Enhancing Flight Delay Prediction Using Residual Neural Networks (ResNets)

Mona Hassan Asiri[1], Abdullah S. AL-Malaise AL-Ghamdi[2], Ayman G. Fayoumi[1], Mahmoud Ragab[3]

[1]*Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia*
[2]*Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Jeddah 34801, Saudi Arabia*
[3]*Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia*
Email: mhussainalasiri@stu.kau.edu.sa

Abstract: The most concern in airline sectors is flight delays because they have a big impact on airlines, passengers, and airports. This study used Residual Neural Network (ResNets), XGBoost, and LightGBM. The aim is to enhance flight delay prediction using ResNets. The performance of ResNets, which is a deep learning model, has been compared to the performances of XGBoost and LightGBM models, which are machine learning models. The dataset used is the domestic flights of United States from January 2019 to August 2023. The confusion matrix is used to make the comparisons between the selected models by summarizing prediction results, which are F1-score, accuracy, sensitivity, and precision. In addition, the validation and models' information, such as file size and prediction time, are used to assess the models' performance. The ResNets models have the best results, followed by LightGBM. The XGBoost has the worst results compared to other models.

Keywords: Residual Neural Network, XGBoost, LightGBM, flight delay, ResNets

1. Introduction

In the airline sector, flight-delays are considered an important concern [1]. They can harm the passengers, cause airlines financial losses, and affect air traffic management. For example, in the US alone, flight-delays annually cost billions in operational inefficiencies and lost productivity [2]. Utilizing accurate models for flight-delay prediction can help airlines optimize their schedules, improve customer satisfaction, and reduce costs. Flight-delay prediction has become a critical area of research using advanced data analysis and machine-learning techniques to enhance scheduling and minimize disruptions. There are many approaches for predicting Flight-Delays: Machine-Learning Models, Deep-Learning Techniques, Delay Propagation Models, and Hybrid Approaches.

Machine-learning models are algorithms widely used for predicting delays based on historical data, e.g., "Random Forest", "Gradient Boosted Trees", and "Support Vector Machines" [3]. The most used features to train the models are departure times, weather conditions, and airport congestion. Recently, Deep Learning Techniques were applied for predicting flight-delays that achieved high accuracy, such as "Long Short-Term Memory (LSTM) networks" and "attention-based models", which are used for sequential data analysis [4]. Delay Propagation Models like the Late Arriving Aircraft Delay (LAAD) framework link previous arrival delays with subsequent departure delays along the same aircraft's itinerary [1]. Iterative predictions improve accuracy by accounting for cascading effects. Hybrid Approaches Combining multiple data sources (e.g., weather data and air traffic information) enhances prediction reliability [4]. For example, integrating ADS-B real-time trajectory data with machine learning models has shown significant improvements in accuracy.

This study aimed to use Residual Neural Network (ResNets) to enhance flight-delay prediction. In addition, the XGBoost and LightGBM had been applied to the same flight data to compare the performance of ResNets to their performance. XGBoost and LightGBM are machine learning models, and ResNets is one of the deep learning techniques. They are classification models, where the flight-delay prediction in this study is a binary classification used only for "departure" flights. The comparison used the resulting cross-validation and optimization metric, which is AUC.

The paper structure is related-work, methodology, implementation, results-discussion, and conclusion-future work.. In the "Related Work" part, brief overviews of prior research in the same field were reviewed. In the Method part, the detailed steps of the case study are presented. In the Implementation part, the setup and the dataset used for the study are described. The outcomes of the implementation part were viewed and analyzed in detail in the results and discussion part. Finally, the Conclusion and Future Work part summarized the findings and proposes directions for future research.

2. Related work

Many studies have used different methods to predict flight-delays, such as machine learning and deep learning [5]. Deep learning techniques provide better precision than conventional machine-learning models, but conventional machine-learning models remain viable with proper settings. Some examples of conventional machine-learning used for flight-delay prediction are "Random Forest" [6], [7], MultiLayer Perceptron (MLP) [8], Decision Tree [9], Gradient Boosting Machine [10], LightGBM[11], and Logistic Regression [12]. Some examples of deep-learning techniques that have been used to predict flight-delays are "Convolutional Neural Networks" [13], "recurrent neural network (RNN)" [14], and "Long Short Time Memory (LSTM)" [15].

The author of [9] used the stacking model to predict flight-delays. The dataset was flight data from "Logan Airport (2019)" in the US, using both departure and arrival flights. The list of machine-learning models were "Decision Tree", "Random Forest", "Gaussian Naïve Bayes", "KNN", "Logistic Regression", and "Stacking" were selected to train the same dataset under the same setting. Multiple indexes, such as "F1 Score," "Accuracy," "Recall," "Precision," "AUC Score," and "ROC Curve," were applied to assess the selected models. The stacking model showed great results in prediction, which could improve the prediction results of flight-delay, followed by Random Forest. The least results of prediction were Naïve Bayes and Logistic Regression.

A study by Jiang, Liu et al. (2020) [8] was conducted to provide a robust flight delay prediction and meaningful patterns. Two datasets were used: "Airline On-Time Performance (AOTP) Data" and "Quality Controlled Local Climatological Data (QCLCD)" (2016). The list of machine-learning models was "Support Vector Machine," "Decision Tree," "Random Forest," and "Multilayer Perceptron". The "Multilayer Perceptron" model gave excellent results followed by SVM and RF. The decision tree model provided the least result.

The authors in [16] used machine-learning models for flight departure delays prediction. The study was applied to "the Nanjing Lukou International Airport" in China from (Mar-2017) to (Feb-2018). The list of used models was LightGBM, extremely randomized trees, SVM, and multiple LR. The LightGBM provided the best prediction results compared to other models used in this study.

The authors in [17] applied machine-learning models to the flight dataset in "Türkiye" integrated with related weather condition data for flight delays prediction. The machine-learning models: "CatBoost", "LightGBM", and "XGBoost" were applied, which are kind of Gradient Boosting. The XGBoost provided the highest prediction results compared to other models. The other models had the same prediction results.

A study by Li Q, Jing R (2022) [18] aimed to provide a novel framework for enhancing flight delay prediction by using ST-RF and LSTM. Random-Forest used as classifier and LSTM used to derive the temporal properties. the suggested model delivered high degree of accuracy, which equals to 92%.

Another study by Chaitanya G, Shaik D et al [7] used ensemble machine-learning models to predict flight delays. The study used "Logistic Regression (LR)", "Neural Networks (NN)", "Random Forest (RF)", and "Decision Trees (DT)". NN+RF and NN+RF+LR were used as ensemble machine learning models. The comparison between all applied models showed that NN+RF+LR provided the highest prediction results compared to other applied models.

In [19], machine-learning models were used for flight delay prediction using flight data in Saudi Arabia. The models used were CatBoost, LightGMB, XGBoost, MLP, and RF. CatBoost's prediction results presented the highest percentage of accuracy among the selected models.

3. Method

This study used Residual Neural Network (ResNets) to enhance the flight delay prediction. In addition, XGBoost and LightGBM were applied to the same flight data to compare ResNets's performance to their performance. XGBoost and LightGBM are machine learning models, and ResNets is one of the deep learning techniques. They are classification models because the flight delay prediction in this study is a binary classification used only for "departure" flights. The comparison used the resulting cross-validation and optimization metric, which is AUC.

