Explainable Artificial Intelligence applied to active runway configuration and dynamic airport capacity prediction

Igor Galhano Gomes*, Mayara Condé Rocha Murça, Marcelo Xavier Guterres
Aeronautics Institute of Technology

*** Corresponding author e-mail address:** igorgalhanog@gmail.com

PAPER ID: SIT2298

ABSTRACT

Despite the widespread use of machine learning methods in a variety of studies in the aviation industry in recent years, there has been limited progress in achieving explainability for the results generated by these models. Given the importance of explainable predictive models for trustworthy decision support in the context of Air Traffic Management (ATM), this study explores the use of Multilayer Perceptron and Random Forests, along with a Local Interpretable Model-agnostic Explanations (LIME) approach, to develop predictive models capable of generating interpretable forecasts of airport runway configuration and dynamic capacity. The machine learning process is based on historical weather, flight demand and actual air traffic movement data for Sao Paulo/Guarulhos International Airport. The predictive models are able to deliver runway configuration forecasts with an accuracy higher than 87.73% and arrival/departure capacity forecasts with errors as low as 1.309. With LIME, consistent explanations are obtained through the models on each prediction. An interactive predictive tool is created to output the runway configuration and capacity forecasts as well as the main contributing factors for each forecast within a 24-hour forecasting horizon. This research aims to improve decision-making by integrating predictive models and explanatory models into applications that provide useful information for ATM.

Keywords: Airport Capacity, Runway Configuration, Machine Learning, Explainable Artificial Intelligence, LIME.

1. INTRODUCTION

Airport capacity can be defined as the maximum number of takeoffs and landings that an airport can handle within a period of time while meeting all air traffic rules and service level criteria. For long-term planning and strategic demand management, a theoretical capacity is calculated based on specific runway configuration and operational conditions using analytical and simulation models. Some of the key factors considered for this estimation are the average runway occupancy time and the required wake turbulence separations between aircraft (DECEA, 2015).

However, airport capacity can vary significantly during the course of a day due to the dynamics of relevant operational factors such as weather. Imbalances between the planned demand and the dynamic capacity can result in significant flight delays because flights that could not be conducted at the scheduled time due to the capacity constraint will be pushed into a waiting queue, along with subsequent flights. Air Traffic Flow Management (ATFM) is a fundamental service provided by the Air Traffic Management (ATM) system to mitigate the impacts of such imbalances. It is accomplished through strategies to adjust the traffic flows in order to match demand with the available capacity at any given moment. Therefore, for the purposes of Air Traffic Flow Management (ATFM), it is necessary to forecast the dynamic airport capacity in real time based on evolving operational conditions. This dynamic capacity estimation is of particular interest in this study.

One of the main factors that affect the dynamic capacity of an airport is the active runway configuration. The runway configuration is the set of runways and their active thresholds that are being used for arrivals and departures at a given moment, as determined by air traffic controllers (ATCOs). It depends on several factors such as wind direction and speed, ceiling/visibility conditions, flight demand, environmental constraints (e.g., noise abatement) and coordination with other airports located nearby (Avery and Balakrishnan, 2015). Accurately predicting the runway configuration is key to accurately predicting dynamic airport capacity.

Several studies have worked towards the development of Machine Learning (ML) models to predict runway configurations and capacity towards supporting ATFM decisions. However, providing only predictions without explanations is insufficient in the context of ATM decision support, as safety is a priority. ATCOs are responsible for the decisions made and, therefore, need to understand the reasons behind the results generated by machine learning models. This allows them to assess the conditions that led to certain model predictions and possibly validate those predictions.

In this context, Explainable Artificial Intelligence (XAI) arises as a suitable approach to make the results of black-box ML models understandable and explainable to humans. XAI models are capable of generating explanations for ML models, maintaining their predictive performance while providing an understanding of the results obtained. This approach increases confidence in the predictions provided by the models, contributing to building trust in automated decision support systems and enhancing their usability.

This work presents a XAI framework to develop explainable machine learning-based predictive models that can generate interpretable forecasts of airport runway configuration and capacity. We use supervised learning algorithms - Multilayer Perceptron (MLP) and Random Forests (RF) - for learning the predictive models from historical demand, weather and traffic data, along with a Local Interpretable Model-agnostic Explanations (LIME) approach for generating forecast explanations. The models are developed for Sao Paulo/Guarulhos International Airport (GRU). As the largest airport in South America, it often needs ATFM measures to balance demand with capacity. The application of the models is demonstrated with a prototype interactive tool that outputs the runway configuration and capacity forecasts as well as the main contributing factors for the forecasts for a 24-hour planning horizon.

The proposed methodology paves the way for integrating these models into applications that offer a wide range of useful information for ATM, leading to improved decision-making and promoting more effective management of airport capacity and air traffic flows.

