Advanced Engineering Informatics 64 (2025) 103067

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Multi-aircraft attention-based model for perceptive arrival transit time prediction

Chris H.C. Nguyen [a], Rhea P. Liem [a,b,*]

[a] The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong Special Administrative Region
[b] Imperial College London, London, United Kingdom

ARTICLE INFO

Keywords:
Attention mechanism
Arrival transit time prediction
Feature importance

ABSTRACT

The states and trajectories of other aircraft are crucial in predicting arrival transit time; yet, current research predominantly concentrates on individual aircraft prediction and inadequately considers other aircraft within the airspace. The oversimplification of existing models raises concerns regarding their relevance and real-time applicability. Indeed, to effectively assist decision-making processes in air traffic management, we need solutions that are accurate, computationally efficient, and consistent with air traffic controller operations. To this end, we leverage the attention mechanism—which has demonstrated success in natural language processing—to appropriately consider all aircraft in the airspace in deriving a perceptive multi-aircraft transit time prediction. To achieve this, we propose a modified attention layer that can realistically mimic aircraft's paying attention to others in a dynamic environment. The introduced model demonstrates a notable reduction in absolute prediction error by approximately 25% compared to state-of-the-art approaches. The functionality and effectiveness of the proposed attention layer are rigorously validated through extensive evaluation during the model's learning process. Additionally, we introduce a model detachment technique in the feature importance analysis to determine the features that influence the attention decision of one flight with respect to another. The promising results highlight the potential of employing the customized attention mechanism in multi-agent systems both within and beyond air transportation research.

1. Introduction

The aviation industry has been recognized as a critical component of the global economy that contributes to economic growth by connecting people and businesses throughout the world. According to a report by the Air Transport Action Group (ATAG) [1], air transport is projected to support a significant number of jobs, estimated at 88.7 million, and contribute a substantial amount of $3.5 trillion to the global economy by 2024. The increasing demand for air travel has markedly boosted the aviation industry, which is expected to persist in the foreseeable future. However, this trend comes with the issue of congestion, particularly in busy airspaces and airports, which can cause delays in flight operations. In Europe, EUROCONTROL reported that the summer of 2024 experienced a 5% surge in air traffic compared to 2023, causing significant delays, with over a third of arrivals being postponed by more than 15 minutes [2].

In this context, an accurate prediction of estimated time of arrival (ETA), along with improvement of infrastructure and technology application, is critical to mitigate congestion and ensure system efficiency in the aviation industry, as suggested by Hamzawi [3] and supported by Wang et al. [4]. Accurate, real-time ETA prediction enables air traffic control (ATC) officers to manage the flow of air traffic, reduce delays, and ensure smooth air traffic management (ATM) [5]. Moreover, ETA prediction facilitates more efficient planning and management of airline operations, which in turn helps reduce costs due to delays and missed connections; these benefits will also extend to airlines' customers. Additionally, ETA prediction can be leveraged to reduce fuel consumption and emissions—thereby promoting sustainability—as demonstrated in a study by Zhang et al. [6].

Traditional ETA prediction methods in aviation are mostly based on deterministic approaches that utilize aircraft performance and trajectory models [7–12]. However, these methods do not account for external factors such as weather, airspace sector densities, and airport congestion, which can directly affect the aircraft's actual flight profile [13]. To address this issue, recent research has delved into machine learning (ML) approaches to enhance ETA prediction in aviation. Among them, the use of the *attention mechanism* has emerged as a promising strategy, having demonstrated success in natural language processing (NLP) tasks since its introduction by Bahdanau et al. [14].

* Corresponding author at: Imperial College London, London, United Kingdom.
E-mail addresses: hcnguyenaa@connect.ust.hk (C.H.C. Nguyen), r.liem@imperial.ac.uk (R.P. Liem).

https://doi.org/10.1016/j.aei.2024.103067
Received 22 April 2024; Received in revised form 5 December 2024; Accepted 15 December 2024
Available online 31 December 2024
1474-0346/

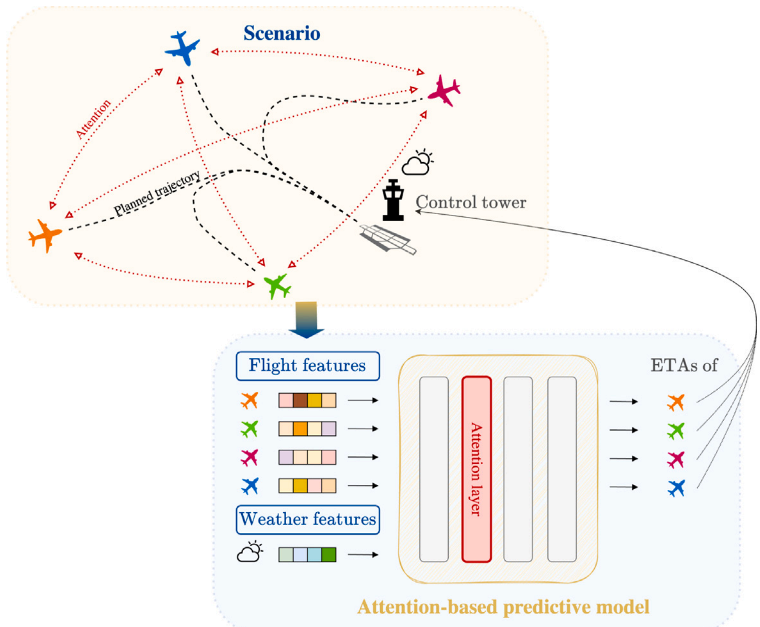


Fig. 1. An illustrative diagram for the paper's core concept. The resulting ETAs could aid ATCOs in organizing the landing sequence; however, this particular aspect is beyond the scope of our paper.

The attention mechanism enables models to selectively focus on specific parts of input data while ignoring others, thus facilitating the extraction of more meaningful features and improving prediction accuracy. As such, attention mechanisms hold immense potential to capture the interdependence between different aircraft and thus improve ETA prediction accuracy. To date, studies in the field have either focused on ETA prediction for individual aircraft without employing attention mechanism [4–6,13,15–19], or applied attention mechanisms to delay prediction at a network-wide level without accounting for individual flights [20,21]. That said, the application of attention mechanism to account for interactions between aircraft in ETA prediction is still an open research area.

Closing this research gap is imperative to substantially improve ATM through more accurate predictions. Therefore, the present study aims to further harness the potential of the attention mechanism to predict individual aircraft's remaining transit time while taking into account other aircraft's characteristics and trajectories in the vicinity, as illustrated in Fig. 1. In this figure, a scenario represents the state of the airspace at a specific timestamp, where each aircraft has its corresponding planned trajectory (as shown by black dashed lines within the yellow box). Based on this scenario, relevant information regarding arrival flights (except for their planned trajectories) and airport weather become inputs to the ETA prediction model, which is illustrated within the blue box. One of the main contributions of this work is the multi-aircraft attention layer in the predictive model, which is derived to mimic the attention interactions among arrival flights (as illustrated by the red dotted double-headed arrows within the yellow box). This additional attention layer can enhance the ETA prediction result accuracy by better capturing the dynamics of the airspace. These predicted ETAs can then be provided to air traffic control officers (ATCOs) and help determine arrival sequence, though this part is beyond the scope of the present paper.

Here, the attention concept is specifically implemented for the ETA prediction task, where the nature of attention differs from how it is commonly modeled in NLP tasks. In language processing, for instance, each word typically has a fixed total attention weight that is distributed towards others. In ETA prediction problems, on the other hand, the total attention weight given by one aircraft cannot be fixed since it depends on its states and traffic conditions. In particular, each aircraft may give full attention to several aircraft at once, and the number of aircraft receiving attention varies depending on its position in the traffic sequence. As such, we need to derive a new attention mechanism that is suitable for the specific problem at hand, which will be elaborated in Section 3.5. The proposed approach is evaluated with actual flight data and compared against commonly used ETA prediction methods. The results demonstrate that the attention mechanism notably improves ETA prediction accuracy, particularly in congested and complex airspaces such as the Hong Kong International Airport (HKIA, ICAO: VHHH). Besides, the paper compares the performance of the proposed model using trajectory-based and position-based flight information. Our findings indicate that the attention-based model's performance with position-based data surpasses that of other models in terms of ETA prediction accuracy. Finally, our results also showcase the effectiveness of the newly developed attention layer when used in a multi-agent complex system.

In our model derivation, the key output is remaining transit time, which estimates the time left to complete the flight. This information can then be used to determine ETA. Hence, the terms ETA and remaining transit time prediction will be referred to in the discussion, sometimes interchangeably, as they are interrelated.

To structure the paper, the following sections are presented. In Section 2, we provide a review of prior research on ETA prediction and the application of attention mechanisms in the field of air transportation. Next, Section 3 offers a thorough explanation of our proposed methodology. The case study used for our model implementation is described in Section 4. In Section 5, we present our results and their comparison against those obtained using common approaches in the field. We also examine the role of input features in the attention mechanism and the impact of attention on the model. We close this section by discussing the model applicability for real-time deployment in air traffic management. Lastly, in Section 6, we summarize our methodology and conclude its contributions.

2. Overview of the current state-of-the-art

Extensive efforts have been invested in predicting ETA for different modes of transportation besides aviation, including watercraft [22] and land vehicles operating on roads [23,24] and rails [25,26]. Similar

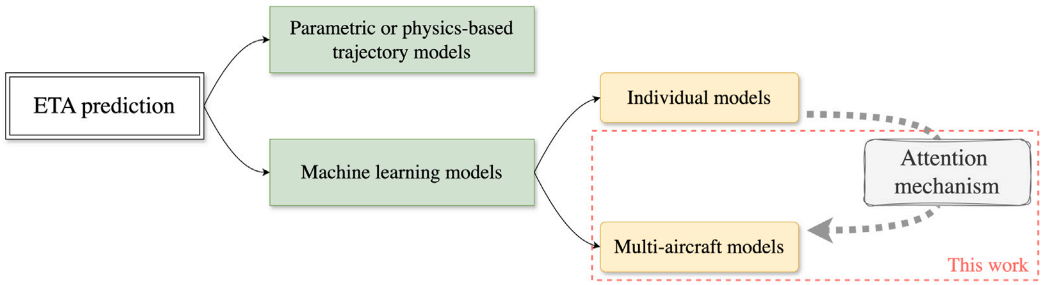


Fig. 2. Overview of ETA prediction methods in the air transportation sector. In the present work, the attention mechanism serves as the link transforming individual prediction efforts into multi-aircraft prediction models.

to other transportation modes, methods for predicting ETA pertaining to aircraft operations can be categorized into two primary groups, as depicted in Fig. 2.

Traditional approaches for ETA prediction in aviation comprise parametric and physics-based trajectory models [7–12]. These methods typically involve computing the flight trajectory (which includes the lateral flight path, altitude, and speed profiles) and subsequently deriving the time required to complete the predicted trajectory. However, the accuracy of these predictions can be affected by uncertainties in atmospheric conditions, airspace, and airport traffic, which may lead to inaccurate ETA predictions at the target airport. Few studies in this category take into account weather conditions and the presence of concurrent arrival aircraft in their ETA predictions; one particular example is the work by Porretta et al. [11]. Their research developed conflict-free trajectories by utilizing predetermined flight intents. Besides, they operated under the assumption of static wind, obtainable through a look-up table. However, it is the generation of flight intents and the interpolation of wind data that impose significant processing demands. Furthermore, in congested airspace during peak traffic hours, forecasting ETAs for individual flights leads to a proportional increase in computational workload with the number of aircraft, making conventional methods less appealing for real-time applications.