1.1. Data source

The dataset used is the domestic flights of the United States from January 2019 to August 2023 downloaded from Kaggle website[20]. It has thirty features and three million rows, the features described in detail in Table 3.1. Its size is "600MB" compressed and 10 GIB in uncompressed CSV format. The feature types are seven categorical, twenty-two numeric, and one date.

Table 3.1: The features of the dataset used in this study

feature	Values	Feature	Values	Feature	Values
FL_DATE	2019-2023	WHEELS_OFF	local, hhmm	DISTANCE	Miles
AIRLINE_CODE	String	WHEELS_ON	local, hhmm	TAXI_IN	Minutes
DOT_CODE	Code	ELAPSED_TIME	Minutes	TAXI_OUT	Minutes
FL_NUMBER	int	CRS_ELAPSED_TIME	Minutes	CANCELLED	0 or 1
DEP_TIME	local, hhmm	AIR_TIME	Minutes	DELAY_DUE_CARRIER	Minutes
CRS_DEP_TIME	local, hhmm	ARR_DELAY	Minutes	DELAY_DUE_WEATHER	Minutes
ARR_TIME	local, hhmm	DEP_DELAY	Minutes	DELAY_DUE_NAS	Minutes
CRS_ARR_TIME	local, hhmm	ORIGIN	Code	DELAY_DUE_SECURITY	Minutes
DIVERTED	0 or 1	DEST	Code	DELAY_DUE_LATE_AIRCRAFT	Minutes
ORIGIN_CITY	City Name	DEST_CITY	City Name		
CANCELLATION_CODE	“A=carrier, B=weather, C=NAS, D=security, F=Late aircraft"				

1.2. Models

As mentioned above, the models used in this study were a deep Residual Neural network (ResNets), an Extreme-gradient Boost tree (XGBoost), and a Light-Gradient Boost machine (LightGBM). XGBoost and LightGBM are machine learning models, and ResNets is a deep learning technique. All used models are classification models. The flight delay prediction in this study is a binary classification used only for departure flights.

1.2.1. Residual Neural network (ResNets)

ResNets are technique of deep neural networks techniques [21,22]. It results from stacking simple residual blocks that are represented as:

$$Yj+1 = Yj + F(Yj , \theta j) \qquad \text{for} \qquad j = 0, ..., N - 1. \qquad (1)$$

Where Yj = the values of the features at the jth layer and θj= the jth layer’s network parameters. There are many applications of ResNets in many fields, like computer vision tasks such as image recognition [21], object detection [23], NLP tasks such as speech synthesis [24], classification prediction of minerals [25], speech recognition [26], and machine translation [27]. In addition,

1.2.2. Extreme-Gradient Boost tree (XGBoost)

XGBoost is a highly effective gradient-boosting machine implementation [28]. It is an algorithm renowned for its great performance in supervised learning because of its optimized and scalable design, which delivers accurate predictions across various applications. Because of its high execution speed, data scientists preferred using XGBoost in problems such as regression or classification problems [29].

1.2.3. Light Gradient Boost machine (LightGBM)

LightGBM is a machine-learning model using gradient descent optimization and decreasing the loss by iterative weight updates [30]. Its outstanding speed and accuracy are due to its unique features like the histogram method to select the optimal split points, the growth strategy of a leaf-wise, and a specialized data storage structure. There are several key components for the LightGBM architecture: Histogram-based Learning, Exclusive Feature Bundling, Parallel and GPU Learning, Gradient-based One-Side Sampling, Leaf-wise Tree Growth, and Regularization.

2. Implementation

This part details the feature selection and engineering procedures and the models' evaluation methods.

2.1. Features Selection and Engineering

The flight data used has thirty features and the selected features are fourteen features which are: Fl_Date, Airline_Code, Origin, Dest, Dep_Time, Dep_Delay, Arr_Time, Arr_Delay, Distance, Delay_Due_Carrier, Delay_Due_Weather, Delay_Due_Nas, Delay_Due_Late_Aircraf, Delay_Due_Security. The next step was null value removal. The null values were removed from these columns to ensure the models would not be trained on incomplete data. The delayed column was created to determine if the flight is delayed or on-time by using the if-statement as follows: If (DEP_DELAY >15, 1,0), which 1 represents the delayed flight and 0 represents the on-time flight. Finally, the processed data was split into three sets: 65% for training, 20% for holdout, and 15% for validation.

2.2. Model Evaluation

The metrics score of learned models is Area Under the Curve (AUC). AUC is defined as a curve comparing the true positive rate of a classifier to its false positive rate, which is varied by the threshold [31]. It is widely used to evaluate the classifier models' performance and compare their results. The values of AUC range from 0 to 1, where (1) refers to perfect classification, (0.5) suggests performance equivalent to random guessing, and less than (0.5) refers to worse-than-random performance [32].

3. Result and Discussion

This section presents the performance of residual Neural Networks (ResNets), XGBoost, and LightGBM after training. Then, the results of the models are compared. The total number of selected features is fifteen and the target features is Delayed.

3.1. Residual Neural Networks (ResNets)

The model used informative features and training schedule, 1 layer and 64 units. The applied structure of ResNets presents in Figure 5.1. The features split into three categories which numeric variables, text variables, and categorical variables. After data preprocessing, Smooth Ridit Transform is used before model start training. The last step is ResNets model.

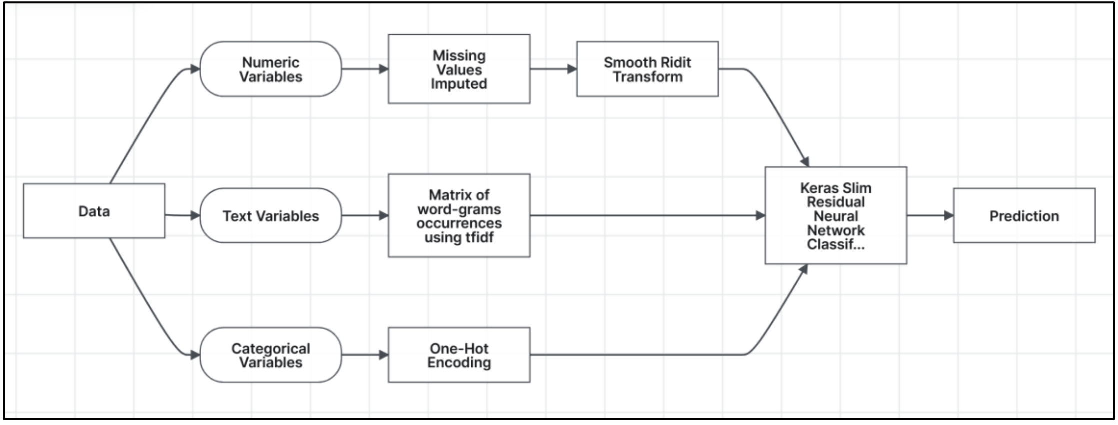


Figure 3.1: The overall structure of ResNets that applied in this study.

The model file size is 37 MB. The prediction time is 0.0528s. The features impact that analyzed by ResNets Classifier using Training Schedule (1 Layer: 64 Units) using 5000 rows is represented in Figure 3.2. The Delay_due_carrier and Delay_due_late_aircraft have the most feature impact on flight delay. Dep_time has a high percentage of impact on delayed flights. Day of month and year have the lowest percentage of feature impact on flight delay.