2. LITERATURE REVIEW

In the aviation industry, ML methods have been employed in a wide range of studies in recent years. However, there has been limited progress in the explainability of the results generated by these models. Explainability is a field that needs to be further explored for these methods to be accepted by users and authorities (Degas et al., 2022).

For runway configuration prediction, discrete choice models were developed to predict runway configuration at LaGuardia Airport (LGA) and San Francisco International Airport (SFO) over 15-minute, 3-hour, and 6-hour planning horizons, using a recursive approach (Avery and Balakrishnan, 2015). In another study, XGBoost and Random Forests models were created to predict runway configuration at the six largest airports in the United States, also using a recursive approach with prediction horizons of up to 6 hours (Rebollo et al., 2021).

In the field of capacity prediction, various studies have employed deep learning approaches to achieve more accurate results. Models such as Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), and Long-Short Term Memory (LSTM) have been developed to predict airport capacity at Hartsfield-Jackson Atlanta International Airport (ATL), which were then used for capacity prediction at Boston Logan International Airport (BOS) through transfer learning (Choi and Kim, 2021). More complex models have also been developed in other studies, such as attention-based deep learning models that combine MLP and LSTM to predict Airport Arrival Rates (AAR) at Hong Kong International Airport (HKG) (Jiao et al., 2023), and convolutional neural network models to predict both runway configuration and AAR in multi-airport systems using gridded weather forecast data (Wang and Zhang, 2021).

In the context of Explainable AI (XAI) applied to ATM, Woo et al. (2022) developed an interactive interface that presents results and explanations of XGBoost predictions for runway exit using trajectory data and meteorological conditions. LIME explanations were generated based on the predictions of the trained models.

The fields of runway configuration prediction, dynamic capacity prediction, and XAI applied to ATM still have significant potential for further exploration. The application of existing XAI techniques and the development of new ones could lead to the creation of various useful applications for ATM. The contribution of this research lies in the integration of ML and XAI methods for the development of interpretable runway configuration and capacity prediction models.

3. METHODOLOGY

Figure 1 provides a schematic overview of the methodological approach, which is detailed in the next subsections.

3.1. Data Collection and Pre-processing

The airport runway configuration and capacity prediction models were created based on historical meteorological, demand and air traffic data for GRU. Weather data was gathered from Meteorological Aerodrome Reports (METAR) and Terminal Aerodrome Forecasts (TAF). Planned flight demand data was collected from the Flight Operations Registration System (SIROS) of the Brazilian Civil Aviation Agency (ANAC). Finally, actual air traffic movement data was gathered from the Air Traffic Control Tower Management System and FlightRadar24.

The data covers six months of operations from July 1, 2022 to December 31, 2022. Data pre-processing was performed to structure the dataset with observations for every 15-minute period. For each 15-minute period, the scheduled and actual arrival and departure rates were calculated. The active runway configuration was determined based on the most frequently used arrival and departure runways in each interval. Missing meteorological data was filled with previous data, since METAR is issued hourly and TAF every 6 hours. Headwind and crosswind components were calculated based on the wind speed, wind direction and runway threshold orientation. The variables are synthesized in Table 1.

The development of the airport capacity prediction model with supervised learning requires knowledge about the actual airport ar-