These limitations have led to the development of new technologies, such as ML and artificial intelligence, which can overcome some of the drawbacks of traditional ETA prediction techniques by leveraging real-time data and incorporating complex factors that affect ETA. Recent studies in the aviation industry have explored the use of ML techniques to improve ETA prediction accuracy. ML algorithms have the potential to identify patterns and relationships in large and complex datasets, which may be difficult for traditional statistical models to capture accurately [27]. Glina et al. [5] were among the first researchers to apply ML, in particular by using quantile regression forest [28], to forecast ETA for individual flights at Dallas/Fort Worth International Airport. Kern et al. [15] subsequently integrated meteorological conditions and air traffic data into their ETA prediction model by utilizing tree-based methodologies. Through meticulous parameter tuning and selective feature incorporation, their model successfully reduced the margin of error distribution in comparison to the standard, commonly used Enhanced Traffic Management System (ETMS). Meanwhile, Ayhan et al. [13] included an even broader range of features, spanning from flight and airport data to atmospheric and meteorological variables, utilizing a variety of ML algorithms for comparative analysis. Nevertheless, their findings generated a contentious inquiry, as the atmospheric features were ranked as the second to sixth most significant factors. Dhief et al. [17] employed a similar methodology to that of Ayhan et al. [13] and obtained more comprehensive results. Wang et al. [4], on the other hand, utilized the stacking technique, which involved combining multiple ensemble models to improve the accuracy of the prediction results. Conversely, Wang et al. [16] contended that constructing distinct prediction models for various trajectory clusters would result in superior accuracy. To achieve this, they utilized the density-based spatial clustering of applications with noise (DBSCAN) algorithm (which was first developed by Ester et al. [29]) for clustering prior to employing a Feed-forward Neural Network (FNN) [30–32] for prediction. Nevertheless, it is important to acknowledge that dividing the data based on clustering may not always be optimal due to several reasons. Firstly, some clusters may still cover a broad spatial area, and traffic patterns may change over time, necessitating modifications to the clustering process. Secondly, clustering can be computationally expensive, which can be a limiting factor in the practicality of this approach. Finally, clustering can introduce errors, which may affect the accuracy of the prediction models. Gui et al. [18] also identified clusters to enhance prediction accuracy by utilizing a novel representation of heading in their clustering technique. However, it is worth mentioning that the clustering process employed therein was tailored to the specific characteristics of Guangzhou International Airport and might not be easily transferable to other airports, which could potentially limit its practicality. The work by Deng et al. [33] offered more advantages, as they adopted Long Short-Term Memory (LSTM) [34] for cluster detection, which was faster than the clustering approaches used by Wang et al. [16] and Gui et al. [18]. However, it should be noted that the prediction results were not as accurate as expected, as the focus of the study was primarily on real-time cluster detection rather than prediction performance. Most recently, Zhang et al. [6] utilized arrival pressure and sequencing pressure, along with a vector of wind magnitude at different altitudes, to capture the dynamic state of the complex but organized terminal airspace. This approach offers significant advantages in improving ETA prediction accuracy. Additionally, Jun et al. [19] employed the Catboost algorithm [35] for a two-step prediction process to evaluate holding probability and delay time. They further extended their approach to optimize the speed profile based on the prediction results to minimize fuel consumption, underscoring the significance of ETA prediction in optimizing air traffic flow.

There is a noticeable trend in the air traffic ETA prediction field towards capturing the dynamic state of the terminal airspace to account for the highly interdependent nature of aircraft operations within Terminal Maneuvering Area (TMA). Initially, the dynamic state was approximated using simple metrics such as the number of aircraft in the airspace, the airport acceptance rate in the previous hour, or even the timestamp [4,13,17]. Later, it was further refined by grouping aircraft together in clusters or Standard Terminal Arrivals (STARs) [16,18]. More recently, the dynamic state was expressed through arrival pressure and sequencing pressure [6]. While the approach represents an improvement over earlier static methods, it still relies on simplified representations of the arrival sequence. This calls for a proper inclusion of *attention* in modeling aircraft interaction to efficiently manage arrival sequences, which will be introduced in this work.

The *attention mechanism*, which was initially proposed by Bahdanau et al. [14], provides a means of adding another layer of information in modeling by paying attention to certain features or entities during the dynamic interaction. It was first applied in the realm of neural machine translation, where it was leveraged to enhance the performance of sequence-to-sequence (often abbreviated to seq2seq) models.

By allowing the model to concentrate on the most important aspects of the input sequence at each time step, the attention mechanism has resulted in substantial advancements in the accuracy of tasks such as machine translation [14,36,37], speech recognition [38,39], and image captioning [40,41]. *De facto*, the attention mechanism is a fundamental component of the architecture of the Generative Pre-trained Transformer [42] series of language models, including the increasingly more popular ChatGPT.

In the field of aviation, there have been several efforts to leverage the attention mechanism to address the inherent interdependence among air vehicles or relevant entities. Yu [43] combined attention mechanism with Gated Recurrent Unit [44] to forecast passenger flow in the short term. Regarding trajectory prediction, Jia et al. [45] developed a model that utilizes LSTM to process the current trajectory and then applied the attention mechanism to determine which time step to focus on. Recently, Cai et al. [21] initiated a project that employed Graph Convolutional Networks [46] to predict flight delays for the Chinese airport network. The work also utilized the attention mechanism to aggregate features from the neighboring nodes. However, the results were still limited since the authors only selected four main airports for the analysis. In contrast, Bao et al. [20] developed a graph-to-sequence model based on the seq2seq architecture [14] for the USA airport network. The model showed promising results for type I airports, which serve more than 500 flights per day. However, further research is needed to improve the model's performance at other airports.

Despite its promising potential, the attention mechanism has not been exploited in air traffic ETA prediction, to the best of our knowledge. This realization has motivated us to conduct this research that employs and modifies the attention mechanism to suit the context of arrival air traffic. The proposed approach is expected to provide fast and accurate ETA predictions in real-time applications, thereby enhancing the ATM efficiency.

In summary, this paper seeks to bridge two main research gaps in ETA prediction. Firstly, it addresses the absence of a suitable method to incorporate the dynamic state of terminal airspace. Drawing inspiration from NLP, this challenge is tackled through an attention mechanism. However, it is crucial to note that the application of attention in the aviation domain differs from its use in NLP, highlighting the second key research gap. Consequently, a tailored attention mechanism will be proposed to account for the distinction.

3. Methodology

In this section, we outline the methodology employed to develop an ML-based model to derive arrival transit time prediction in the terminal airspace that is apposite for real-time applications. After describing the problem and providing an overview of the model, different types of neural networks are presented, including an attention mechanism with modifications. Lastly, we present the metrics used to evaluate and compare our model performances.

3.1. Problem statement

Our prediction model focuses on aircraft in TMA and surrounding airspace. Fig. 3 depicts a complete aircraft operation, where an aircraft flies through the TMA at the origin airport during departure and continues to the en-route airspace before entering the TMA at the destination for arrival. In essence, the en-route airspace refers to the airspace between TMAs and is primarily designed to handle aircraft during the cruise phase of their flights. On the other hand, TMAs typically experience higher traffic density and more maneuvering requirements, which are primarily caused by the convergence of arriving aircraft from different directions. Hence, estimating the transit time within and around TMAs, where a significant portion of aircraft maneuvers take place, presents a greater challenge, which motivates us to develop our model for this specific airspace region.

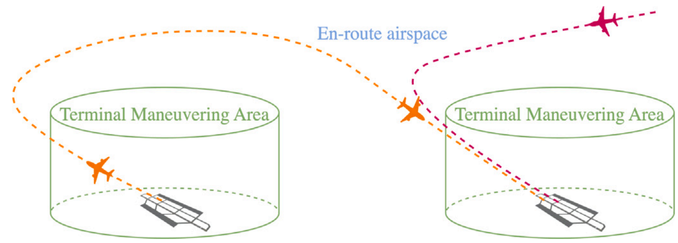


Fig. 3. Schematic depiction of TMAs and en-route airspace.

The collective airspace within and surrounding a TMA is commonly referred to as terminal airspace [47]. Throughout this paper, unless otherwise specified, we use the term "terminal airspace" to refer to the airspace within and around a TMA.

We consider a scenario where a set of n arrival aircraft $\mathbf{X}_m = \{x_{i,m} | i = 1, \ldots, n\}$ are present within the terminal airspace of the destination airport at time point m, all of which are affected by the same weather conditions. Our objective is to predict the corresponding remaining transit time for all n aircraft, $\mathbf{Y}_m = \{y_{i,m} | i = 1, \ldots, n\}$, which measures the duration between the current time and the arrival time at the destination airport.

Specifically, at a particular timestamp m, predictions are simultaneously made for all aircraft in $\mathbf{X}_m$. These predictions rely on inputs that consist of both the information regarding all aircraft within the designated airspace and the corresponding weather conditions at the airport at the associated timestamp. The detailed breakdown of these inputs will be discussed in Section 4.2.

3.2. Model overview

The proposed remaining transit time prediction model is illustrated in Fig. 4, featuring four main components, namely a flight embedding network $\mathcal{F}$ (in light yellow), a weather embedding network $\mathcal{W}$ (in green), an attention layer $\mathcal{A}$ (in the shaded gray area), and an output network $\mathcal{O}$ (in blue). The model operates as follows: first, flight information vectors are fused into $\mathcal{F}$, resulting in flight-embedded vectors $\mathbf{f}$. These vectors are then fed into the attention layer $\mathcal{A}$ to simulate the interaction among aircraft. After this step, each aircraft receives a vector that represents the level of attention it should pay to other aircraft in the airspace. Meanwhile, $\mathcal{W}$ takes a weather information vector as input and converts it into a weather-embedded vector $\mathbf{w}$. For each flight, the attention and weather vectors, $\mathbf{a}$ and $\mathbf{w}$, are concatenated with the flight-embedded vector $\mathbf{f}$. The concatenated vector is then fused into $\mathcal{O}$ to obtain the anticipated remaining transit time.

The proposed transit time prediction model incorporates various types of neural networks, including FNN, LSTM, and an attention layer. The weather embedding and output network are both FNNs, while the flight embedding network can be either an LSTM or an FNN depending on the consideration of the flight's state, either the previous one or the current position. As previously mentioned, a modified attention layer is introduced in this work to be able to accommodate a multi-aircraft system, where the number of aircraft can vary. Further details on these neural networks will be presented in the subsequent subsections.

3.3. Feed-forward neural networks

FNN is one of the artificial neural network architectures widely used in various fields for function approximation and classification tasks. The architecture of FNNs is composed of an input layer, one or more hidden layers, and an output layer, as illustrated in Fig. 5. The input layer receives input data that are propagated forward through the hidden layers to the output layer. Each hidden layer contains a set of neurons, which are connected to the neurons in the previous and next

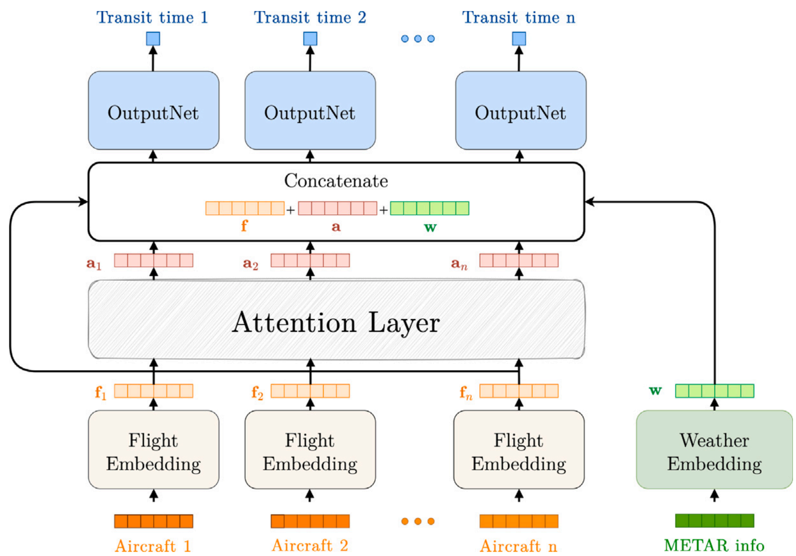


Fig. 4. The architecture of the developed predictive model. For each inference, the remaining transit time is evaluated based on flight information (for all aircraft in the airspace) and weather conditions.

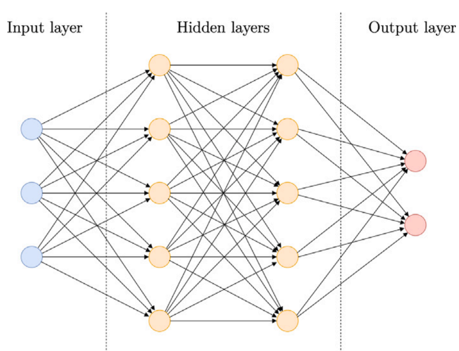


Fig. 5. An illustration of an FNN architecture. The input layer is positioned on the left, followed by two hidden layers, and the output layer is located on the right-hand side. The number of hidden layers shown here is arbitrarily selected for illustration purposes only.

layers through weighted connections. FNNs are called "feed-forward" because the output of each layer is fed forward as input to the next layer without any feedback connections. The hidden layers of FNNs are responsible for feature extraction and abstraction, while the output layer is responsible for producing the final output of the network. FNNs are trained using backpropagation, which involves iteratively adjusting the weights of the connections between neurons to minimize the error between the predicted output and the actual one.