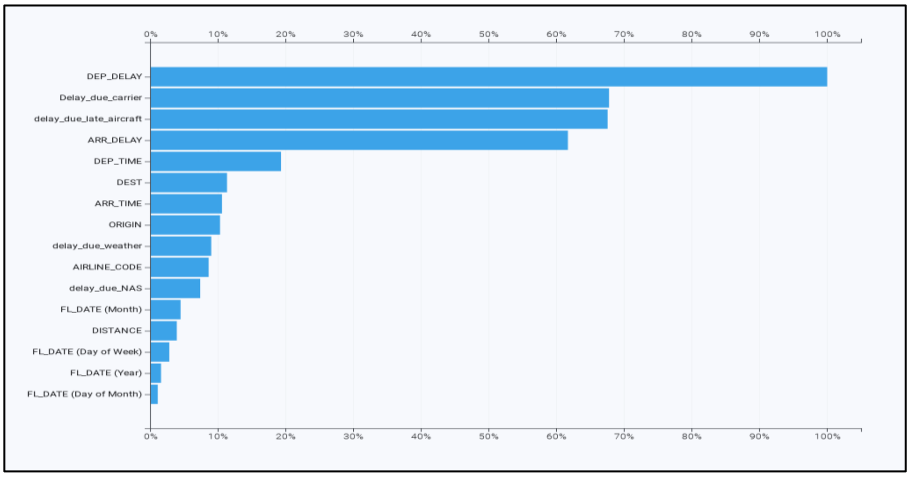


Figure 3.2: The features impact that analyzed by (ResNets)

The all-data SHAP distribution of ResNets has been calculated and shown in Figure 5.3. SHAP distribution visualizes the distribution and density of SHAP scores for individual feature values. The following features, Delay_due_carrier, Delay_due_late_aircraft, Dep_time, and Dest, have a high score of impact on flight delay compared to other selected features.

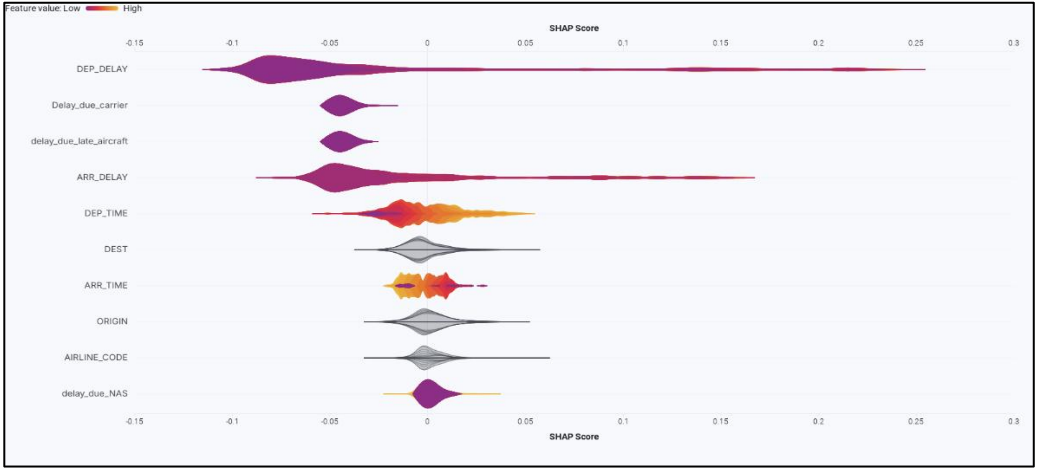


Figure 3.3: All-data SHAP distribution (ResNets)

The model's performance evaluation is 0.99 for validation, 0.9900 for cross-validation, and 0.99 for holdout. The lift chart of the ResNets model, which compares actual and predicted values to visualize how well a model performs for different ranges of the target variable, is shown in Figure 3.4.

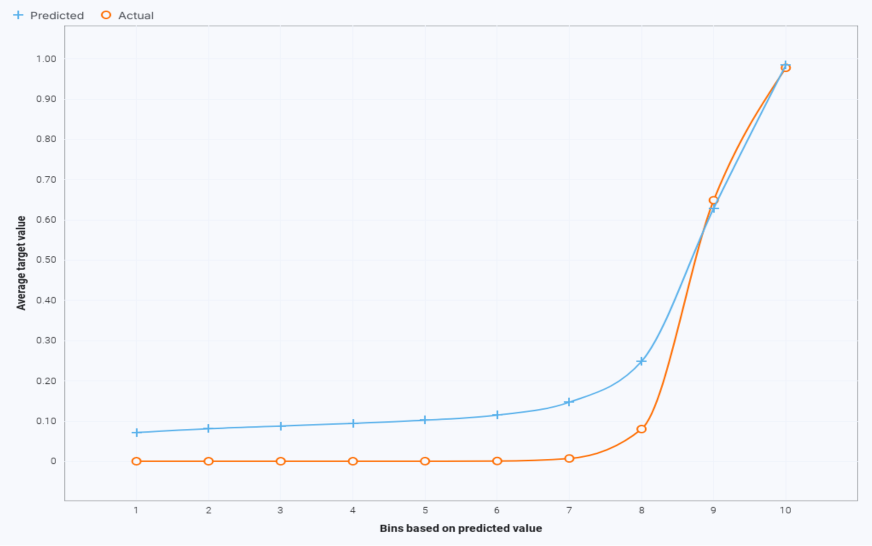


Figure 3.4: The lift chart of ResNets model.

The closer the ROC curve is to the top-left corner of the graph, the more accurate the test. This is because the sensitivity in the top-left corner equals one, and the false-positive rate is 0 (specificity = 1).

So, the ROC Curve of ResNets is considered an excellent curve. It is presented in Figure 3.5, and the prediction distribution of the ResNets model is shown in Figure 3.6.