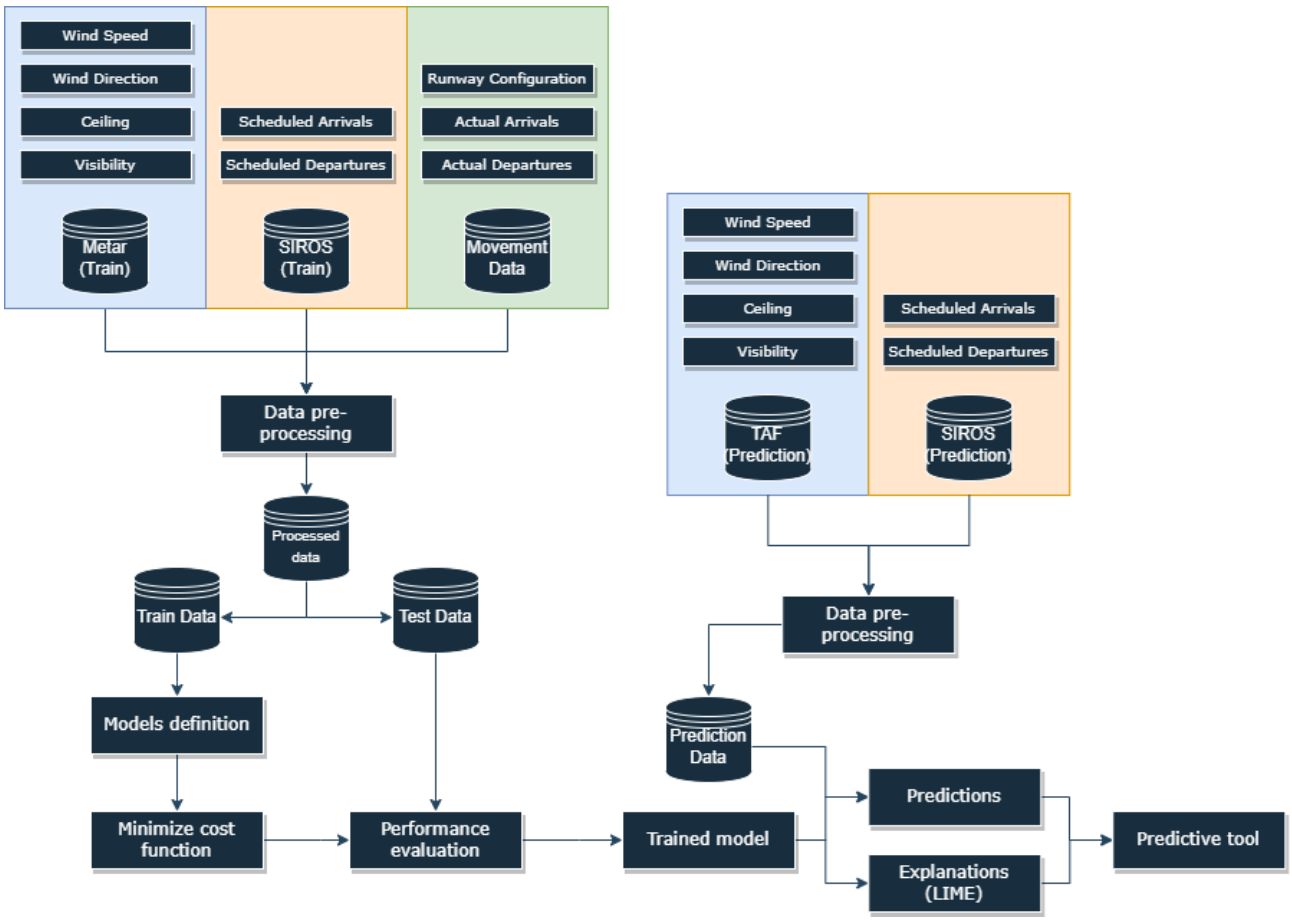


Figure 1 Methodology flowchart.

Table 1 Variables created for modeling

Group	Name	Type
Temporal	local_time	categorical
	day_of_the_week	categorical
Actual operations	runway˙configuration	categorical
	actual_arrivals	numerical
	actual_departures	numerical
Planned operations	scheduled_arrivals	numerical
	scheduled_departures	numerical
Weather	headwind	numerical
	crosswind	numerical
	visibility	numerical
	ceiling	numerical

rival/departure rate (output ”truth”). For this, we used air traffic movement data and developed an empirical method for estimating the capacity from observed airport throughput performance. The purpose of using this method was to identify the periods for which the demand was high enough and the airport was operating under pressure so that arrival/departure rates observed reflected the airport capacity. Otherwise, capacity could be underestimated due to unloaded flight periods. First, the number of aircraft simultaneously in transit from the gate to the runway at the beginning of each interval and the runway throughput for this interval were used to estimate a departure saturation curve for the airport. This curve allows for the estimation of the number of departing aircraft on the ground that leads to runway throughput saturation. After determining this quantity, the data was filtered for periods when saturation occurs, as the number of departures occurring in those periods becomes a good representation of the airport's departure capacity at that moment. Similarly, the number of arriving aircraft within a 40 NM radius from the airport cen-

ter was determined for each time interval, and a saturation curve for landings was estimated. A new database was subsequently generated, containing the time intervals for which arrival saturation occurs.

As additional data preprocessing for modeling and forecasting, one-hot encoding was used for categorical variables, and min-max normalization in the range [0,1] was applied to numerical variables.

3.2. Predictive Modeling

For learning the predictive models, 75% of the dates were used for training, with a portion of this data later separated for validation, and 25% for testing. A classification model was developed for runway configuration prediction and two regression models were developed for arrival and departure capacity prediction, respectively, based on a supervised learning approach. The output variables were *runway_configuration*, *actual_arrivals* and *actual_departures*, while the features were the remaining variables in Table 1. For the supervised learning process, the Multilayer Perceptron (MLP) and the Random Forests (RF) algorithms were used. As non-parametric methods, they allow for non-linear mapping between input and output.

For training the runway configuration prediction model, binary cross-entropy was chosen as the loss function, and accuracy was employed as the primary performance metric. For the capacity prediction model training, the loss function used was the Mean Squared Error (MSE), while the Root Mean Squared Error (RMSE) was the main performance metric chosen. The hyperparameter tuning was performed through manual search for MLP and grid search for RF based on cross-validation.

Recursive runway configuration models were also developed to achieve improved performance for shorter forecasting horizons. These models use the runway configuration predicted for the previous time period as a feature to predict the current one.