3.4. Long short-term memory networks

LSTM is a type of recurrent neural network (RNN) that has gained popularity in recent years due to its ability to effectively capture long-term dependencies in sequential data. LSTMs are particularly useful when dealing with data that exhibit temporal dynamics, such as speech [48], text [37], and time-series data [49]. Unlike traditional RNNs, LSTMs, as shown in Fig. 6, have a memory cell that allows information to be retained or forgotten over a prolonged period, enabling them to model long-term dependencies more effectively. LSTMs consist of a set of gates that control the flow of information into and out of the memory cell, including an input gate, an output gate, and a forget gate. These gates, along with the memory cell, allow LSTMs to selectively store and retrieve information at each time step.

One of the aims of this study is to investigate whether considering both current and previous states of an aircraft can improve the accuracy of prediction results compared to using the current state alone. The hypothesis is that taking into account the past states of the aircraft would allow us to better understand its behavior, thereby producing more accurate predictions of its future states, including ETA. This research question will be addressed by comparing the performance of LSTM-based models with those based solely on the current state of the aircraft.

3.5. Attention layer

The attention mechanism operates akin to a differentiable key-value memory model [50,51]. When used in our model, the attention layer takes a set of flight embedding vectors to produce a set of output vectors that focus on impactful flights to themselves. In a way, this mechanism serves to mimic the consideration of other flights in making ATC decisions. Concretely, the attention of aircraft i towards aircraft j is computed via the query vector $\mathbf{q}_i$, the key vector $\mathbf{k}_j$, and the value vector $\mathbf{v}_j$. The query vector $\mathbf{q}_i$ from aircraft i serves as the reference used to extract pertinent details from the key-value pairs of aircraft j, where the key vector $\mathbf{k}_j$ represents encoded input data and the value vector $\mathbf{v}_j$ holds the corresponding contextual information. Hence, the computing procedure of the attention vector of aircraft i, $\mathbf{a}_i$, proceeds as follows. First, the corresponding attention weight of aircraft i towards aircraft j, $\alpha_{i,j}$, is obtained by comparing $\mathbf{q}_i$ with $\mathbf{k}_j$ through a bilinear mapping (*i.e.,* a similarity function), followed by a *softmax* operation to ensure that the attention weights sum to one:

$$\alpha_{i,j} \propto \exp\left(\frac{\mathbf{q}_i \mathbf{k}_j^\top}{\sqrt{d_K}}\right). \tag{1}$$

In Eq. (1), d_K is the dimension of both $\mathbf{q}_i$ and $\mathbf{k}_j$. Subsequently, the corresponding attention vector $\mathbf{a}_i$ is computed as a weighted sum of the value vectors, where the weights are contingent upon the attention weights:

$$\mathbf{a}_i = \sum_{j \neq i} \alpha_{i,j} \mathbf{v}_j. \tag{2}$$

In this work, we propose a new attention layer that differs from the standard version in NLP by incorporating a *tanh* operation, instead of *softmax* as in Eq. (1), to determine attention weights. This new architecture is shown in Fig. 7. The new model is derived based on our realization that constraining the sum of attention weights to

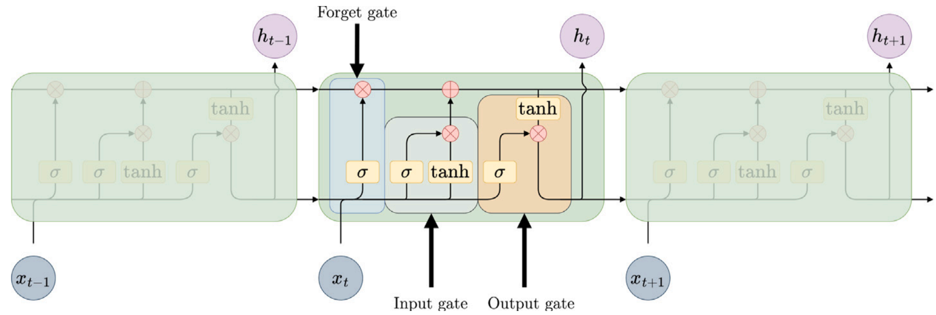


Fig. 6. An illustration of three LSTM cells, representing the time steps $t-1$, t, and $t+1$, which are fundamental components of LSTM networks. The forget gate regulates the information flow from the previous cell state, the input gate manages the information flow from the input, and the output gate modulates the information flow from the cell state to the output.

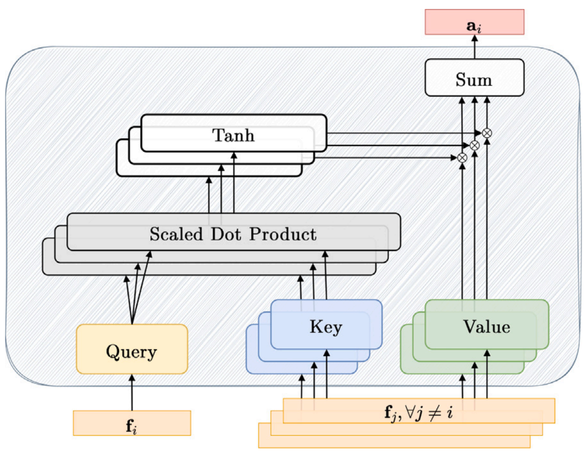


Fig. 7. Customized attention layer with *tanh*-weighted attention.

one may not accurately represent certain real-world situations. For instance, in arrival air traffic scenarios, an aircraft typically experiences longer waiting time when multiple aircraft are positioned ahead of it. Consequently, it may need to allocate attention to more than one aircraft in its immediate surroundings. Conversely, an aircraft that is about to land may not require attention towards other aircraft at all. Using *softmax*, which forces the attention weights to sum up to one, may not be suitable in these cases as it always requires aircraft to distribute their attention with a total weight of one among all other aircraft. In contrast, the *tanh* operation in the proposed attention layer computes the attention weight towards aircraft j based on the query of the focal aircraft and the key of aircraft j only. This design allows for the independent assignment of attention weights to different aircraft. With this rationale, Eq. (1) is then modified as:

$$\alpha_{i,j} = \frac{1}{2}\left[\tanh\left(\frac{\mathbf{q}_i\mathbf{k}_j^{\top}}{\sqrt{d_K}}\right)+1\right], \tag{3}$$

where an offset of one and a scaling of $\frac{1}{2}$ are applied to ensure that the attention weights fall within the desired range of zero to one. To the best of our knowledge, this derivation is new and constitutes an original contribution of the present work.

Additionally, in this study, the attention weight of an aircraft to itself is always set to zero, regardless of whether the standard or modified attention layer is used. This constraint is imposed to prevent an aircraft from excessively influencing its own prediction. By excluding self-attention, the model can effectively concentrate on capturing the relationships and dependencies between different aircraft in the airspace.

3.6. Evaluation metrics

In this study, we use mean absolute error (MAE) and root mean squared error (RMSE) as our primary training and evaluation metrics. MAE measures the average magnitude of the errors in our predictions, while RMSE measures the square root of the average squared error. Specifically, they are calculated as follows:

$$MAE = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{|\mathbf{X}_m|}\sum_{i\in\mathbf{X}_m}|y_{i,m}-\hat{y}_{i,m}|\right), \tag{4}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{|\mathbf{X}_m|}\sum_{i\in\mathbf{X}_m}\left(y_{i,m}-\hat{y}_{i,m}\right)^2\right)}, \tag{5}$$

where $y_{i,m}$ and $\hat{y}_{i,m}$ respectively indicate the ground truth and estimated remaining transit time of aircraft i in scenario m, which has the set of aircraft $\mathbf{X}_m$ in the airspace. Meanwhile, M denotes the number of scenarios corresponding to timestamps in our dataset.

4. HKIA case study

This section presents a detailed case study that specifically examines the performance of the proposed model at HKIA. First, we introduce the airspaces under consideration, followed by the feature selection for the prediction model. Lastly, several baselines are established for comparative analysis purposes.

4.1. Region of interest

As highlighted in our previous study [52], TMA represents a critical bottleneck in managing air traffic flow. Building upon this understanding, we apply our prediction methodology to the TMA within 50 NM around the airport, as visually illustrated by the green region in Fig. 8. This region is highly organized, with the arrival sequence being almost entirely determined by ATCOs. Yet, traffic patterns within this airspace are complex, which calls for a special attention in predicting the transit time accurately.

To further evaluate the performance of our method in a more comprehensive context encompassing a larger number of aircraft and diverse airspace characteristics, we expand our study area to cover a 100 NM radius around the airport. This expanded area is represented by the blue region in Fig. 8, which excludes the green region described above. The primary objective of this expansion is to study the applicability of our method in different regions with different shapes and sizes, without being restricted to the definition of TMA. Within this extended airspace, the distribution of aircraft is less dense, resulting in less complicated traffic patterns. Historical data analysis reveals that flight

Fig. 8. In-study airspace with STARs. The green region represents the control area bordered by a circle of 50 NM and the border with Mainland China. The blue region depicts a 100 NM area around the airport, excluding the green region, to investigate the model performance beyond the strict definition of TMA.

trajectories within this region primarily consist of straight routes, occasionally incorporating holding patterns. While this expansion naturally aligns with the concept of enlarging the time control perimeters beyond the terminal area, or Extended Metering,[1] the specific discussion on this application and its implication is beyond the scope of the present paper.

TMA is structured by STARs, as depicted in Fig. 8. Each STAR connects an entry waypoint to an Initial Approach Fix (IAF). In the case of HKIA, there are four entry waypoints, namely ABBEY, BETTY, CANTO, and SIERA. At the time when the flight data used in this study were collected, HKIA operated two parallel runways, one for takeoff and the other for landing. These runways correspond to two landing directions, namely 07 and 25, which are associated with IAFs TD and LIMES, respectively. Due to the proximity of the entry waypoint SIERA to the airport, there are two STARs originating from SIERA for each landing direction. This is necessary to allow ATCOs to have flexibility in maneuvering aircraft in the event of congestion. Hence, there are a total of ten STARs established in the studied TMA.

4.2. Model features

To effectively train and validate our proposed model, it is essential to carefully select pertinent flight and weather features that influence the duration of aircraft landing. This subsection discusses what these features are, along with their respective data sources. Table 1 summarizes the features used in our model derivation, which include flight and weather features. They are described in details below.

4.2.1. Flight features

The landing time of an aircraft is determined by a confluence of factors, including its position and dynamic features such as longitude, latitude, altitude, groundspeed, and vertical speed. Depending on whether a trajectory-based or position-based model is employed, the previous position and dynamic states can be fused into the prediction model. Additionally, the remaining transit time is also affected by the aircraft's category (*i.e.*, jumbo, heavy, medium, or light), the flight range category (*i.e.*, long, medium, or short haul), and the holding time (which is known in advance to pilots, as will be further described below). Furthermore, the landing direction is also another factor that cannot be overlooked in determining the remaining transit time. Using these features (as listed in Table 1) makes this prediction model suitable for real-time applications since they are promptly accessible to ATCOs at the time of prediction. While some induced features (such as arrival trajectory cluster and sequencing pressure) have been proven to be useful [4,6,16], our study utilizes only readily available features without manipulation. The main purpose is to demonstrate the underlying capability of the model itself.

In the context of TMA operation, a holding pattern is defined as a predesignated flight path that an aircraft follows under air traffic control instructions, typically on the onset of congestion near the airport. Holding patterns are typically set up in a circular or racetrack-shaped configuration around specific waypoints located near the boundary of TMA. Specifically for the terminal operations in HKIA, pilots are typically informed of the duration of holding patterns (when necessary) by ATCOs approximately five minutes prior to entering TMA. This information is calculated by the arrival manager (AMAN). Considering the range of aircraft speed before they reach the TMA (ranging from 350 to 550 knots), a time span of five minutes corresponds approximately to a distance of 25 to 40 NM. The assumption that holding time is known in advance will be further discussed in Section 5.3.1.

The process of extracting holding time begins with detecting loops in the two-dimensional trajectory of latitude and longitude of the aircraft, following an in-house procedure previously developed by Lui et al. [53]. Once a holding pattern is detected, we interpolate the time of entry and exit from the pattern. With this information at hand, it becomes relatively straightforward to determine, at a specific

[1] https://www.faa.gov/air_traffic/publications/atpubs/foa_html/chap18_section_25.html (last accessed on December 04, 2024).