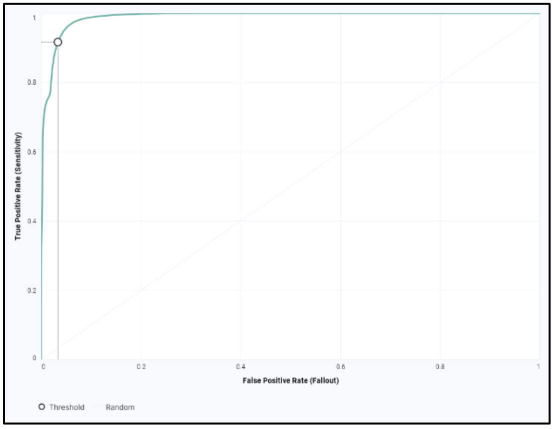


Figure 3.5: The ROC curve of ResNets

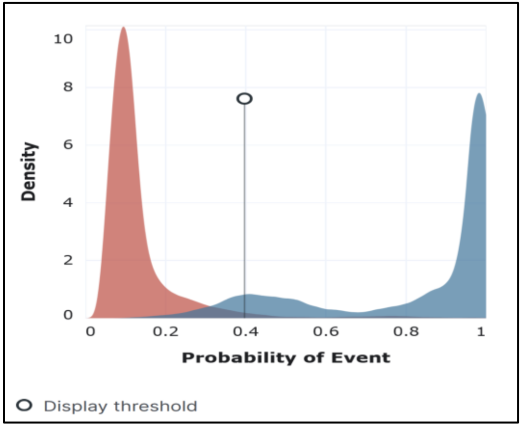
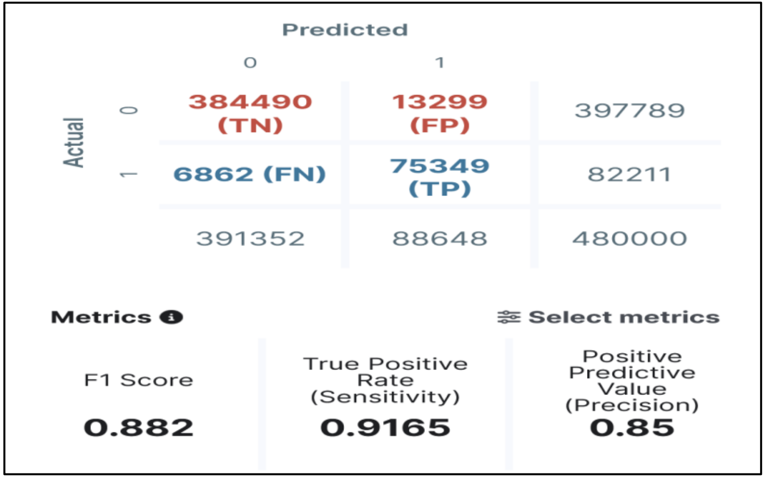


Figure 3.6: The prediction distribution of the ResNets model .

The last point in ResNets model is the resulting confusion matrix. It is presented in Figure 3.7. The confusion matrix presents that the F1 Score is 88%, precision is 85%, and sensitivity is 91%. Finally, the ResNets' accuracy is equal to 96%..



Predicted

Actual	0	1	
0	384490 (TN)	13299 (FP)	397789
1	6862 (FN)	75349 (TP)	82211
	391352	88648	480000

Metrics ℹ

☰ Select metrics

F1 Score	True Positive Rate (Sensitivity)	Positive Predictive Value (Precision)
0.882	0.9165	0.85

Figure 3.7: the confusion matrix of ResNets model.

3.3 Extreme Gradient Boost Tree (XGBoost)

The model used informative features and training schedule, 1 layer and 64 units. The applied structure of XGBoost presents in Figure 3.8. The features split into two categories which numeric variables and categorical variables. After data preprocessing, the XGBoost model is modeled.

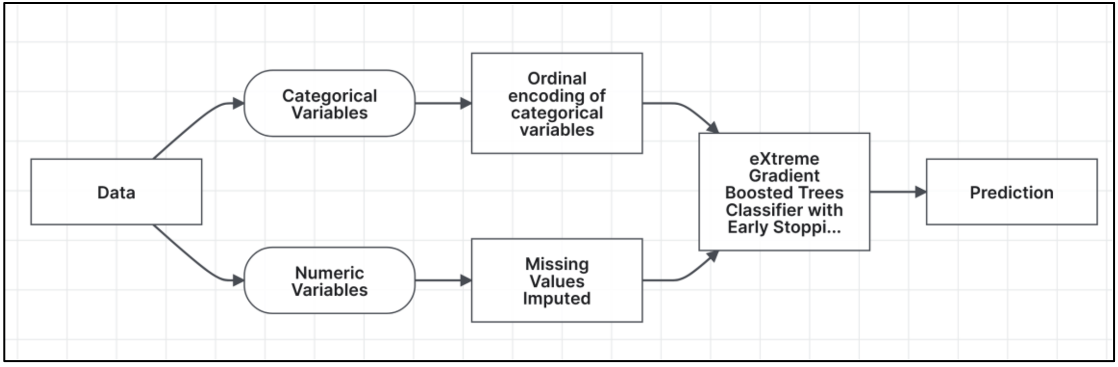


Figure 3.8: the overall structure of XGBoost .

The model file size is 154.202 MB. The prediction time is 0.2044s. The features impact that analyzed by XGBoost Classifier with learning rate =0.13 and Fast Feature Binning is represented in Figure 3.9. The Arr_delay has the greatest impact than other features followed by Dep_time, Arr_time, and Distance. Delay_due_weather presented the lowest impact on the flight delay, followed by day_of_week, day_of_month, and year.

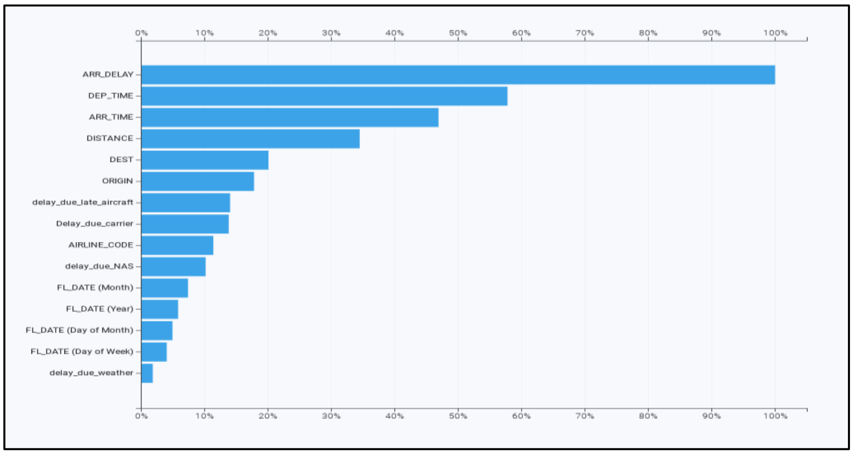


Figure 3.9: The features impact analyzed by XGBoost model.

The all-data SHAP distribution of (XGBoost) model has been calculated and shown in Figure 3.10. The SHAP distribution visualizes the distribution and density of SHAP scores for individual feature values. The distance has a high score of impact on flight delay compared to other selected features.

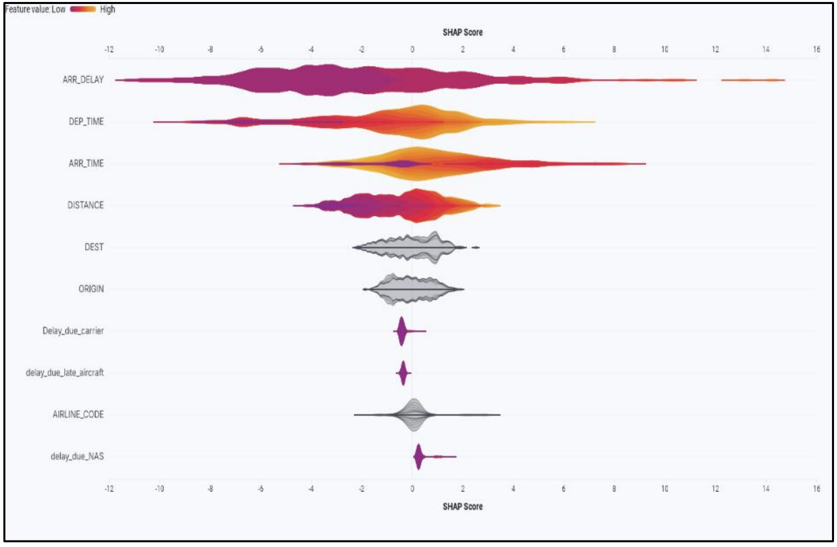


Figure 3.10: All data SHAP distribution (XGBoost)

The model's performance evaluation is 0.9858 for validation, 0.9860 for cross-validation, and 0.9860 for holdout. The lift chart of the XGBoost model, which compares actual and predicted values to visualize how well a model performs for different ranges of the target variable, is shown in Figure 3.11.