With the trained models, weather forecast data from TAF was used for forecasting over a 24-hour prediction horizon, and the performance metric was recalculated for predictions on the test data using TAF.

3.3. Predictive Model Explanation

To obtain interpretable airport runway configuration and capacity forecasts, a Local Interpretable Model-agnostic Explanations (LIME) approach was used to generate explanations for the predictions delivered by the machine learning models. LIME is a group of model-agnostic interpretable models for local explanations. It relies on training simpler surrogate models, such as linear models, to generate explanations for a single prediction.

Explaining a model means providing textual or graphical representations that enable the qualitative interpretation of the relationship between the model's prediction and the variables used to generate that prediction (Ribeiro et al., 2016). Specifically for LIME, explanations involve quantifying the contribution of the most relevant variables in a given prediction, which is determined by the weights obtained for each variable in the surrogate model.

Let f be the trained model to be explained, and g be the surrogate model used to generate explanations, belonging to a model family G. Also, let x be a data instance, and π_x be the neighborhood of x. LIME algorithm optimizes the parameters of g to determine the explanation as follows:

$$\xi(x) = argmin_{g \varepsilon G}(L(f,g,\pi_x) + \Omega(g)), \quad (1)$$

where L is a loss function that measures how unfaithful g is in approximating f in the neighborhood defined by π_x, and Ω is a measure of the complexity of the explanation g (Ribeiro et al., 2016).

In this study, Ridge Regression was used for explanations, based on a weighted Mean Squared Error loss function with an L2 regularization term to penalize complexity. The neighborhood data was obtained through Gaussian sampling.

4. RESULTS AND DISCUSSION

4.1. Exploratory Data Analysis

GRU has two closely spaced parallel runways that are typically operated in two main configurations: $10R|10L$ (i.e., landings on runway threshold 10R and takeoffs on runway threshold 10L) and $28L|28R$ (i.e., landings on runway threshold 28L and takeoffs on runway threshold 28R). Figure 2 displays the frequency of use of these runway configurations based on the processed data.

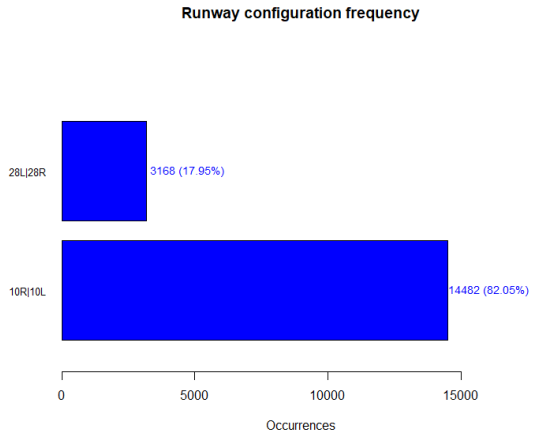


Figure 2 Frequency of use of main runway configurations at GRU.

Figure 3 shows the extremely high correlation between GRU's active runway configuration and headwind component. As expected, runway configurations are typically chosen to avoid tailwind components. Therefore, *headwind* is expected to be a relevant feature for runway configuration prediction.

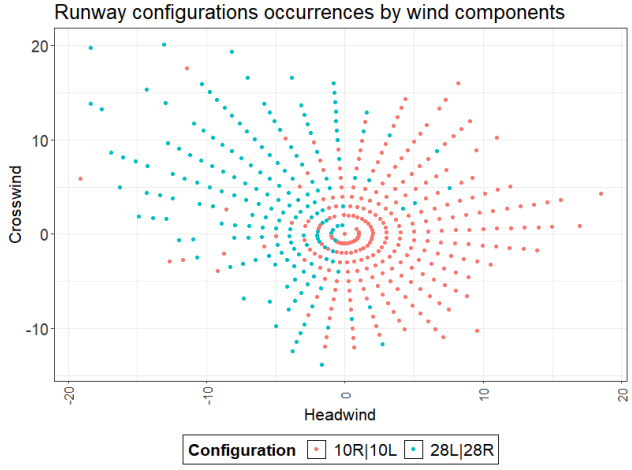


Figure 3 Runway configuration occurrences and observed wind components at GRU.

After the initial data preprocessing described in Section 3., we obtained the arrival and departure throughput performance curves for GRU, as shown in Figures 4 and 5. The airport is found to be saturated for arrivals when the number of aircraft in the terminal area exceeds 9. Similarly, the airport is found to be saturated for departures when the number of aircraft in transit from the gate to the runway exceeds 8.