Table 1
Input features of the ETA prediction networks.

Category	Feature	Description
Position/Trajectory	Latitude	Relative latitude to the airport
	Longitude	Relative longitude to the airport
	Altitude	Altitude of the aircraft in feet
	HA Sine	Sine of the heading angle
	HA Cosine	Cosine of the heading angle
	Ground speed	Ground speed in knots
	Vertical speed	Vertical speed in feet/second
	Holding	Whether the aircraft is in holding
	Holding time	Time duration (in seconds) that the aircraft is expected to be in holding
	Go around	Whether the aircraft is attempting to land again
Specification	Aircraft category	Jumbo, heavy, medium, or light
	Range category	Long, medium, or short haul
	Runway direction	Which direction the aircraft is assigned to land (07 or 25)
Weather	Wind speed	Wind speed in knots
	Wind sine	Sine of the wind direction
	Wind cosine	Cosine of the wind direction
	Visibility	Visibility in miles
	Wind gust	Wind gusts in knots
	Sky coverage	The maximum sky coverage of the four observed levels
	WX Code	Present weather code

timestamp, whether an aircraft is in a holding pattern and to estimate the remaining time within it.

In our study, flight trajectory data are sourced from Flightradar24,[2] which employs automatic dependent surveillance-broadcast (ADS-B) technology to collect data. The data, with a time granularity of one minute, are obtained for the period spanning from April 20, 2018 to July 04, 2018, encompassing 36,968 arrival flights. Given the prediction model's purpose, the information of all flights in the airspace at each timestamp available in the historical dataset is treated as a single data point, which will be complemented with weather features in the following subsection. However, due to missing timestamps during data collection, we obtain only 91,881 data points for the aforementioned period, which are deemed sufficient for the specific problem at hand.

For a more detailed data description, Fig. 9 displays flight trajectories to HKIA over a single day, where the complexity of the traffic flow pattern is evident. Despite the intended purpose of STARs as the designated approach routes for aircraft, a substantial number of flight trajectories deviate from these predefined paths due to various maneuvering requirements. Factors such as congestion can result in longer trajectories compared to their corresponding STARs. In such scenarios, holding patterns are frequently utilized at HKIA to manage delay. Trajectories with holding patterns account for approximately 27% of the total flights in our dataset. The total holding durations of such flights are summarized in the histogram shown in Fig. 10. It is important to note that a holding pattern typically requires six minutes to complete one round, and an aircraft may be instructed to enter holding patterns multiple times. In other cases, an aircraft may experience a shorter trajectory if it is prioritized to proceed straight to the airport. The extensive use of holding patterns and other maneuvers to handle the vast volume of traffic adds complexity to predicting ETA for aircraft.

Fig. 11 illustrates the mean and standard deviation (SD) of the remaining transit time within the blue and green airspace around the airport (referring to Fig. 8). This visualization is carried out on the test dataset. To generate this figure, we capture the remaining transit time of each aircraft at all available timestamps in the relevant dataset, correlated with the respective latitude and longitude coordinates. The airspace is then discretized into small grid cells, and we calculate the mean and SD of the remaining transit time within each cell. Finally, a Gaussian interpolation method is applied for a smoother depiction of the data.

[2] https://www.flightradar24.com/ (last accessed on December 04, 2024).

Upon examining the figure displaying the mean remaining transit time on the left-hand side of Fig. 11, it becomes apparent that aircraft within TMA typically require considerably less time compared to those outside the TMA boundaries (as signified by the lighter color). Nonetheless, an interesting anomaly occurs in the southwest part of the TMA, where a concentrated cluster of dark red dots is observed, indicating the high mean transit time associated with some aircraft. This anomaly is attributed to the proximity of the entry waypoint SIERA, located on the boundary of Hong Kong airspace. Aircraft entering the TMA via this waypoint need to perform most of their maneuvers within the TMA airspace. Another noteworthy observation is the racetrack shapes with darker colors observed at the TMA boundary, indicating the presence of holding patterns. Beyond this boundary, there is a notable variation in colors, signifying the varied remaining transit times of aircraft even in close proximity.

The right-hand side figure of Fig. 11 strongly reinforces the earlier observation, even at a cell-level analysis. Darker colors in this representation indicate a higher degree of variation in the remaining transit times within each cell. A consistent trend with the mean figure is discernible. In particular, there is a sharp change in SD at the TMA boundary, where holding patterns can be seen more clearly. Plus, notable variance can be observed even in mere cells. In other words, factors beyond horizontal spatial location exert a considerable influence on the remaining transit times of aircraft.

4.2.2. Weather features

Weather has a significant impact on air traffic operations, and many studies [6,13,15,17,19] have incorporated various weather features into their models to capture and evaluate this influence. However, including too many weather features can make the model more computationally demanding and prone to overfitting. In our study, we focus on demonstrating the advantages of using attention mechanisms in transit time prediction rather than investigating the direct impact of weather. Therefore, we specifically consider weather conditions around the airport, as they largely influence the arrival slot and consequently the arrival time.

Besides, the range of dominant weather conditions at each airport can vary significantly. For instance, comparing the weather conditions at HKIA and London Heathrow International Airport provides a notable contrast, as presented by Lui et al. [54]. HKIA experiences frequent rainfall and thunderstorms due to its humid subtropical climate, while Heathrow encounters freezing conditions and visibility challenges. To derive suitable weather condition representations for a particular airspace, researchers may choose to select a specific set of

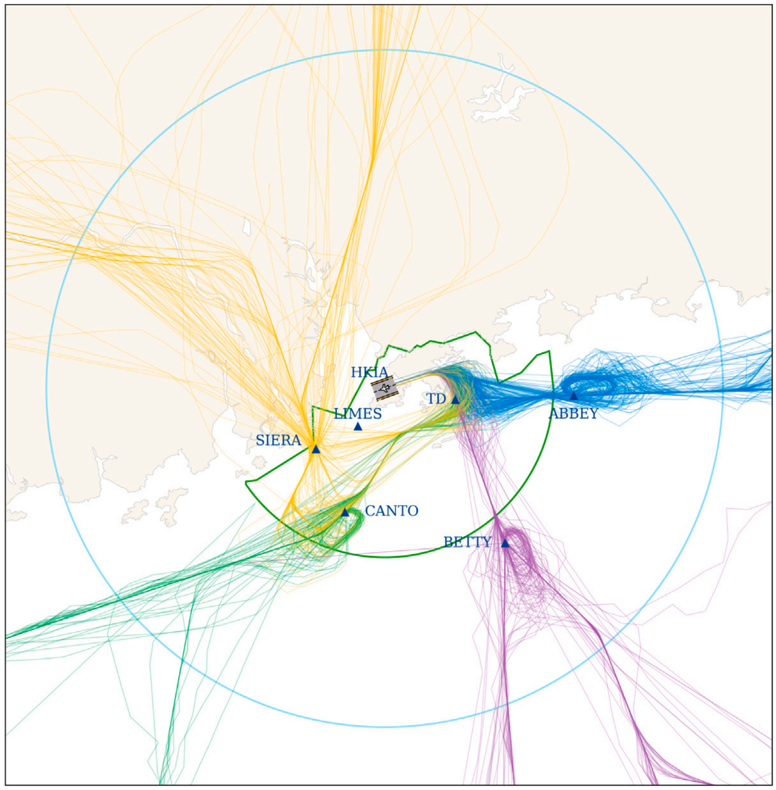


Fig. 9. Actual trajectories on 2018-05-27 when the landing direction 25 (at HKIA) were used. They are clustered and color-coded based on their respective entry waypoints.

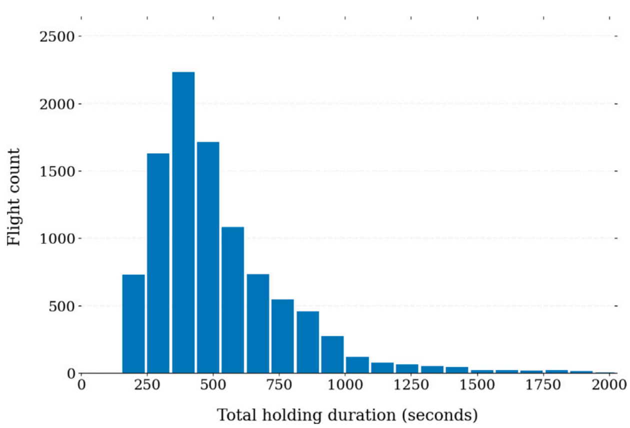


Fig. 10. Histogram of holding duration in the dataset.

weather features to incorporate into their models. The selected set of weather features in this study is summarized in Table 1.

To obtain the necessary weather data, we utilize the Meteorological Terminal Air Report (METAR), a weather report issued by airports around the world to provide their current weather conditions. The METAR data for HKIA are obtained from Iowa State University,[3] which provides free access to historical weather data.

4.2.3. Feature scaling

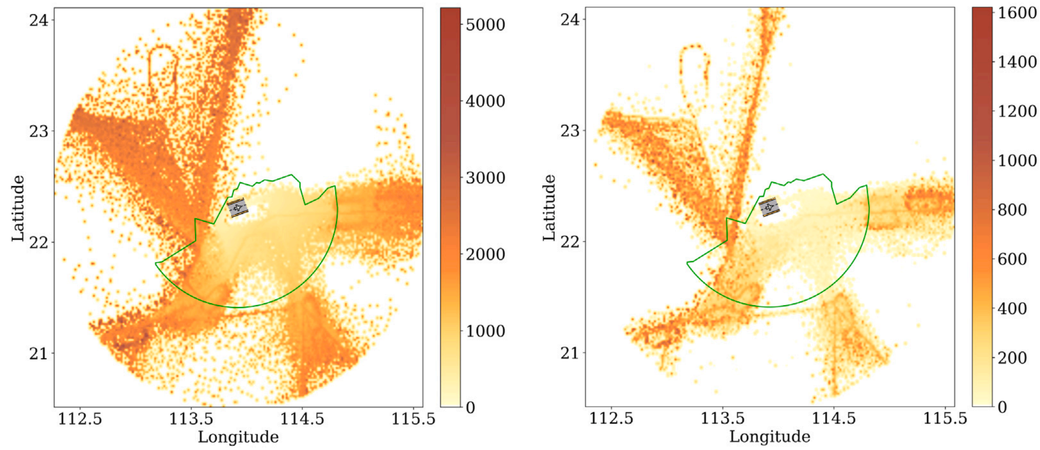
Feature scaling is a critical preprocessing step in ML that involves normalizing or scaling data features to ensure that they are on a similar scale to prevent the algorithm putting more emphases on some features in the learning process due to their larger magnitudes. As such, feature scaling can help improve the accuracy and efficiency of the ML algorithm, making it easier to train and providing more accurate predictions.

Considering the benefits, this preprocessing step is applied to the features used in our model derivation. Note that categorical features and features that exhibit values around zero (such as relative latitude, relative longitude, and other features represented by sine and cosine functions) are excluded in the feature scaling process. In particular, one hot encoding is employed to represent categorical features, which include holding pattern, aircraft category, range category, runway direction, sky coverage, and weather code (details of the categorical options of these features can be found in Table 1). The remaining features undergo scaling using either the max-min normalization method or the variance scaling method, contingent upon the data's specific characteristics and distribution. When the data are normally distributed, scaling using variance might be appropriate, as it transforms the data to have a mean of zero and a variance of one. Alternatively, when the data are not normally distributed, scaling using max-min normalization may prove more effective.

4.3. Baseline and variant ML models

In our study, we introduce a *baseline* model, delineated in Fig. 12, which does not incorporate any attention layers. This model amalgamates the flight-embedded vector $\mathbf{f}$ with the weather-embedded vector $\mathbf{w}$ and subsequently fuses them into the OutputNet $\mathcal{O}$. Additionally, we explore two variants of the baseline model: one adorned with the inclusion of *queue information*, noted as *baseline v.1*, and another incorporating queue information and time of day, referred to as *baseline v.2*. The *queue information* in this context refers to the pseudo landing order of aircraft, which is determined by arranging the aircraft in ascending order based on their distances to the airport. Meanwhile, the cyclic nature of traffic density at different *time of day* (to account for the peak and non-peak hour) is represented using sine and cosine functions.

[3] https://mesonet.agron.iastate.edu/request/download.phtml (last accessed on December 04, 2024).