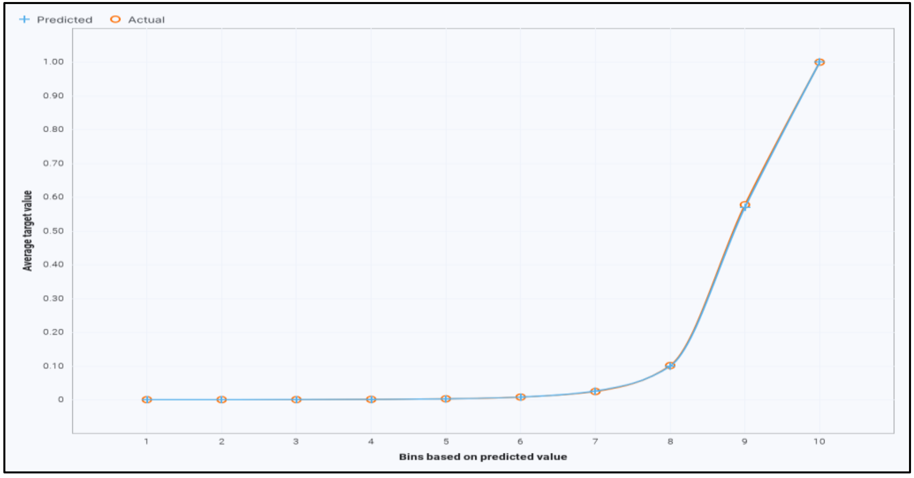


Figure 3.11: The lift chart of XGBoost model.

The closer the ROC curve is to the top-left corner of the graph, the more accurate the test. This is because the sensitivity in the top-left corner equals one, and the false-positive rate is 0 (specificity = 1).

So, the ROC Curve of ResNets is considered an excellent curve. It is presented in Figure 3.12, and the prediction distribution of the XGBoost model is shown in Figure 3.13.

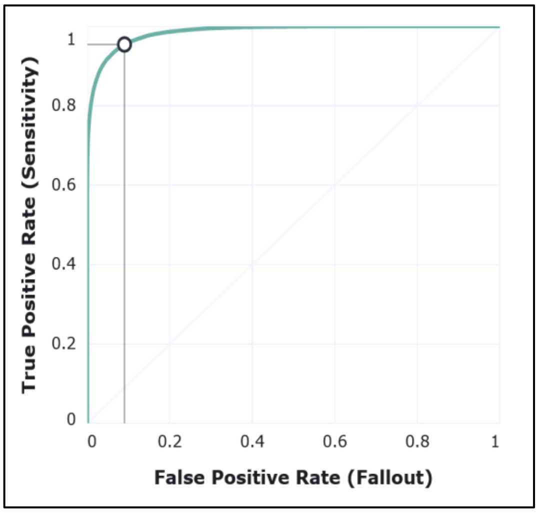


Figure 3.12: The ROC curve (XGBoost)

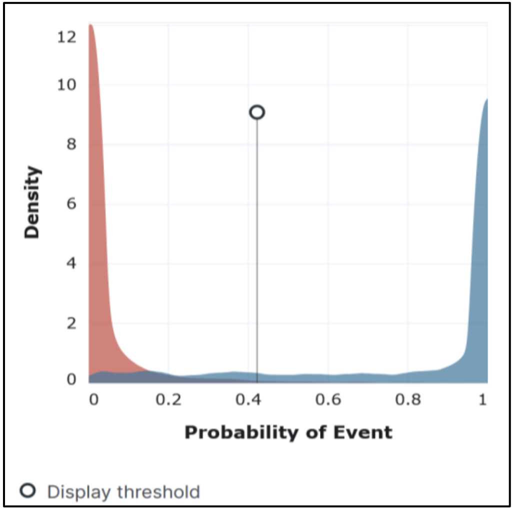
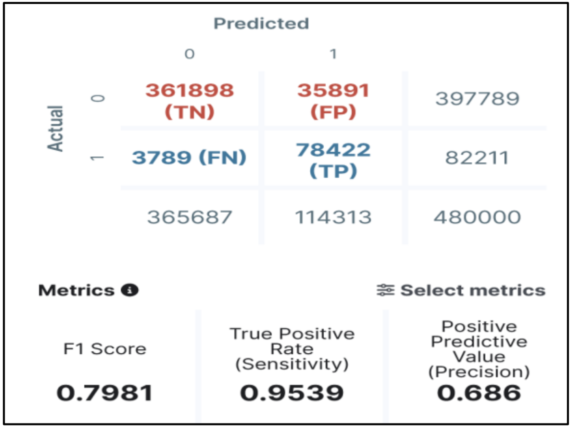


Figure 3.13: The prediction distribution of the ResNets model .

The last point in XGBoost model is. It is presented in Figure 3.13. The confusion matrix presents that the F1 Score is 79%, precision is 68%, and sensitivity is 95%. Finally, the XGBoost's accuracy is equal to 92%.



	Predicted 0	Predicted 1	
Actual 0	361898 (TN)	35891 (FP)	397789
Actual 1	3789 (FN)	78422 (TP)	82211
	365687	114313	480000

Metrics **Select metrics**

F1 Score	True Positive Rate (Sensitivity)	Positive Predictive Value (Precision)
0.7981	0.9539	0.686

Figure 3.13: The resulting confusion matrix (XGBoost)

3.2. Light Gradient Boost Tree (LightGBM)

The model used informative features. The applied structure of LightGBM presents in Figure 3.14. The features split into two categories which numeric variables and categorical variables. After data preprocessing, the XGBoost model is modeled.

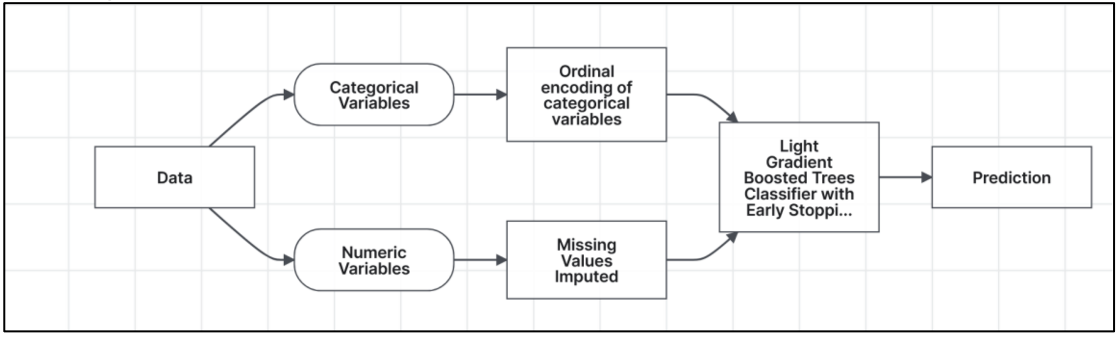


Figure 3.14: The overall structure of LightGBM model.