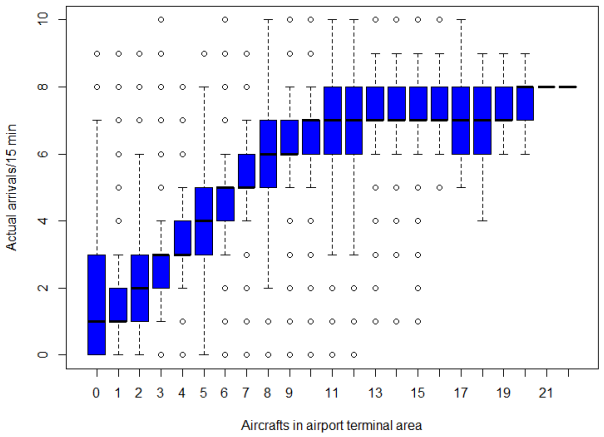


Figure 4 Arrival throughput performance curve.

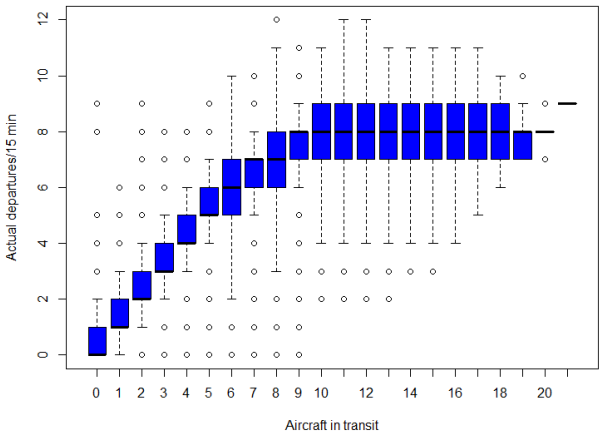


Figure 5 Departure throughput performance curve.

Considering the demand levels described above, the data was then filtered to keep only the observations associated with the periods when the airport was operating at capacity. Figures 6 and 7 show that the filtering process achieved the desired effect, as the distribution of arrival and departure throughput rates became less impacted by the hourly demand patterns. Therefore, it is reasonable to assume that the arrival and departure rates reflect the airport capacity for the filtered instances.

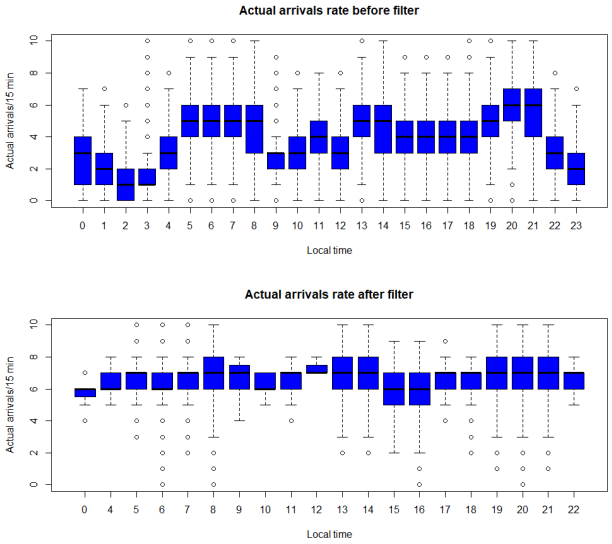


Figure 6 Distribution of arrival rates by hour of the day before and after the filtering process.

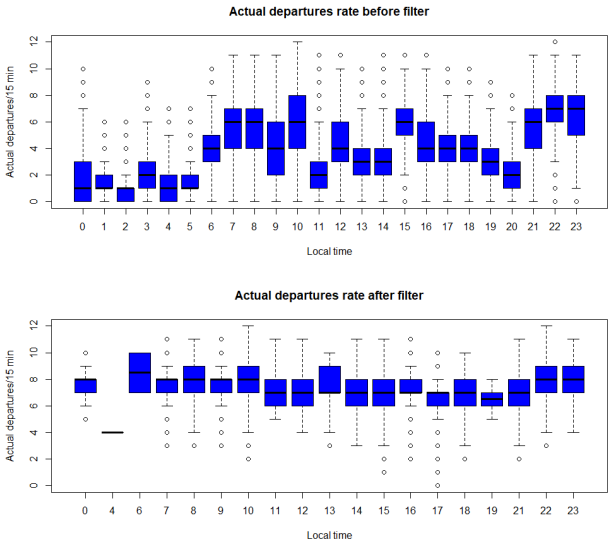


Figure 7 Distribution of departure rates by hour of the day before and after the filtering process.

4.2. Predictive Model Performance

For MLP, through manual hyperparameter tuning with 100 epochs, the architectures shown in Table 2 achieved the best performance. Dropout regularizers were applied to the hidden layers with a probability of 10%. Training was performed using the Adam algorithm with a batch size of 128 and a learning rate of 0.01, with 10% of the training data used for validation.