(a) Mean. (b) Standard deviation.

Fig. 11. Local mean and standard deviation of transit time (in seconds) of flights from the test dataset across the considered airspace.

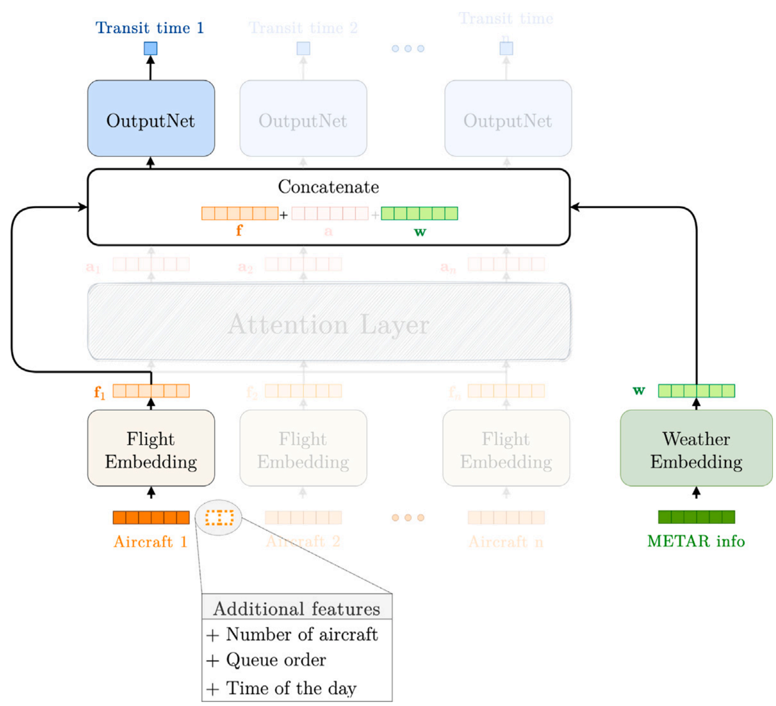


Fig. 12. Illustration of the commonplace FNN model for individual flight ETA prediction.

Fundamentally, the baseline models can be considered on par with current state-of-the-art models. Specifically, these baseline models are primarily developed based on the principles outlined by Wang et al. [16]. In contrast to their strategy of training separate models for distinct trajectory clusters, we opt not to incorporate clustering information at all, since determining a flight's cluster is feasible only upon its arrival at the airport—a detail unavailable at the time of inference. Furthermore, we utilize a more comprehensive array of features compared to their approach, encompassing the specification and weather features detailed in Table 1.

In contrast, our proposed model, referred to as *tanh-Attention* (as discussed in Section 3.2), integrates an attention layer that can potentially enhance the model's performance. The proposed model also includes two variations: *Standard Attention* and *LSTM embedding*, both sharing the same architecture as illustrated in Fig. 4. The former integrates a standard attention layer, which employs a *softmax* operator to dynamically allocate weights to different aircraft proportionally based on their relevance. Meanwhile, the *LSTM embedding* utilizes the proposed attention layer in conjunction with an LSTM network for the Flight Embedding $\mathcal{F}$. This allows the model to analyze a time series comprising five consecutive data points, ranging from $(t-4)$ to t in minutes. LSTM possesses the ability to capture temporal patterns in the flight trajectory, which is expected to further enhance the model's predictive capabilities [33].

All models in our study employ the same input features, except for the *LSTM embedding* model which handles time series data. It is worth noting that all components within these models have similar numbers of parameters, such that they share a similar level of complexity. To provide better clarity, we have listed all the models investigated in our study in Table 2.

5. Results

This section presents our analysis results and compares the capability of the developed transit time prediction models against the baseline

Table 2
Comparison of model structures between proposed models and baselines. The complexity of models listed herein is comparable, in that all components within these models have similar numbers of parameters.

Model	Components			
	$\mathcal{F}$	$\mathcal{W}$	$\mathcal{A}$	$\mathcal{O}$
Baseline	FNN	FNN	✗	FNN
Baseline v.1 (with queue)	FNN	FNN	✗	FNN
Baseline v.2 (with time and queue)	FNN	FNN	✗	FNN
Standard Attention	FNN	FNN	Standard	FNN
tanh-Attention	FNN	FNN	Proposed	FNN
LSTM embedding	LSTM	FNN	Proposed	FNN

models. To gain further insight into the role of the attention layer in the model, the investigation on how the model allocates attention over time during the learning process is conducted and presented. Lastly, a feature importance analysis is carried out to quantify the relative importance of input features to model outputs. The results of this analysis can also reveal the intuitiveness of our model, i.e., whether the outcomes are aligned and consistent with our expectations and realistic operations.

5.1. Model's predictive capability

All the models listed in Table 2 are trained using PyTorch for a total of 30 epochs, and their outcomes are displayed in Table 3. To avoid look-ahead bias, we split the dataset chronologically into two parts, one for training and the other for testing, with an 80:20 ratio. As mentioned in Section 4.1, we consider different airspace regions in our study, which are shown as the blue and green region in Fig. 8. In particular, we test our model within the green region, blue region, and the combined blue + green region, as listed in Table 3. Specifically, when mentioning the green region, all aircraft within the TMA are integrated into the model and are factored into the training of model parameters. In the context of the blue region, all aircraft within a radius of 100 NM are considered in the model; nevertheless, solely the metric errors of those within the blue region (i.e., the ring-shaped region) are leveraged for the refinement of model parameters. Meanwhile, in the case of the blue + green region, all aircraft within a 100 NM radius are integrated into the model and are considered for the training of model parameters. The minimum error in each scenario is highlighted in orange. Our proposed approach, *tanh-Attention*, consistently produces the lowest errors across all scenarios. The detailed results are presented and discussed below.

Within the green region, the *baseline* model exhibits errors of 59 s and 120.8 s in terms of MAE and RMSE, respectively. Despite providing additional information about the airspace, the variants of the baseline model do not yield any improvements. This can be attributed to two reasons. Firstly, the pseudo queue information is not useful in the context of the HKIA TMA, as being closer to the airport does not necessarily indicate an earlier landing time due to the presence of specific STARs. Secondly, the time-of-day information can only provide a general indication of whether it is a peak hour or not, without being able to clearly describe the semantic context of the airspace. In contrast, *Standard Attention* achieves notably lower levels of errors, although still worse than the *tanh-Attention*. Our proposed model, *tanh-Attention*, reduces the RMSE to 113.0 s and the MAE to 45.1 s, which corresponds to an approximately 24% reduction in MAE over the baselines. It also outperforms the *Standard Attention* model with a reduction of nearly 5 s on average. This outcome validates the argument that the original attention layer, which uses the *softmax* operation for calculating attention weights, is not appropriate for this air traffic problem. The reasons behind this unsuitability will be investigated further in Section 5.2.1. On the other hand, the *LSTM embedding* model performs the worst among all the proposed models (though it still outperforms the *baseline* variants), exhibiting the highest MAE and RMSE values. Despite incorporating an LSTM to augment the input with both current and previous states and utilizing the modified attention layer, which has previously shown effectiveness, this approach does not outperform the one using only the current state. The reduced accuracy and increased computational demands associated with the *LSTM embedding* model deem it ineffective for the specific problem at hand.

In the blue region, not surprisingly, the queue information improves the prediction performance. In this region, the distribution of aircraft is sparse, and most flight trajectories are straight towards the airport. Therefore, the order of aircraft based on their distance to the destination becomes highly informative, leading to an approximately 15% reduction in MAE and RMSE for the two variants of the *baseline* model compared to the original one. In contrast, the time-of-day information does not provide any substantial improvement, for a similar reason provided earlier. Similarly, *Standard Attention* and *LSTM embedding* do not yield any surprising results in this region. In fact, these models even lose their advantages compared to the *baseline* variants. On the other hand, the superior performance of *tanh-Attention* becomes more pronounced in this region, with approximately a 26% reduction in MAE compared to the *baseline* model and a 12% decrease compared to its variants.

Experiments conducted in the extended airspace that includes both green and blue regions consistently yield similar trends as observed in the individual regions. At the worst performance, the *baseline* achieves an MAE of 103.3 s and an RMSE of 199.9 s. In contrast, the *tanh-Attention* model provides solutions with an MAE of 71.3 s and an RMSE of 139.9 s. These results clearly demonstrate that the *tanh-Attention* model outperforms the baseline variants, which are considered the current state-of-the-art models. Specifically, the *tanh-Attention* reduces the MAE by approximately 25% and the RMSE by 20% compared to the best of the baseline variants.

Fig. 13 demonstrates the difference between the RMSE of the *tanh-Attention* and the SD inherent in the input data across the extended airspace. This comparison provides insights into the model's performance in relation to the data variability and its efficacy compared to using mean transit times, with SD acting as a reference point for RMSE [55]. Conceptually, RMSE bears similarity to SD in that while SD gauges how far actual values deviate from the mean, RMSE employs a similar formula to assess the gap between actual values and the model's predictions for those values. A proficient model should generally yield more accurate predictions than a simple mean-based estimate for all predictions. Hence, RMSE is expected to better capture and mitigate the inherent randomness in the data compared to SD. Indeed, in most areas, RMSE shows lower values than SD (i.e., the quantities shown in Fig. 13 are negative), depicted by the green color. The clear appearance of holding patterns represented by green dots indicates the model's ability to discern holding-related features effectively. Although there are exceptions, mostly outside TMA, their impacts are minimal. Towards the eastern side of the TMA, SD values are slightly lower than RMSE, likely attributed to the pre-established landing procedures for aircraft in that region. In contrast, the model excels on the western side with higher traffic complexity, demonstrating strong performance in managing the intricacies of that airspace section.

Even though the *tanh-Attention* demonstrates superior performance compared to the machine learning baseline models, a crucial step remains: comparing it with the industry standard, which in this case is the AMAN system. AMAN is the main actor predicting aircraft arrival times and subsequently coordinating the landing sequence by providing suggested flight intents to ATCOs and pilots. However, the commercial nature of AMAN systems often limits access for research purposes—a constraint that holds true for our research team as well. This final validation step would be required prior to the full deployment of the method presented in this paper, which requires close cooperation with the relevant governing body controlling ATM.

Table 3
Comparison of the baseline, proposed approach, and its variants based on evaluation metrics. In each scenario, the error that is highlighted in orange represents the smallest value.

Model	Green region		Blue region		Green + Blue	
	MAE (s)	RMSE (s)	MAE (s)	RMSE (s)	MAE (s)	RMSE (s)
Baseline	59.0	120.8	162.9	249.8	102.2	189.3
Baseline v.1	59.1	127.9	140.4	221.8	96.3	174.5
Baseline v.2	59.7	132.0	141.2	214.3	95.7	180.6
Standard Attention	49.7	115.2	135.4	206.8	77.0	155.6
tanh-Attention	45.1	109.9	120.1	185.5	71.3	139.9
LSTM embedding	50.3	118.0	140.8	222.7	85.7	147.0

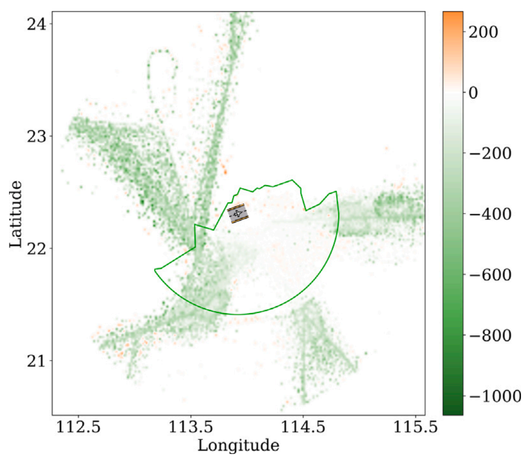


Fig. 13. Comparison between SD inherent in the input data and RMSE of the *tanh-Attention* across the considered spatial coverage. The values represented are RMSE minus SD. The visualization method employed is akin to that of Fig. 11.

5.2. Model explainability

To investigate how well the *tanh-Attention* model can deliver the intended operations, we explore the evolution of attention allotment among aircraft during the training process (Section 5.2.1). In addition, we also examine the model's relative feature importance in two distinct airspaces to gain physical insights of the model (Section 5.2.2).