The model file size is 97.958 MB. The prediction time is 0.1714s. The feature impact for LightGBM Classifier is represented in Figure 3.15. The following features, Dep_time and distance, Dest, and Origin, have a high percentage of impact on flight delay compared to other selected features.

Delay_due_weather presented the lowest impact on the flight delay, followed by day_of_week, day_of_month, and year.

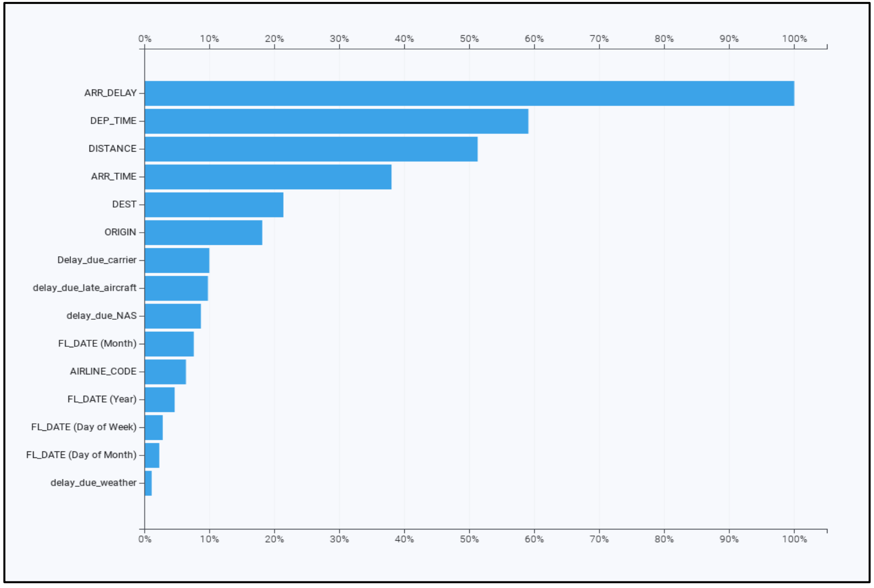


Figure 3.15: the features impact analyzed by LightGBM Model.

The all-data SHAP distribution has been calculated and shown in Figure 3.16. The SHAP distribution of LightGBM model visualizes the distribution and density of SHAP scores for individual feature values. The Dep_time, and Distance have a high score of impact on flight delay compared to other selected features.

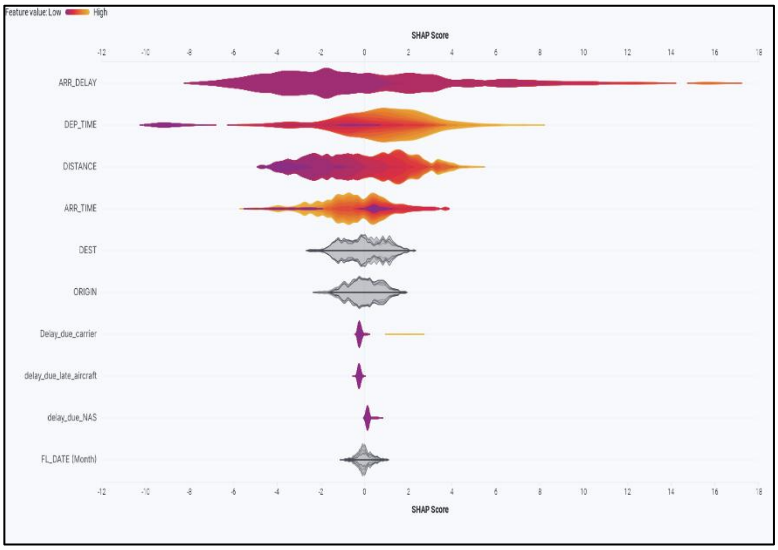


Figure 3.16: All-data SHAP distribution (LightGBM).

The model's performance evaluation is 0.9866 for validation, 0.9867 for cross-validation, and 0.9866 for holdout. The lift chart of the LightGBM model, which compares actual and predicted values to visualize how well a model performs for different ranges of the target variable, is shown in Figure 3.17.

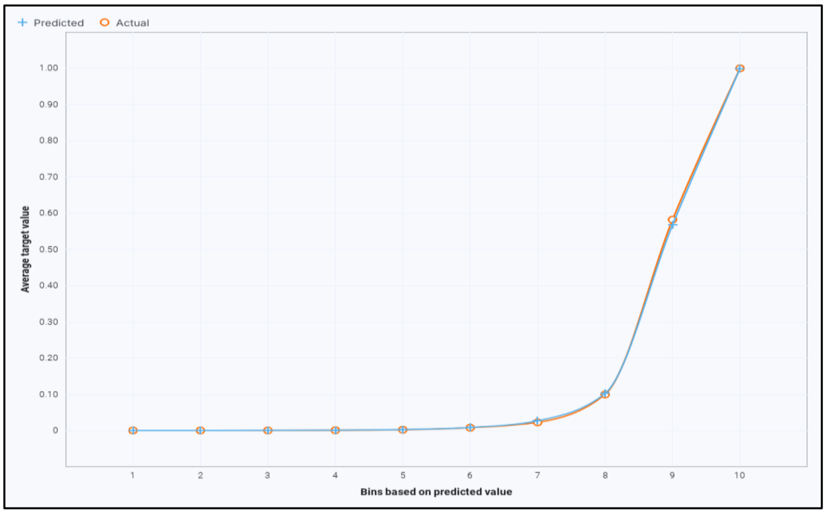
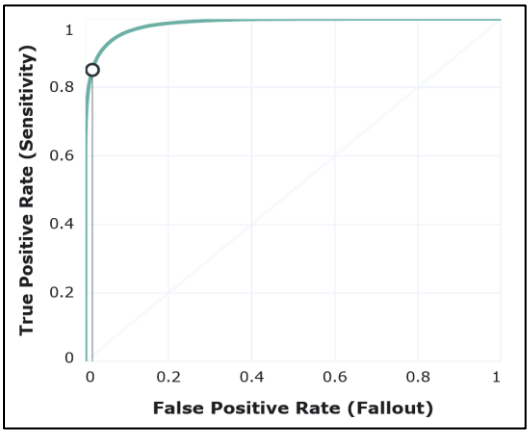


Figure 3.17: The lift chart (LightGBM).

The closer the ROC curve is to the top-left corner of the graph, the more accurate the test. This is because the sensitivity in the top-left corner equals one, and the false-positive rate is 0 (specificity = 1). So, the ROC Curve of LightGBM model is considered an excellent curve. It is presented in Figure 3.18, and the prediction distribution of the LightGBM model is shown in Figure 3.19.