Table 2 MLP models' architecture

Prediction model	*Neurons*	*Activation function*
Runway configuration	[60, 35, 1]	[tanh, tanh, sigmoid]
Arrival capacity	[60, 35, 1]	[tanh, tanh, linear]
Departure capacity	[60, 35, 1]	[tanh, tanh, linear]

For the RF models, 500 trees were used to train with 4-fold cross-validation and 2 repetitions to optimize the number of features used to generate decisions at each tree node. Figures 8, 9 and 10 display the importance obtained for each feature in the trained RF models. We observed that the most relevant feature for the runway configuration prediction model was *headwind*, consistent with the results shown by the exploratory data analysis. For the capacity prediction models, the most relevant feature was *local_time*.

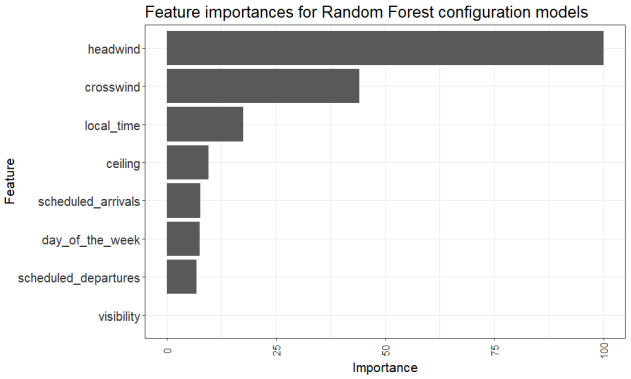


Figure 8 Feature importance for the RF runway configuration prediction model.

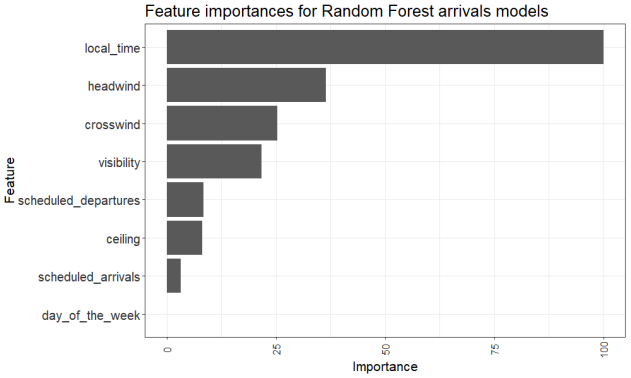


Figure 9 Feature importance for the RF arrival capacity prediction model.

The predictive performance of the models on the test data is summarized in Table 3. Over-

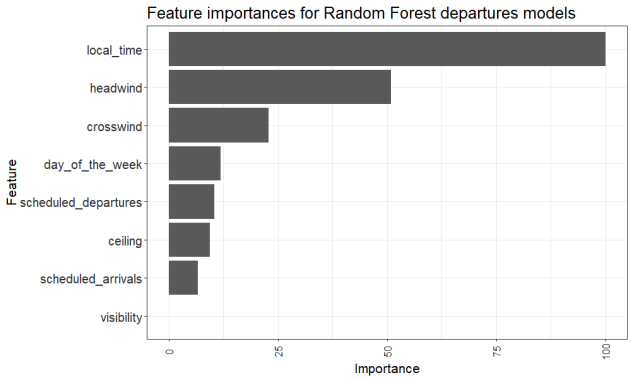


Figure 10 Feature importance for the RF departure capacity prediction model.

all, the MLP models showed a slightly better performance than the RF models, with an accuracy of 87.73% for runway configuration predictions and a RMSE of 1.113 and 1.309 for arrival and departure capacity predictions, respectively.

Table 3 Predictive model performance

Performance	*MLP*	*RF*
Runway configuration prediction accuracy	87.73%	87.05%
Arrival capacity prediction RMSE	1.113	1.156
Departure capacity prediction RMSE	1.309	1.389

Table 4 shows the accuracy obtained for the recursive MLP and RF runway configuration prediction models at different forecasting horizons. The performance of these models is also compared to that of a baseline model, which predicts the runway configuration as the current one (constant prediction) throughout the entire forecasting horizon. It is observed that for short-term forecasting horizons, the previous configuration serves as a good estimate for the next configuration, yielding even better performance than the learned models for a 15-minute horizon. However, for longer forecast horizons, a significant improvement in predictive performance is obtained with the machine learning models.

Table 4 Runway configuration prediction model performance as a function of the forecast horizon

Forecasting horizon	*MLP*	*RF*	*Baseline*
15 minutes	98.52%	98.68%	98.73%
1 hour	98.02%	98.29%	95.13%
2 hours	96.92%	97.49%	91.13%
3 hours	95.73%	96.51%	88.09%
6 hour	92.10%	92.42%	80.65%
12 hours	88.09%	85.12%	76.94%
24 hours	85.63%	80.63%	75.65%

The machine learning models were also compared with an alternative baseline model that predicts the runway configuration based on the headwind signal (if greater than 0, $10R|10L$, and $28L|28R$, otherwise). Figure 11 summarizes the predictive performance of all models as a function of the forecast horizon. It is clearly noticeable that the learned models provided an improvement in prediction accuracy when compared to the baseline models, with the exception of very short-term predictions (15-minute forecast horizon). The longer the forecast horizon, the larger the performance improvement.