5.2.1. Attention weights

To demonstrate the functionality and impact of the attention mechanism, we select two specific data samples, one from the training dataset and another from the testing dataset. Fig. 14 showcases the training sample, representing the real scenario that occurred on 2018-04-30 at 12:50 UTC. In this particular instance, there were ten aircraft present in the green region, each identified by their designated order of landing. Meanwhile, Fig. 15 visualizes the chosen test sample, recorded at 2018-07-03 10:19 UTC, which involves a total of 12 aircraft.

To evaluate our model performance, Fig. 16 provides insights into the evolution of attention weights among the aircraft during the training process, taken at several representative epochs. These attention weights, ranging from 0 to 1, indicate the level of attention allocated by one aircraft to another. A weight of 0 signifies no attention, while a weight of 1 represents full attention. Initially, at epoch 0, the aircraft demonstrate a uniform distribution of attention, wherein each aircraft pays full attention to all others. However, as the training progresses, the model begins to adapt its attention mechanism. By epoch 4, a notable change occurs, with the first scheduled-to-land aircraft starting to disregard the presence of other aircraft entirely. Subsequently, at epoch 8, another shift arises, with aircraft 1 and 2 showing a lack of attention towards other aircraft. However, it is not until epoch 16 that the trend whereby early-coming aircraft ignoring later-coming ones becomes evident across the entire system. By the 24th epoch of training, the desired attention pattern begins to emerge, with the model effectively allocating full attention to leading aircraft. Nevertheless, it takes a few more epochs for the model to stabilize and establish consistent attention weights, as demonstrated at epoch 28.

A similar analysis to evaluate the learning process of the attention mechanism is also performed using the test data sample at 2018-07-03 10:19, and the results are presented in Fig. 17. Overall, the evolution of attention weights follows a similar pattern to the previous case. However, the model exhibits some uncertainties when dealing with specific pairs of aircraft, namely 8 and 9 (circled in magenta), as well as 10 and 11 (circled in teal). In the case of aircraft 8 and 9, although it is more likely for aircraft 9 to enter the landing queue before aircraft 8 (refer to the designated landing sequence shown in Fig. 15), the information indicating that aircraft 8 is descending leads the model to speculate that aircraft 8 might land before aircraft 9. Consequently, aircraft 8 and 9 allocate a minimal amount of attention to each other, deviating from the zero and full attention observed in other pairs of aircraft. Similarly, for the pair of aircraft 10 and 11, the proximity of aircraft 11 to the airport gives the impression that it would land before aircraft 10. However, in reality, aircraft 10 obtains landing precedence over aircraft 11 due to its lighter category. This discrepancy causes aircraft 10 to pay full attention to aircraft 11, while aircraft 11 pays no attention to aircraft 10. Nevertheless, the model still manages to correctly predict the landing sequence between the two aircraft, albeit with limited accuracy in estimating their ETAs.

To compare the learning process with that of the standard attention mechanism, Fig. 18 presents the attention patterns among the aircraft using the latter. Notably, apart from the first aircraft paying full attention to the second aircraft, all other aircraft allocate their entire attention exclusively to the first landing aircraft. This phenomenon arises from the *softmax* operation employed, which prioritizes flights with high scaled dot product values. The key provided by the first aircraft, when combined with the queries of the other aircraft, generates considerably higher values compared to the rest. Consequently, the first aircraft consistently receives the spotlight. Regarding the first aircraft itself, its attention to itself is forced to be zero. As a result, the second scheduled-to-land aircraft emerges as the primary influencer. In summary, the standard attention mechanism proves to be ineffective in this context, as it fails to distribute attention appropriately among the aircraft.

5.2.2. Feature importance

Studying the feature importance of the model can provide valuable insight into whether attention truly contributes to the prediction. In this work, this is achieved by means of permutation feature importance [56], which is a widely used technique that provides a reliable and interpretable measure of feature importance. It operates by randomly shuffling the values of a single feature while keeping the others unchanged, and then measuring the resulting decrease in model performance. This approach allows us to assess the impact of each feature on the model's predictive accuracy.

When the model involves attention mechanisms, however, the traditional calculation method of feature importance might not be suitable, for reasons described below. Undoubtedly, considering the importance of attention in conjunction with other features is essential, even if it is not explicitly included as an input to the model. The reason for

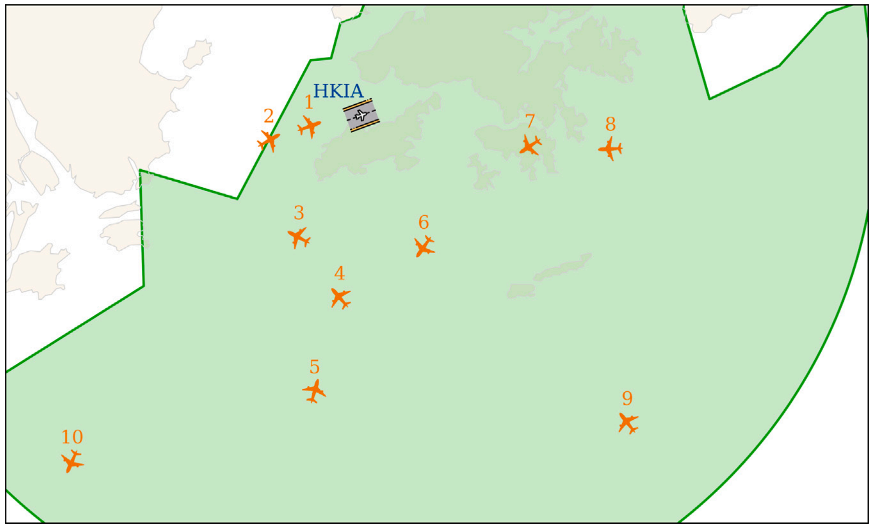


Fig. 14. Traffic sample in the training dataset at 2018-04-30 12:50.

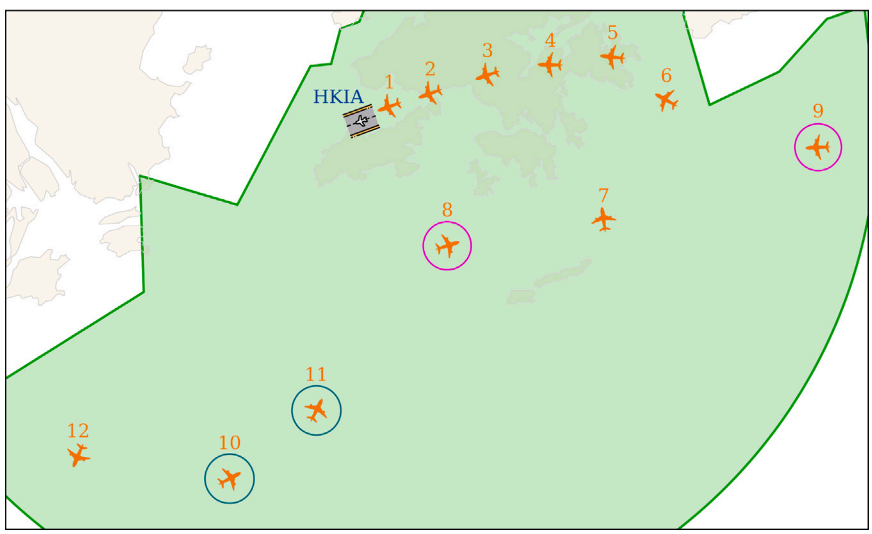


Fig. 15. Traffic sample in the test dataset at 2018-07-03 10:19.

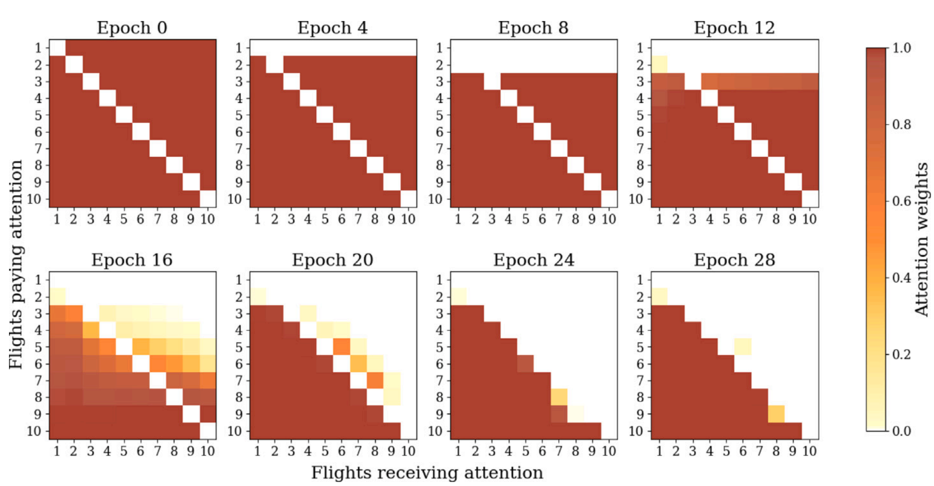


Fig. 16. Evolution of attention among aircraft in the training dataset at 2018-04-30 12:50.

this significance lies in the fact that attention can capture semantic information that greatly influences the model's predictions. Ignoring attention in the feature importance analysis would lead to the feature importance values for the features listed in Table 1 summing up to one.

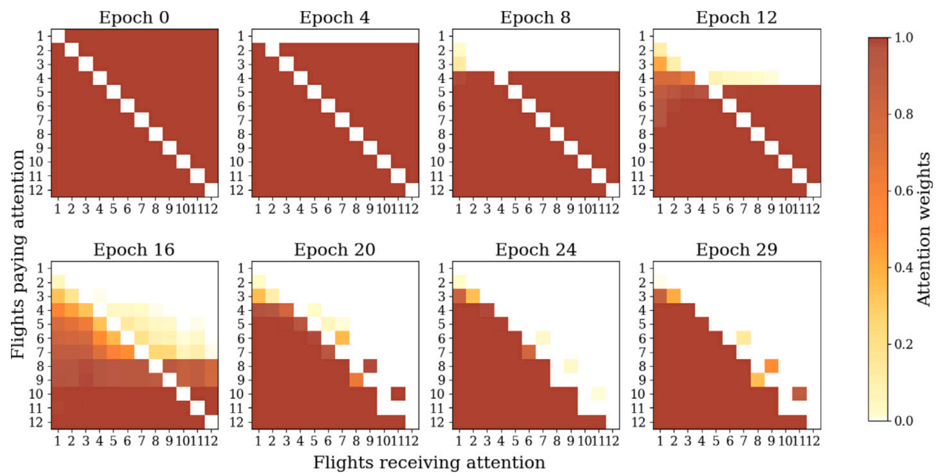


Fig. 17. Evolution of attention among aircraft in the test dataset at 2018-07-03 10:19.

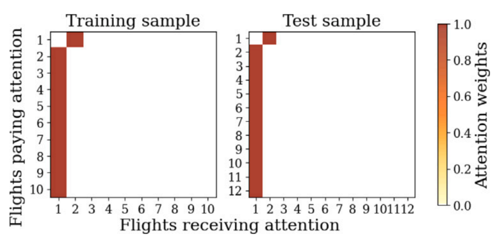


Fig. 18. Attention weights among aircraft calculated by the *Standard Attention* model at epoch 28.

Consequently, in scenarios where two aircraft share similar positions and specifications and are subjected to the same weather conditions, the model will provide the same predicted ETA, regardless of their underlying semantic differences. Clearly, this outcome is not desired, as it fails to account for the potential influence of attention in capturing the unique scenario of each aircraft.

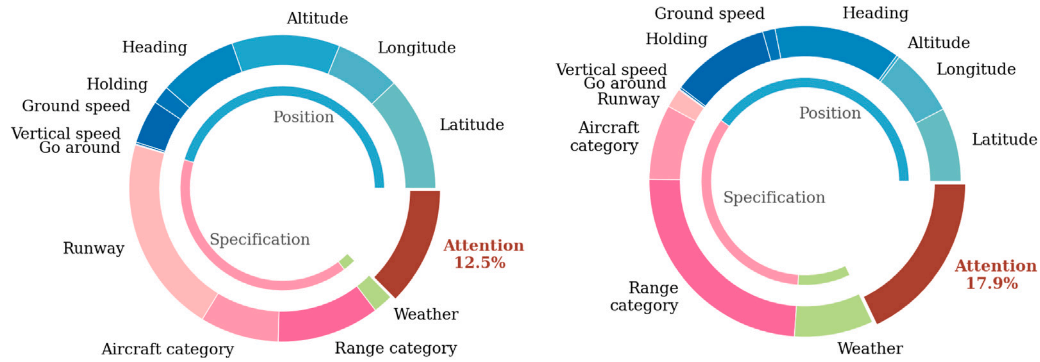
Besides, traditional feature importance calculation methods typically require fixed-length input feature vectors, which poses a challenge in the proposed model due to the varying number of aircraft. Herein, the attention layer takes the attentions from all other aircraft as inputs and evaluates the attention vector by taking their sum. Since the number of aircraft can vary, the number of inputs to the attention layer, and consequently to the entire model, also varies. To ensure that the input vector going into the attention layer always has a fixed length, the attention vector calculation needs to be done *offline* using a pre-trained attention layer. This approach is hence referred to as the *detachment technique*. In other words, the attention layer is removed from the original network in Fig. 4, to be replaced by a pre-computed attention vector. Subsequently, the attention vector is fused directly into the output network. The revised network closely resembles the baseline model, except for the direct integration of the attention vector into the OutputNet.