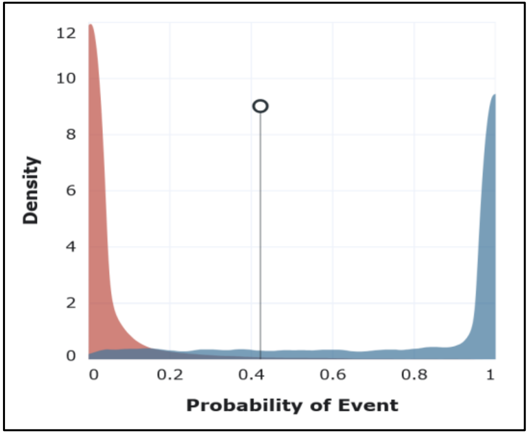
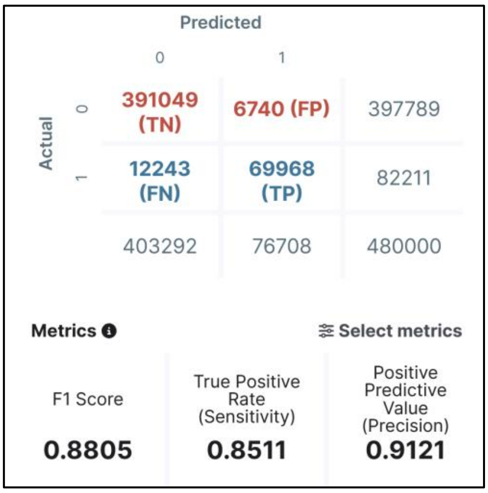


Figure 3.19: The prediction distribution of the LightGBM model.

Figure 3.18: The ROC Curve (LightGBM).

The last point in LightGBM model is the resulting confusion matrix. It is presented in Figure 3.20. The confusion matrix presents that the F1 Score is 88%, precision is 91%, and sensitivity is 85%. Finally, the LightGBM accuracy is equal to 96%.



Actual	Predicted 0	Predicted 1	
0	391049 (TN)	6740 (FP)	397789
1	12243 (FN)	69968 (TP)	82211
	403292	76708	480000

Metrics 🛈 ☰ Select metrics

F1 Score	True Positive Rate (Sensitivity)	Positive Predictive Value (Precision)
0.8805	0.8511	0.9121

Figure 3.20: The confusion matrix (LightGBM).

3.4 Comparison between Models

The results of confusion matrix of all models show that the ResNets model is the best model among other selected models in this study. The detailed confusion matrix results of all models are presented in Table 5.1. The results of the F1 score of ResNets and LightGBM models are equal and the best result (88%) compared to the result of the XGBoost model(79%). The sensitivity results showed that the XGBoost model had the best result (95%), followed by ResNets (91%). The sensitivity of lightGBM is the lowest (85%). The precision results showed that the LightGBM model is the best result (91%), followed by the ResNets model (85%). The XGBoost model's precision result is the lowest (68%). Finally, the accuracy results of trained models showed that the ResNets and LightGBM models are equal and have the best results (96%), whereas the XGBoost model has the lowest accuracy (92%).

Table 3.1: Confusion matrix of all models.

Model	F1 Score	Sensitivity	Precision	Accuracy
Residual Neural Networks (ResNets)	0.88	0.91	0.85	0.96
XGBoost	0.79	0.95	0.68	0.92
LightGBM	0.88	0.85	0.91	0.96

In Figure 3.21, All models presented by using bar chart. There are four bar charts which are F1 score chart, Sensitivity chart, Precision chart, and Accuracy chart.

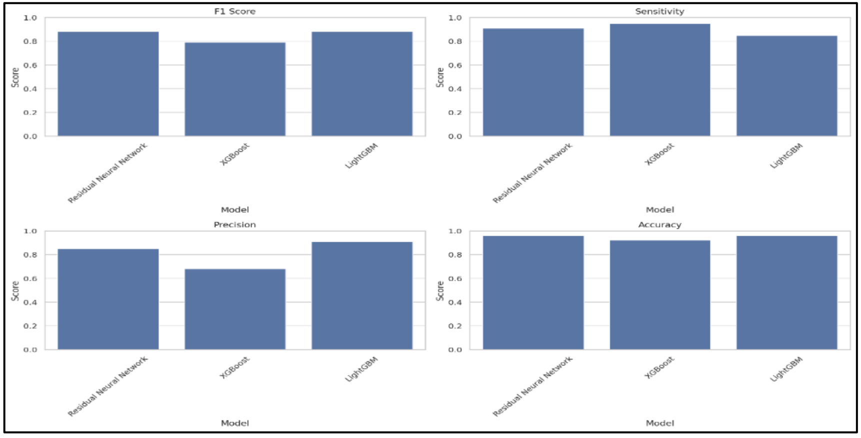


Figure 3.21: The results of confusion matrix for all models.

The validation results of all models are declared in Table 3.2. In addition, the file size and prediction time for each model is presented. The results show that the ResNets model has the best results among other selected models, followed by LightGBM. The XGBoost has the worst results among the other selected models. The validation results showed that the ResNets had the best result (0.99), and the XGBoost and LightGBM models had the worst result (0.98). The file size of ResNets has the lowest result, which is 37 MB; that is, the ResNets model has the best result compared to the other trained models. In addition, the prediction time of ResNets models is the lowest result, equal to 0.0528s; that is, the ResNets model is the best and fastest model compared to other trained models in this study.

Table 3.2: The validation results and models information.

Model	Validation	File Size	Prediction Time
Residual Neural Networks (ResNets)	0.99	37 MB	0.0528s
XGBoost	0.985	154.202 MB	0.2044s
LightGBM	0.986	97.958 MB	0.1714s

4. Conclusion and Future work

The Residual Neural Network (ResNets) model has been used in this paper to enhance the flight delay prediction. In addition, the XGBoost and LightXGM were used to compare the performance of the ResNets model with their performance results. The ResNets is one of the deep learning techniques. While the XGBoost and LightGBM are machine-learning models. The proposed technique contains fourteen significant features selected for the selected models. In addition, a new column has been created: "delayed". In this study, the flight delay prediction is a binary classification problem. The target feature was a delay with two values: (1) a delayed flight and (0) an on-time flight. The selected models' performances have been measured by calculating the confusion matrix, cross-validation, file size, and prediction time. The confusion matrix results were presented regarding "F1-score", "Accuracy", "precision", and "sensitivity". The prediction models used can seriously enhance the prediction results of flight delays. Multiple simulations have been run to show how well the ResNets model performs optimally with various measures. The results showed that ResNets was the best model compared to XGBoost and LightGBM, followed by the LightGBM model. The XGBoost model gave the worst prediction results among the selected models. Thus, the dataset of flight with the selected features and deep-learning techniques (ResNets) can be applied as a valuable tool for flight delay prediction.

In the future, modified deep-learning models, different feature-selection, and effective classification models can be applied to enhance results. The combination of deep learning techniques can also be applied with the same data or with different datasets.

References

[1] M. Li, “Air Traffic Delay Prediction Based on Machine Learning and Delay Air Traffic Delay Prediction Based on Machine Learning and Delay Propagation Propagation,” 2018.

[2] T. Chaudhuri, S. Zhang, and Y. Zhang, “Attention-based Deep Learning Model for Flight Delay Prediction using Real-time Trajectory,” 2024.

[3] Y. Tang, “Airline Flight Delay Prediction Using Machine Learning Models,” ACM International Conference Proceeding Series, pp. 151–154, Oct. 2021, doi: 10.1145/3497701.3497725.