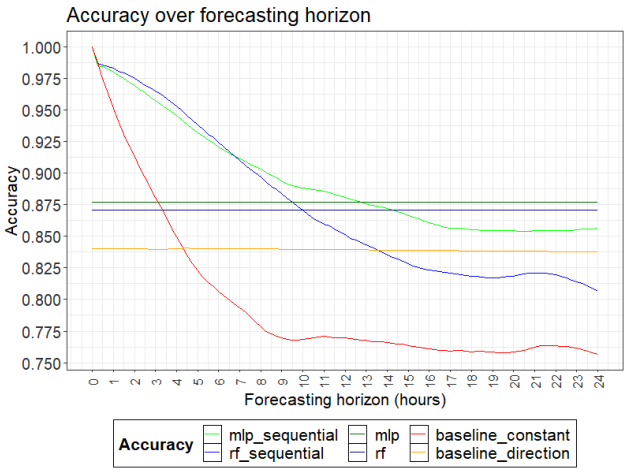


Figure 11 Accuracy as a function of the forecast horizon for the runway configuration prediction models

4.3. Explained Predictions

The trained models were used to generate explainable predictions using LIME with TAF meteorological data for a 24-hour prediction horizon. An interactive tool was created to display runway configuration and arrival/departure capacity predictions and explanations hourly based on TAF forecasts, as shown in Figure 12.

For a specific instance (2023-06-27 06:00:00) characterized by Low Instrument Flight Rules (LIFR) weather condition (Figure 14), models' explanations were obtained and are shown in Figure 13. We observed consistency in the models' explanations for the analyzed prediction instance, as they captured the influence of low visibility and low ceiling on the reduced departure capacity forecast.

5. CONCLUSIONS

This work combined Explainable Artificial Intelligence (XAI) and machine learning techniques to develop explainable predictive models capable of generating interpretable forecasts

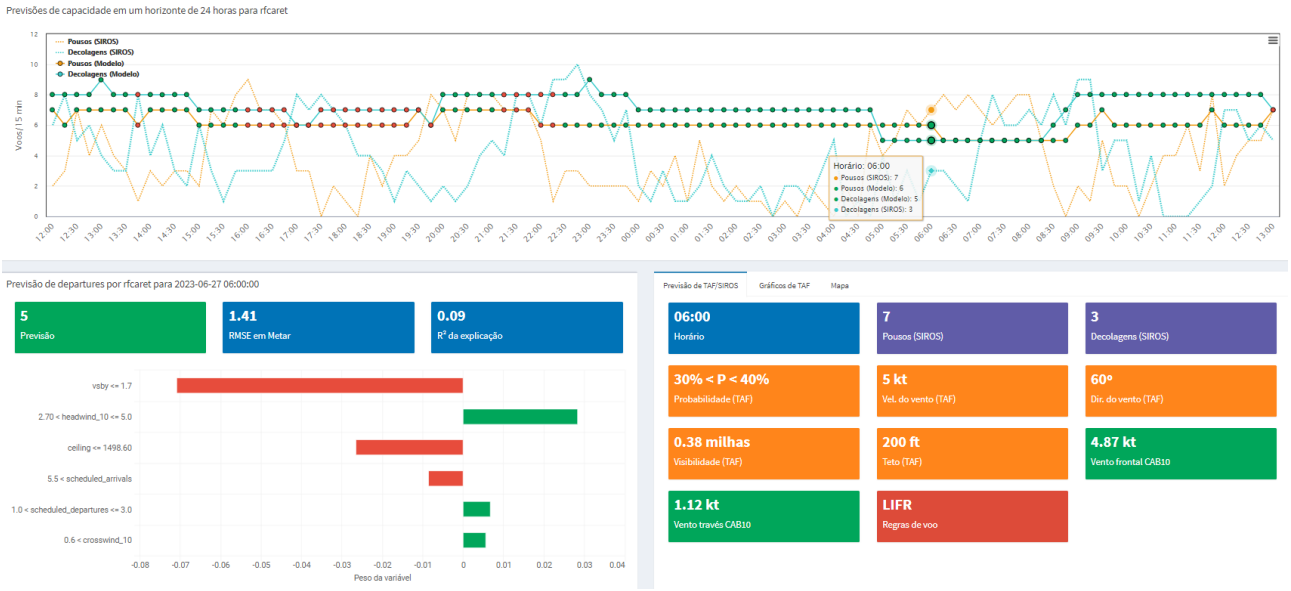


Figure 12 Interactive tool with explained runway configuration and capacity predictions for a 24-hour forecasting horizon.