Fig. 19 displays the variations in feature importance between the blue and green regions. The aim is to highlight the importance of the attention mechanism in predicting ETA. It is observed that the attention mechanism has a greater contribution in the blue region. This distinction arises from the fact that inside the TMA (*i.e.,* the green region in Fig. 8), the landing sequence is more predictable and the airspace is well-structured, unlike in the blue region. In highly organized systems, the value of attention decreases because aircraft are more likely to follow their assigned landing slots, reducing the need for capturing intricate relationships.

Compared to the green region, there are several noticeable changes in feature importance beyond attention when the model is used for the blue region. Firstly, the importance of altitude decreases since most aircraft in the blue region are likely to fly at similar altitudes. In contrast, holding-related features gain greater importance as holding patterns are predominantly executed outside or at entry waypoints of the TMA. Moreover, there are substantial variations in the importance of specification features. Runway information becomes less important in the blue region, while aircraft and range categories gain greater significance. It is worth noting that aircraft and range categories are highly correlated (*i.e.,* a smaller aircraft typically flies a short-haul flight and a larger aircraft a long-haul flight), and both factors play a crucial role in determining the optimal landing sequence. Lastly, when the model is applied in the blue region, it places more emphasis on weather-related features. Recall that the blue region covers a larger space than the green one, as shown in Fig. 8, and consequently, aircraft within the former fly a longer distance to the airport than those in the latter. Intuitively, the longer an aircraft flies, the more time the weather can affect its landing process, resulting in a larger impact on ETA prediction.

With the attention mechanism being a key component in the proposed model, we wish to investigate which features contribute to the model's ability to learn attention weights among aircraft in a perceptive manner. Specifically, we focus on analyzing the factors that affect the attention weights between aircraft. To recall, the attention weight of aircraft i towards aircraft j is determined by the value obtained after applying the *tanh* operation to the scaled dot product of query i and key j.

The feature importance analyses for attention weights in both regions are depicted in Fig. 20. It reveals that the attention layer gives greater consideration to position-related features when determining whether or not to pay attention when an aircraft is outside the TMA region. However, the contribution of each feature within this category varies due to the structural characteristics of the airspaces under study. In the TMA region, where two out of three primary arrival directions are from south to north, latitude becomes more important than longitude in influencing the attention weights. On the other hand, in the blue airspace, where the incoming directions of aircraft are more uniformly



(a) In the green region. (b) In the blue region.

Fig. 19. Permutation feature importance of the *tanh-Attention* model.

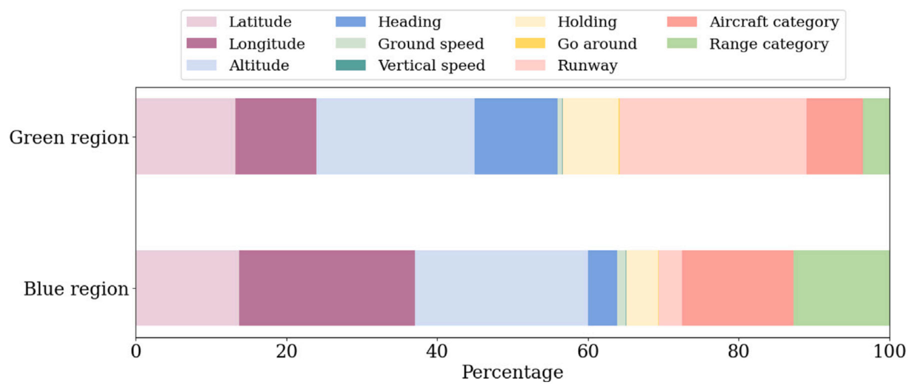


Fig. 20. Permutation feature importance of attention weights computed by the *tanh-Attention*.

distributed, the importance of latitude and longitude in the attention mechanism becomes more balanced. Meanwhile, the importance of altitude is even more pronounced in the blue region. Although altitude may not directly impact ETA predictions for aircraft outside the TMA, as discussed earlier, it plays a crucial role in the decision-making process of whether an aircraft should allocate attention to other aircraft. This is because the altitude of an aircraft, to a certain extent, implies its readiness to land. Regarding specifications, while the importance of the runway is highlighted in deciding which aircraft to pay attention to within the TMA region, aircraft and range categories become important factors outside the TMA. Indeed, these specifications serve as criteria for forming the landing sequence in typical operations.

Insights gained from feature importance analysis align well with intuition and actual ATC practices. This positive outcome provides confidence in the model's ability to capture relevant factors and make informed decisions, rendering it a promising tool for supporting ATC operations.

5.3. Actual applications and limitations

This subsection is dedicated to exploring the implementation requirements of our models in real-time, particularly under the assumption of pre-known holding time. It also discusses considerations for practical application and acknowledges current limitations related to the lack of disturbance modeling and the absence of robustness quantification. While general discussions are included, detailed solutions and investigations are beyond the scope of the current paper and will be a subject for future work.

5.3.1. Assumption of pre-known holding time

While assuming that pilots have access to information regarding holding durations for all remote aircraft during the model training might not sound very realistic, this factor does not hinder the practical implementation of our model as intended. As showcased in Section 5.2.1, aircraft primarily prioritize the monitoring of other aircraft ahead of them. Consequently, the absence of holding information for aircraft that are 100 NM away from the airport does not affect the accuracy of ETAs for those in closer proximity to the airport, which is our main focus. In fact, the ETAs of nearer aircraft carry greater significance in terms of arrival slot allocation. In brief, the model's attention allocation pattern ensures that the assumption during training does not undermine its intended use in real-world applications.

5.3.2. Model deployment in real-time

Our models are developed to be suitable for real-time deployment without requiring further model adjustments or data manipulation. Adhering to the data normalization procedures detailed in Section 4.2.3 is imperative, while complex data engineering is not essential. For this deployment to be successful, one important requirement is that ATCOs must have immediate access to the required flight information detailed in Table 1. By leveraging the pretrained model and factoring in its low, almost negligible computational requirements, ETA results can be obtained in real-time.

5.3.3. Model application in air traffic management and robustness

As previously highlighted, accurate ETA predictions play a pivotal role in enhancing air traffic flow upon arrival. In this subsection, we will delve into two key practical applications where a reliable remaining transit time prediction model, especially one that is capable of real-time operation, proves beneficial.

Firstly, in certain indirect arrival sequencing methodologies, ETA predictions serve as the foundation for determining the optimal landing sequence [57,58]. Initially, predicted ETAs are utilized to structure the landing order. Subsequently, the feasibility of nearby aircraft swaps is assessed by validating additional physical and operational constraints to ascertain the most efficient arrival landing sequence, aiding ATCOs in their decision-making process.

Secondly, ETA predictions can contribute to supporting green aviation practices [19]. By proactively anticipating the airport's future status, these predictions empower ATCOs to strategically delay aircraft positioned far from the airport, effectively mitigating TMA congestion. This approach not only helps conserve fuel—thus promoting sustainable aviation practices—but also streamlines the workload of ATCOs in managing TMA congestion.

Moreover, armed with insights into the TMA status, it becomes feasible for ATCOs to decide on the optimal runway configuration (*e.g.*, switching a runway from takeoff to landing mode and vice versa). This optimization is particularly relevant in airports equipped with mixed-mode runways, such as HKIA with its new third runway.

One key challenge in the application of these models in real-world scenarios is the presence of disturbances, such as environmental turbulence, equipment measurement errors, and other unforeseen factors. These disturbances can result in changes to the model inputs, subsequently affecting the outputs. Particularly concerning is the phenomenon where small deviations in inputs can lead to significant variations in outputs, potentially causing unpredictable outcomes. This underscores the importance of ensuring the robustness of these models to handle such disturbances effectively, maintaining their reliability and accuracy in practical applications.

Several pertinent methods exist for evaluating a model's robustness, including perturbation tests, directional expectation tests, confidence measurement, slice-based evaluation, and others. For a detailed understanding of these techniques, one can refer to the work by Huyen [59]. While these methodologies have not been extensively investigated in this study, they present promising avenues for future research. Indeed, the current work is part of an ongoing endeavor that will lead to model deployment and application by relevant practitioners. Along with this effort, the true performance of the model in operational settings will continue to be evaluated and validated.

6. Conclusion

The present investigation aimed to enhance ETA prediction accuracy by introducing an attention-based neural network model that incorporates the airspace's semantic information in addition to aircraft's operational parameter features. The developed model is unique in that the modified attention layer was derived to be used with a variable total attention weight to realistically mimic aircraft's paying attention to others in a dynamic environment. The proposed model architecture effectively reduced absolute prediction errors by up to 25% compared to current advanced ML models and by more than 7% compared to the standard attention layer in the extended airspace. However, we found that the addition of an LSTM to incorporate both current and previous states did not yield better results; instead, it decreased prediction accuracy and increased computational demands.

To effectively support ATCOs in their real-time decision-making process, the derived model needs to be able to mimic their intuitions. In other words, the role of ML techniques employed herein is to make the decision-making process more scalable and with reduced subjectivity while ensuring consistent decisions with those made by humans. Indeed, our results showed that adding the attention layer could better align the model performance with actual practices. In fact, the "importance" of attention vector reached 12.5% and 17.9% in the two studied airspaces, as revealed by the feature importance analysis. To make this analysis possible with the modified attention layer, a new detachment technique was introduced. Consistent with our intuition, the role of attention increased in the less-controlled, less-structured region, which referred to the larger airspace (i.e., the blue region). In general, factors that tend to have higher variations and frequencies—such as weather conditions and holding in the blue region—would consequently contribute more to the prediction, as opposed to factors that are more controlled—such as flight altitude prior to entering TMA. When performed on attention weight, the feature importance analysis revealed that when the aircraft was in the blue region, it paid more attention to decision factors for landing sequencing, such as position, aircraft category, and range category. Conversely, as the aircraft entered the green region (TMA), features related to the distance to the airport, including position and runway direction, became crucial determinants.

Our work showcased the potential of attention-based models—once tailored to suit the nature of aircraft operations—to optimize real-time ETA predictions and improve airport operational efficiency. These accurate ETA predictions could help ATCOs manage air traffic congestion and make sequencing decisions, as discussed in Section 5.3.3. While the model was designed to be generalizable, some customizations (especially on airspace shapes and weather conditions) might be required when applying it to other airports and regions to capture their characteristics more effectively. We have carefully included key operational factors in the model derivation; however, there might be some potential disturbances that, when unmodeled, could affect the model performance. For a more realistic application, especially considering the stochasticity of real-world air traffic operations, future work should focus on improving and validating the robustness of the model under operational disturbances.

Beyond air transportation applications, the developed attention mechanism could lend itself to other multi-agent systems that require variable total attention weights to account for their dynamic environment, such as other sequencing systems where the total number of agents could vary.

Abbreviations

ADS-B	Automatic Dependent Surveillance-Broadcast
ATC	Air Traffic Control
ATCO	Air Traffic Control Officers
ATM	Air Traffic Management
ETA	Estimated Time of Arrival
ETMS	Enhanced Traffic Management System
FNN	Feed-Forward Neural Network
HKIA	Hong Kong International Airport
IAF	Initial Approach Fix
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
METAR	Meteorological Aerodrome Report
ML	Machine Learning
NLP	Natural Language Processing
NM	Nautical Mile
SD	Standard Deviation
STAR	Standard Terminal Arrival
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
TMA	Terminal Maneuvering Area
$\mathcal{A}$	Attention layer
$\mathcal{F}$	Flight embedding network
$\mathcal{O}$	Output network
$\mathcal{W}$	Weather embedding network
$\mathbf{X}$	Aircraft set at a time point
$\mathbf{Y}$	Ground-truth remaining transit time

CRediT authorship contribution statement

Chris H.C. Nguyen: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Rhea P. Liem:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was supported by the Innovation and Technology Commission (Project No. ITS/016/20). We would like to acknowledge the support from the University Grants Committee of the Hong Kong Special Administrative Region for providing financial support to the first author through the Hong Kong PhD Fellowship Scheme (Reference No. PF20-50039).

