enlightening by this idea, we treat features as a pattern and rearrange its shape as a map so that it can be used as the input of convolutional neural networks, expecting that useful patterns can be found by the built model. For feature engineering, Standard Scaler is used due to the better performance on neural network models and LabelCount encoder proves to be the only encoding method suiting this model due to a large increase of accuracy(over 23%) compared with others in experiments. This actually makes sense since Convolutional Neural Networks rely on the spatial characteristic of data, while encoding methods that increase the dimensions of features ruffle the spatial information and impede feature learning if the feature map is not carefully designed previously.

The architecture of this model is presented in Fig.10. Dense connections between convolutional layers are used along with 1x1 convolution kernels so that the features can be reused in different levels and the training of neural networks will be easier, where the original idea is from ResNet [16] and DenseNet [17]. Moreover, the accuracy using this architecture is about 0.4% higher than that of a plain network according our experiments, suggesting the advantages of a feature learning process that is just mentioned. The training and validating configurations are the same as previous models. Besides, the batch size is 256, the initial learning rate is 0.0035, and the number of training epochs is 130. Cross entropy is selected as the loss function with Adam optimizer. PReLU [18] is used as the activation function instead of ReLU where the slope of function can be learned to fit the negative part of the scaled input data with a negligible increase in training cost. The final prediction accuracy for this model is 89.32% which is slightly better than that of Multilayer Perceptron, suggesting the advantages of applying neural network methods and feasibility of further researches based on deep learning models.

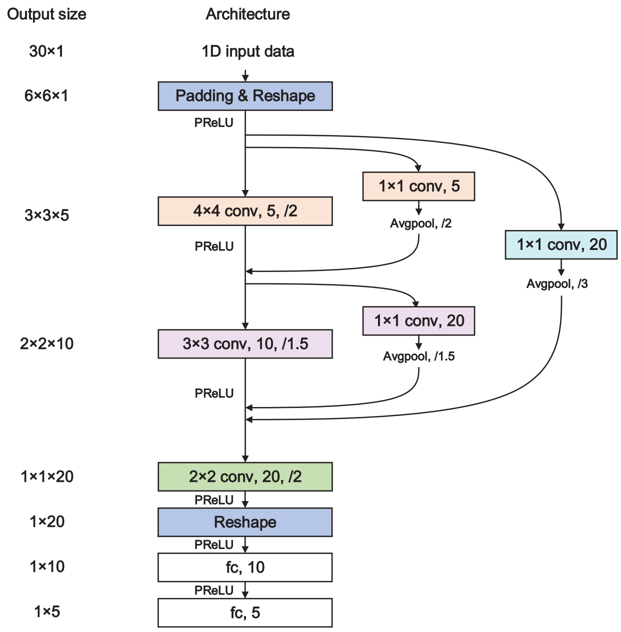

Fig. 10. Proposed Convolutional Neural Network Architecture

V. Conclusion And Future Work

This paper provides a comprehensive aviation data analytic regarding flight delay. QCLCD and AOTP data are used to construct a new dataset with both flight information and weather condition. Then this dataset is further explored and some useful pattern toward flight delay is shown. After that, prediction models using machine learning techniques are built and compared, with the best accuracy of 89.07% in Multilayer Perceptron model. Furthermore, the feasibility of using convolutional neural networks is discussed and such a model is built with 89.32% accuracy, showing a research trend toward flight delay prediction. For the future work, we suggest to pre-design the arrangement of features when using encoding methods increasing the feature dimensions in order to leverage the advantage of convolution neural network on high dimensional feature space.

Acknowledgment

This research was supported by the Center for Advanced Transportation Mobility (CATM), USDOT Grant 69A3551747125.

References

[1] G. Dartmann, H. Song, and A. Schmeink, *Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things*. Elsevier, 2019.

[2] 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.

[3] Y. Sun, H. Song, A. J. Jara, and R. Bie, “Internet of things and big data analytics for smart and connected communities,” *IEEE Access*, vol. 4, pp. 766–773, 2016.

[4] F. Kong, J. Li, B. Jiang, T. Zhang, and H. Song, “Big data-driven machine learning-enabled traffic flow prediction,” *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 9, p. e3482, 2019.

[5] R. Dhanawade, M. Deo, N. Khanna, and R. V. Deolekar, “Analyzing factors influencing flight delay prediction,” in *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2019, pp. 1003–1007.

[6] 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.

[7] 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.

[8] 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)*, 2019, pp. 102–107.

[9] M. Lambelho, M. Mitici, S. Pickup, and A. Marsden, “Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions,” *Journal of Air Transport Management*, vol. 82, p. 101737, 2020.

[10] 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.

[11] Y. Liu, Y. Liu, M. Hansen, A. Pozdnukhov, and D. Zhang, “machine learning to analyze air traffic management actions: Ground delay program case study,” vol. 131, pp. 80–95, 2019.

[12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[13] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, “A deep learning approach to flight delay prediction,” in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 2016, pp. 1–6.

[14] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, “Flight delay prediction based on aviation big data and machine learning,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2020.

[15] Q. Luo, J. Li, S. Tang, R. Deng, Q. Liu, F. Xu, C. Li, H. Zhao, X. Weng, and N. Zhou, “Impact on flight delay predictions based on route network of multi-airport under joint operations in multi airport,” 03 2017, pp. 151–159.

[16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.