Figure 13 LIME explanations for MLP and RF departure capacity models for a specific prediction instance.

of airport runway configuration and capacity towards trustworthy automated decision support in air traffic flow management. Based on historical demand, weather and traffic data for Sao Paulo/Guarulhos International Airport, the predictive models were learned using the Multilayer Perceptron (MLP) and Random Forests (RF) algorithms. A Local Interpretable Model-agnostic Explanations (LIME) approach was then used to generate forecast explanations.

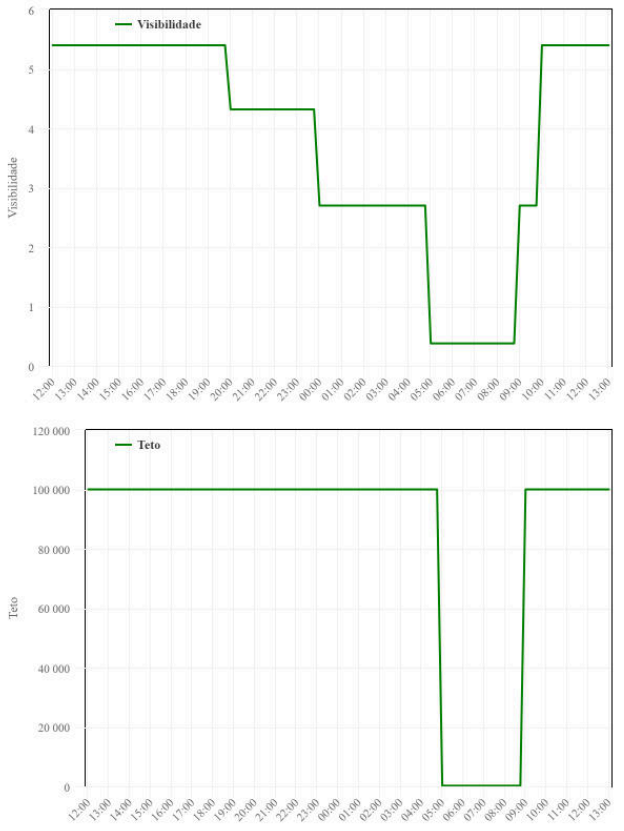


Figure 14 TAF visibility and ceiling forecast for a specific prediction instance.

The learned models showed good predictive performance, with the MLP models delivering runway configuration predictions with an accuracy of 87.73% and arrival and departure capacity predictions with a RMSE of 1.113 and 1.309, respectively. Moreover, a significant improvement in predictive performance was observed when the learned models were compared with baseline models, especially for longer forecast horizons. The LIME framework proved to be useful in quickly determining the variables that most influenced the model predictions, and the main explanations were found to be consistent across the different models learned.

For future work, potential predictive performance improvements might be explored with the use of other machine learning algorithms suitable for sequence data, such as Recurrent Neural Networks (RNN). Different surrogate models might be considered to generate local explanations with LIME in order to identify the one that best fits locally the predictions. Finally, the methodology presented in this research can be extended to other applications to develop interpretable predictive models that can be integrated into novel trustworthy decision support tools in aviation.

References

Avery, J. and Balakrishnan, H. (2015). Predicting Airport Runway Configuration: A Discrete-Choice Modeling Approach. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar*, pages 1–11, Lisbon, Portugal.

Choi, S. and Kim, Y. J. (2021). Artificial neural network models for airport capacity prediction. *Journal of Air Transport Management*, 97:102146.

DECEA (2015). Portaria DECEA n° 78/DGCEA, de 23 de março de 2015. Aprova a reedição do MCA 100-14, Manual que trata da 'Capacidade do Sistema de Pistas'. Boletim do Comando da Aeronáutica.

Degas, A., Islam, M. R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M. U., Begum, S., Rahman, M. A., Bonelli, S., Cartocci, G., Flumeri, G. D., Borghini, G., Babiloni, F., and Aricò, P. (2022). A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory. *Applied Sciences*, 12(3).

Jiao, Q.-Y., Li, N., Zheng, Z.-G., Feng, Z., Ren, G.-S., and Qiang, Y.-G. (2023). Attention Mechanism-Based Deep Learning Method for Predicting Airport Acceptance Rate: A Case Study of Hong Kong Airport. *Journal of Aeronautics, Astronautics and Aviation*, 55(2):185–200.

Rebollo, J., Khater, S., and Coupe, W. J. (2021). A Recursive Multi-step Machine Learning Approach for Airport Configuration Prediction. In *AIAA AVIATION 2021 FORUM*, page 2406.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA. Association for Computing Machinery.

Wang, Y. and Zhang, Y. (2021). Prediction of runway configurations and airport acceptance rates for multi-airport system using gridded weather forecast. *Transportation Research Part C: Emerging Technologies*, 125:103049.

Woo, C. J., Goh, S. K., Alam, S., Ferdaus, M. M., and Ellejmi, M. (2022). A runway exit prediction model with visually explainable machine decisions. In *2022 International Conference on Research in Air Transportation (ICRAT 2022)*, pages 1–9.