# MACHINE LEARNING FOR IMPROVING AIR MOBILITY UNDER EMERGENCY SITUATIONS

**FINAL REPORT**

**JANUARY 2024**

**AUTHORS**

**DR. YONGXIN LIU**
**DR. DAHAI LIU**
**YUJING ZHOU**
**EMBRY-RIDDLE AERONAUTICAL UNIVERSITY**

**US DEPARTMENT OF TRANSPORTATION GRANT 69A3551747125**

## DISCLAIMER

| 1. Report No. | 2. Government Accession No. | | 3. Recipient's Catalog No. |
|---|---|---|---|
| 4. Title and Subtitle<br>Machine Learning for Improving Air Mobility under Emergency Situations | | | 5. Report Date<br>January 2024<br><br>6. Source Organization Code |
| 7. Author(s)<br>Yongxin Liu<br>Dahai Liu<br>Yujing Zhou | | | 8. Source Organization Report No.<br>CATM-2024-R3-ERAU |
| 9. Performing Organization Name and Address<br><br>Center for Advanced Transportation Mobility<br>Transportation Institute<br>1601 E. Market Street<br>Greensboro, NC 27411 | | | 10. Work Unit No. (TRAIS)<br><br>11. Contract or Grant No.<br>69A3551747125 |
| 12. Sponsoring Agency Name and Address<br><br>University Transportation Centers Program (RDT-30)<br>Office of the Secretary of Transportation–Research<br>U.S. Department of Transportation<br>1200 New Jersey Avenue, SE<br>Washington, DC 20590-0001 | | | 13. Type of Report and Period Covered<br>Final Report:<br>October 2021 – January 2024<br><br>14. Sponsoring Agency Code<br>USDOT/OST-R/CATM |
| 15. Supplementary Notes: | | | |
| 16. Abstract<br><br>This project investigates the integration of advanced machine learning and optimization techniques to improve air traffic management during emergencies. Over three phases, the study utilized AI to predict flight delays, optimize evacuation processes, and plan efficient evacuation flight paths, enhancing the resilience of air mobility systems in crises. The first phase developed an explainable GRU neural network to predict weather-related airport capacity constraints using historical data. The second phase employed Particle Swarm Optimization (PSO) to optimize air travel for emergency evacuations, demonstrating cost-effective and rapid mobilization of resources. The final phase introduced a hybrid model combining a genetic algorithm with a neural network to enhance evacuation flight planning. The study's findings highlight the potential of AI in emergency air mobility and offer recommendations for policymakers, airline operators, and researchers to advance these technologies and their practical applications. | | | |
| 17. Key Words<br><br>Machine learning, emergency evacuation, optimal scheduling, explainable AI | 18. Distribution Statement<br><br>Unrestricted; Document is available to the public through the National Technical Information Service; Springfield, VT. | | |
| 19. Security Classif. (of this report)<br><br>Unclassified | 20. Security Classif. (of this page)<br><br>Unclassified | 21. No. of Pages<br><br>87 | 22. Price<br><br>… |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

## EXECUTIVE SUMMARY


This project presents a comprehensive examination of the integration of advanced machine learning and optimization techniques to enhance air traffic management during emergency scenarios. This multi-phase study, encompassing three distinct but interconnected phases, leverages cutting-edge artificial intelligence (AI) to predict flight delays, optimize evacuation processes, and plan evacuation flight paths efficiently, thereby contributing significantly to the resilience and responsiveness of air mobility systems in the face of natural disasters and other emergencies.

The initial phase of the project focused on developing an explainable machine learning model, specifically a gated recurrent unit (GRU) neural network, to predict weather-related airport capacity constraints. By analyzing three years of historical weather and flight data, the model successfully predicted the number of flights arriving with Estimated Departure Clearance Time (EDCT) delays. This phase emphasized the importance of accurate, real-time data analysis in forecasting airport capacity constraints, highlighting the potential for further improvements in model accuracy through the inclusion of additional data sources.

In the second phase, the project shifted focus towards optimizing the use of air travel for emergency evacuations through a cost-aware approach. Utilizing Particle Swarm Optimization (PSO), the study identified optimal flight selections to maximize the efficiency of evacuation efforts from the Daytona Beach International Airport. This phase demonstrated the feasibility of rapidly mobilizing air travel resources in emergency situations, prioritizing cost-effectiveness and operational capacity to facilitate swift and large-scale evacuations.

The final phase of the project explored the use of a hybrid model combining a genetic algorithm (GA) with a neural network (NN) to plan evacuation flights. This innovative approach significantly reduced computational overhead while improving the efficiency and effectiveness of evacuation planning. The findings underscore the potential of combining different AI methodologies to enhance decision-making processes in emergency air mobility management.

Our recommendations for policy makers include investing in data infrastructure, supporting technological innovation, and updating regulatory frameworks to incorporate advanced predictive tools. Airline operators are advised to adopt these advanced predictive and optimization tools, participate in data sharing initiatives, and invest in flexible resource management systems. Researchers are encouraged to focus on the generalizability of models, the development of explainable AI, and the practical impact of these technologies on emergency air mobility.