[4] S. Gholami and S. Khashe, “Flight Delay Prediction Using Deep Learning and Conversational Voice-Based Agents.” [Online]. Available: http://asrjetsjournal.org/

[5] A. Makalesi, R. Article Anıl AYAYDIN, Ma. AKÇAYOL Bilgisayar Mühendisliği, G. Üniversitesi, and T. ürkiye, “Learning Based Forecasting of Cancellation, Delay and Orientation on Flights,” 2021, doi: 10.17671/gazibtd.1060646.

[6] D. Shah, A. Lodaria, D. Jain, and L. D’Mello, “Airline Delay Prediction using Machine Learning and Deep Learning Techniques,” International Journal of Recent Technology and Engineering (IJRTE), vol. 9, no. 2, pp. 1049–1054, Jul. 2020, doi: 10.35940/IJRTE.B4047.079220.

[7] G. S. Chaitanya, Dr. S. Shaik, P. Visalakshi, and G. Rakshitha, “Comparative study of ensemble machine learning techniques for airline delay prediction,” International Journal of Engineering in Computer Science, vol. 6, no. 1, pp. 14–21, Jan. 2024, doi: 10.33545/26633582.2024.V6.I1A.105.

[8] Y. Jiang, Y. Liu, D. Liu, and H. Song, “Applying Machine Learning to Aviation Big Data for Flight Delay Prediction,” in Proceedings - IEEE 18th International Conference on Dependable, Autonomic and Secure Computing, IEEE 18th International Conference on Pervasive Intelligence and Computing, IEEE 6th International Conference on Cloud and Big Data Computing and IEEE 5th Cybe, 2020, pp. 665–672. doi: 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00114.

[9] J. Yi, H. Zhang, H. Liu, G. Zhong, and G. Li, “Flight Delay Classification Prediction Based on Stacking Algorithm,” J Adv Transp, vol. 2021, 2021, doi: 10.1155/2021/4292778.

[10] K. Kiliç and J. M. Sallan, “Study of Delay Prediction in the US Airport Network,” Aerospace 2023, Vol. 10, Page 342, vol. 10, no. 4, p. 342, Apr. 2023, doi: 10.3390/AEROSPACE10040342.

[11] Z. Wang, C. Liao, X. Hang, L. Li, D. Delahaye, and M. Hansen, “Distribution Prediction of Strategic Flight Delays via Machine Learning Methods,” Sustainability 2022, Vol. 14, Page 15180, vol. 14, no. 22, p. 15180, Nov. 2022, doi: 10.3390/SU142215180.

[12] T. Nibareke and J. Laassiri, “Using Big Data-machine learning models for diabetes prediction and flight delays analytics,” J Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00355-0.

[13] J. Qu, T. Zhao, M. Ye, J. Li, and C. Liu, “Flight Delay Prediction Using Deep Convolutional Neural Network Based on Fusion of Meteorological Data,” Neural Process Lett, vol. 52, no. 2, pp. 1461–1484, Oct. 2020, doi: 10.1007/S11063-020-10318-4/METRICS.

[14] T. O. ; Yhdego et al., “Analyzing the Impacts of Inbound Flight Delay Trends on Departure Delays Due to Connection Passengers Using a Hybrid RNN Model,” Mathematics 2023, Vol. 11, Page 2427, vol. 11, no. 11, p. 2427, May 2023, doi: 10.3390/MATH11112427.

[15] Y. Yuan, Y. Wang, and C. S. Lai, “Multi-Attribute Data-Driven Flight Departure Delay Prediction for Airport System Using Deep Learning Method,” Aerospace 2025, Vol. 12, Page 246, vol. 12, no. 3, p. 246, Mar. 2025, doi: 10.3390/AEROSPACE12030246.

[16] B. Ye, B. Liu, Y. Tian, and L. Wan, “A methodology for predicting aggregate flight departure delays in airports based on supervised learning,” Sustainability (Switzerland), vol. 12, no. 7, Apr. 2020, doi: 10.3390/SU12072749.

[17] N. T. Irmak Hatıpoğlu , Ömür Tosun, “FLIGHT DELAY PREDICTION BASED WITH MACHINE LEARNING,” Scientific Journal of Logistics, vol. 18, no. 1, pp. 97–107, 2022.

[18] Q. Li and R. Jing, “Flight delay prediction from spatial and temporal perspective,” Expert Syst Appl, vol. 205, p. 117662, Nov. 2022, doi: 10.1016/J.ESWA.2022.117662.

[19] M. Alfarhood, R. Alotaibi, B. Abdulrahim, A. Einieh, M. Almousa, and A. Alkhanifer, “Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines,” International Journal of Aerospace Engineering, vol. 2024, no. 1, p. 3385463, Jan. 2024, doi: 10.1155/2024/3385463.

[20] "Flight Delay and Cancellation Dataset (2019-2023)." Accessed: Apr. 10, 2025. [Online]. Available: https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023?select=flights_sample_3m.csv

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016. Accessed: Apr. 10, 2025. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[22] B. Chang, L. Meng, E. Haber, L. Ruthotto, D. Begert, and E. Holtham, "Reversible Architectures for Arbitrarily Deep Residual Neural Networks," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, pp. 2811–2818, Apr. 2018, doi: 10.1609/AAAI.V32I1.11668.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN".

[24] A. van den Oord et al., "WaveNet: A Generative Model for Raw Audio," Sep. 2016, Accessed: Apr. 10, 2025. [Online]. Available: https://arxiv.org/abs/1609.03499v2

[25] P. Theerthagiri, A. U. Ruby, and J. George Chellin Chandran, "Prediction and classification of minerals using deep residual neural network," Neural Comput Appl, vol. 36, no. 4, pp. 1539–1551, Feb. 2023, doi: 10.1007/S00521-023-09141-4.

[26] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, and U. San Diego, "Aggregated Residual Transformations for Deep Neural Networks," 2017. Accessed: Apr. 10, 2025. [Online]. Available: https://github.com/facebookresearch/ResNeXt

[27] Y. Wu et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," Sep. 2016, Accessed: Apr. 10, 2025. [Online]. Available: https://arxiv.org/abs/1609.08144v2

[28] A. Ibrahem Ahmed Osman, A. Najah Ahmed, M. F. Chow, Y. Feng Huang, and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," Ain Shams Engineering Journal, vol. 12, no. 2, pp. 1545–1556, Jun. 2021, doi: 10.1016/J.ASEJ.2020.11.011.

[29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4.

[30] K. Ramalingam et al., "Light gradient boosting-based prediction of quality of life among oral cancer-treated patients," BMC Oral Health, vol. 24, no. 1, pp. 1–8, Dec. 2024, doi: 10.1186/S12903-024-04050-X/TABLES/1.

[31] T. Yang and Y. Ying, "AUC Maximization in the Era of Big Data and AI: A Survey," ACM Comput. Surv, vol. 55, 2022, doi: 10.1145/3554729.

[32] H. Yang, A. Ligmann-Zielinska, Y. Dou, M. G. Chung, J. Zhang, and J. Liu, "Complex Effects of Telecouplings on Forest Dynamics: An Agent-Based Modeling Approach," Earth Interact, vol. 26, no. 1, pp. 15–27, Jan. 2022, doi: 10.1175/EI-D-20-0029.1.