Data availability

Data will be made available on request.

References

[1] Air Transport Action Group (ATAG), Aviation: Benefits beyond borders, 2020, https://aviationbenefits.org/downloads/aviation-benefits-beyond-borders-2020/. (Accessed on 4 December 2024).
[2] EUROCONTROL, European aviation trends paper on summer 2024 performance, 2024, https://www.eurocontrol.int/article/european-aviation-trends-paper-summer-2024-performance. (Accessed on 4 December 2024).
[3] S.G. Hamzawi, Lack of airport capacity: Exploration of alternative solutions, Transp. Res. A 26 (1) (1992) 47–58, http://dx.doi.org/10.1016/0965-8564(92)90044-8.
[4] Z. Wang, M. Liang, D. Delahaye, Automated data-driven prediction on aircraft estimated time of arrival, J. Air Transp. Manag. 88 (2020) 101840, http://dx.doi.org/10.1016/j.jairtraman.2020.101840.
[5] Y. Glina, R. Jordan, M. Ishutkina, A tree-based ensemble method for the prediction and uncertainty quantification of aircraft landing times, in: American Meteorological Society–10th Conference on Artificial Intelligence Applications to Environmental Science, New Orleans, LA, 2012.
[6] J. Zhang, Z. Peng, C. Yang, B. Wang, Data-driven flight time prediction for arrival aircraft within the terminal area, IET Intell. Transp. Syst. 16 (2) (2022) 263–275, http://dx.doi.org/10.1049/itr2.12142.
[7] J. Krozel, C. Lee, J. Mitchell, Estimating time of arrival in heavy weather conditions, in: Guidance, Navigation, and Control Conference and Exhibit, 1999, p. 4232, http://dx.doi.org/10.2514/6.1999-4232.
[8] T. Mueller, J. Sorensen, G. Couluris, Strategic aircraft trajectory prediction uncertainty and statistical sector traffic load modeling, in: AIAA Guidance, Navigation, and Control Conference and Exhibit, 2002, p. 4765, http://dx.doi.org/10.2514/6.2002-4765.
[9] P.T. Wang, C.R. Wanke, F.P. Wieland, Modeling time and space metering of flights in the national airspace system, in: Proceedings of the 2004 Winter Simulation Conference, 2004, Vol. 2, IEEE, 2004, pp. 1299–1304, http://dx.doi.org/10.1109/WSC.2004.1371463.
[10] L. Xi, Z. Jun, Z. Yanbo, L. Wei, Simulation study of algorithms for aircraft trajectory prediction based on ADS-B technology, in: 2008 Asia Simulation Conference-7th International Conference on System Simulation and Scientific Computing, IEEE, 2008, pp. 322–327, http://dx.doi.org/10.1109/ASC-ICSC.2008.4675378.
[11] M. Porretta, M.-D. Dupuy, W. Schuster, A. Majumdar, W. Ochieng, Performance evaluation of a novel 4D trajectory prediction model for civil aircraft, J. Navig. 61 (3) (2008) 393–420, http://dx.doi.org/10.1017/S0373463308004761.
[12] J.V. Benavides, J. Kaneshige, S. Sharma, R. Panda, M. Steglinski, Implementation of a trajectory prediction function for trajectory based operations, in: AIAA Atmospheric Flight Mechanics Conference, 2014, p. 2198, http://dx.doi.org/10.2514/6.2014-2198.
[13] S. Ayhan, P. Costas, H. Samet, Predicting estimated time of arrival for commercial flights, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 33–42, http://dx.doi.org/10.1145/3219819.3219874.
[14] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014, http://dx.doi.org/10.48550/arXiv.1409.0473, arXiv preprint arXiv:1409.0473.
[15] C.S. Kern, I.P. de Medeiros, T. Yoneyama, Data-driven aircraft estimated time of arrival prediction, in: 2015 Annual IEEE Systems Conference (SysCon) Proceedings, IEEE, 2015, pp. 727–733, http://dx.doi.org/10.1109/SYSCON.2015.7116837.
[16] Z. Wang, M. Liang, D. Delahaye, A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area, Transp. Res. C 95 (2018) 280–294, http://dx.doi.org/10.1016/j.trc.2018.07.019.
[17] I. Dhief, Z. Wang, M. Liang, S. Alam, M. Schultz, D. Delahaye, Predicting aircraft landing time in extended-TMA using machine learning methods, in: Proceedings of 9th International Conference for Research in Air Transportation, ICRAT, 2020.
[18] X. Gui, J. Zhang, Z. Peng, C. Yang, Data-driven method for the prediction of estimated time of arrival, Transp. Res. Rec. 2675 (12) (2021) 1291–1305, http://dx.doi.org/10.1177/03611981211033295.
[19] L.Z. Jun, S. Alam, I. Dhief, M. Schultz, Towards a greener Extended-Arrival Manager in air traffic control: A heuristic approach for dynamic speed control using machine-learned delay prediction model, J. Air Transp. Manag. 103 (2022) 102250, http://dx.doi.org/10.1016/j.jairtraman.2022.102250.
[20] J. Bao, Z. Yang, W. Zeng, Graph to sequence learning with attention mechanism for network-wide multi-step-ahead flight delay prediction, Transp. Res. C 130 (2021) 103323, http://dx.doi.org/10.1016/j.trc.2021.103323.
[21] K. Cai, Y. Li, Y.-P. Fang, Y. Zhu, A deep learning approach for flight delay prediction through time-evolving graphs, IEEE Trans. Intell. Transp. Syst. 23 (8) (2021) 11397–11407, http://dx.doi.org/10.1109/TITS.2021.3103502.
[22] G. Fancello, C. Pani, M. Pisano, P. Serra, P. Zuddas, P. Fadda, Prediction of arrival times and human resources allocation for container terminal, Marit. Econ. Logist. 13 (2011) 142–173, http://dx.doi.org/10.1057/mel.2011.3.
[23] W.-H. Lin, J. Zeng, Experimental study of real-time bus arrival time prediction with GPS data, Transp. Res. Rec. 1666 (1) (1999) 101–109, http://dx.doi.org/10.3141/1666-12.
[24] S.I.-J. Chien, Y. Ding, C. Wei, Dynamic bus arrival time prediction with artificial neural networks, J. Transp. Eng. 128 (5) (2002) 429–438, http://dx.doi.org/10.1061/(ASCE)0733-947X(2002)128:5(429).
[25] M. Yaghini, M.M. Khoshraftar, M. Seyedabadi, Railway passenger train delay prediction via neural network model, J. Adv. Transp. 47 (3) (2013) 355–368, http://dx.doi.org/10.1002/atr.193.
[26] N. Marković, S. Milinković, K.S. Tikhonov, P. Schonfeld, Analyzing passenger train arrival delays with support vector regression, Transp. Res. C 56 (2015) 251–262, http://dx.doi.org/10.1016/j.trc.2015.04.004.
[27] H. Ij, Statistics versus machine learning, Nat. Methods 15 (4) (2018) 233, http://dx.doi.org/10.1038/nmeth.4642.
[28] N. Meinshausen, G. Ridgeway, Quantile regression forests, J. Mach. Learn. Res. 7 (6) (2006).
[29] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Vol. 96, No. 34, 1996, pp. 226–231.
[30] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133, http://dx.doi.org/10.1007/BF02478259.
[31] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol Rev 65 (6) (1958) 386, http://dx.doi.org/10.1037/h0042519.
[32] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, http://dx.doi.org/10.1038/323533a0.
[33] C. Deng, H.-C. Choi, H. Park, I. Hwang, Trajectory pattern identification and classification for real-time air traffic applications in Area Navigation terminal airspace, Transp. Res. C 142 (2022) 103765, http://dx.doi.org/10.1016/j.trc.2022.103765.
[34] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.
[35] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, Adv. Neural Inf. Process. Syst. 31 (2018).
[36] M.-T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, 2015, http://dx.doi.org/10.48550/arXiv.1508.04025, arXiv preprint arXiv:1508.04025.
[37] Y. Wu, M. Schuster, Z. Chen, Q.V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., Google's neural machine translation system: Bridging the gap between human and machine translation, 2016, http://dx.doi.org/10.48550/arXiv.1609.08144, arXiv preprint arXiv:1609.08144.
[38] J.K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, Y. Bengio, Attention-based models for speech recognition, Adv. Neural Inf. Process. Syst. 28 (2015).
[39] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, Y. Bengio, End-to-end attention-based large vocabulary speech recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2016, pp. 4945–4949, http://dx.doi.org/10.1109/ICASSP.2016.7472618.
[40] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in: International Conference on Machine Learning, PMLR, 2015, pp. 2048–2057.
[41] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, L. Zhang, Bottom-up and top-down attention for image captioning and visual question answering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077–6086, http://dx.doi.org/10.1109/CVPR.2018.00636.

[42] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training, 2018, OpenAI, URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

[43] J. Yu, Short-term airline passenger flow prediction based on the attention mechanism and gated recurrent unit model, Cogn. Comput. 14 (2) (2022) 693–701, http://dx.doi.org/10.1007/s12559-021-09991-x.

[44] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, http://dx.doi.org/10.48550/arXiv.1406.1078, arXiv preprint arXiv:1406.1078.

[45] P. Jia, H. Chen, L. Zhang, D. Han, Attention-LSTM based prediction model for aircraft 4-D trajectory, Sci. Rep. 12 (1) (2022) 1–11, http://dx.doi.org/10.1038/s41598-022-19794-1.

[46] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, http://dx.doi.org/10.48550/arXiv.1609.02907, arXiv preprint arXiv:1609.02907.

[47] 2015 Airspace Concept and Strategy for the ECAC States and Key Enablers, EUROCONTROL, 2008.

[48] A. Graves, N. Jaitly, A.-r. Mohamed, Hybrid speech recognition with deep bidirectional LSTM, in: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, IEEE, 2013, pp. 273–278, http://dx.doi.org/10.1109/ASRU.2013.6707742.

[49] H. Sak, A.W. Senior, F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling, 2014, http://dx.doi.org/10.48550/arXiv.1402.1128, arXiv preprint arXiv:1402.1128.

[50] A. Graves, G. Wayne, I. Danihelka, Neural turing machines, 2014, http://dx.doi.org/10.48550/arXiv.1410.5401, arXiv preprint arXiv:1410.5401.

[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[52] C.H. Nguyen, N.L. Go, K.Y. Hui, R. Liem, Tactical routing for air transportation in HKIA terminal maneuvering area, in: 26th International Conference of Hong Kong Society for Transportation Studies, HKSTS, 2022.

[53] G.N. Lui, T. Klein, R.P. Liem, Data-driven approach for aircraft arrival flow investigation at terminal maneuvering area, in: AIAA Aviation 2020 Forum, 2020, p. 2869, http://dx.doi.org/10.2514/6.2020-2869.

[54] G.N. Lui, K.K. Hon, R.P. Liem, Weather impact quantification on airport arrival on-time performance through a Bayesian statistics modeling approach, Transp. Res. C 143 (2022) 103811, http://dx.doi.org/10.1016/j.trc.2022.103811.

[55] G. Shmueli, P.C. Bruce, I. Yahav, N.R. Patel, K.C. Lichtendahl Jr., Data Mining for Business Analytics: Concepts, Techniques, and Applications in R, John Wiley & Sons, 2017.

[56] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[57] Z. Du, J. Zhang, B. Kang, A data-driven method for arrival sequencing and scheduling problem, Aerospace 10 (1) (2023) 62, http://dx.doi.org/10.3390/aerospace10010062.

[58] Y. Pang, P. Zhao, J. Hu, Y. Liu, Machine learning-enhanced aircraft landing scheduling under uncertainties, Transp. Res. C 158 (2024) 104444, http://dx.doi.org/10.1016/j.trc.2023.104444.

[59] C. Huyen, Designing Machine Learning Systems, O'Reilly Media, 2022, URL: https://books.google.com.vn/books?id=EThwEAAAQBAJ. (Accessed on 4 December 2024).