# 1. BACKGROUND AND LITERATURE REVIEW

In the United States, there are approximately 45,000 flights managed by the Federal Aviation Administration (FAA) every day. About 5,400 aircraft operate simultaneously in the sky at peak operational times, with more than 2,800,000 passengers flying daily in and out of U.S. airports (FAA, 2021a). Under the complex airspace system with a vast flow of aircraft and passengers, emergency situations could happen at any time, anywhere. Even a small error, like the wrong size bolt, can cause emergency situations in aviation operations. In 1990, an improperly installed windshield in the cockpit on BAC One-Eleven from British Airways was broken while the aircraft was cruising. The pilot was partially blown out of the aircraft and almost caused a crash, which risked hundreds of lives on the aircraft. (Ranter, n.d.). Emergency situations in aviation can impact the efficiency of normal operations and air mobility severely. Moreover, delays, incidents, or accidents resulting from emergency situations will make airline companies and airports lose both their reputation and economy.

Large emergency situations, including extreme weather, military operations, and abnormal situations involving crewmembers or air traffic control can cause large-scale delays, which can affect air mobility greatly over a wide range. Analyzing the causes behind the emergencies that can cause large-area delays and designing a corresponding decision-making assistant system is one essential ways to alleviate the impact of emergency situations in the aviation industry. Researchers and computer scientists have studied how different reasons and factors affect air mobility through various techniques. This study will review earlier research and provide an updated view of using machine learning to improve air mobility in emergency situations.

In this chapter, we will review earlier studies that aim to improve air mobility under emergency situations and provide a better view of using machine learning techniques and algorithms, including reinforcement learning, deep learning, random forest, artificial neural network, etc., to improve air mobility in emergency situations. In response to the issue, designing a decision-making assistant system has been widely studied to alleviate the negative impact of perturbations on aviation air mobility from a global-view perspective. Through various techniques, researchers and computer scientists have studied how different reasons and factors affect air mobility.

Different from other works related to the machine learning algorithm, our paper not only defines and analyzes air mobility and emergency situations under different environments and cause factors but also reviews and examines the opportunities and challenges of each type of machine learning technique. Although deep neural networks can remember important input and capture the nonlinearity, thus providing precise prediction, it is weak in predicting and learning based on a long time period (Hochreiter & Schmidhuber, 1997). On the other hand, holding too many aircraft on the ground as one of the traffic control solutions can damage both economies, reputation, and the environment is also a challenge that is considered in the paper. Using the multi-agent reinforcement learning method is one of the corresponding ways mentioned in the later chapter, which has the advantage of maximizing multiple agents' gain in the environment without losing balance.

### 1.1 Air Mobility

Air mobility was originally used as a concept in the Army and represents the ability to ensure the balance of firepower, mobility, support, etc. (Tolson, 1973). In the modern aviation industry, this concept has been adopted to describe the ability to ensure the regular process of flights to maintain the normal airport capacity.

In the real-world aviation environment, the regular process of flight for air mobility is complicated and should be discussed and achieved from multiple aspects. There are airlines that own and operate flights, airports where passengers and aircraft arrive and depart, pilots who fly the aircraft, and air traffic control who protect the flights from collision and ensure normal operation. Every aspect and part of the operation process should be considered. Understanding the regular process of aircraft operation is essential for maintaining and optimizing the normal flight process. Referring to Wang and Zhao's paper (2020), the general flight phases can be divided into seven phases: planning phase, takeoff phase, climb phase, cruise phase, descent phase, approach phase, and taxi phase (Figure 1).

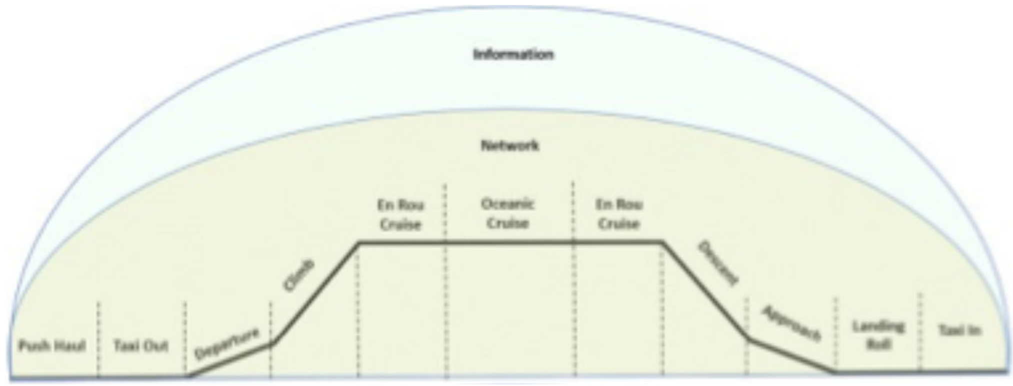



*Figure 1.* Division and composition of flight tasks and phases (Wang & Zhao, 2020, p14)

Understanding different flight phases is important for protecting air mobility from emergency situations. According to Airbus Accident Statistics (2021), fatal accidents happened in the approach and landing phases the most, while the most incidents with no fatalities happened in the landing and takeoff phases. The results indicate the importance and complexity of each different flight phase, which is significantly influenced by the pilots, air traffic control, airspace and airport capacity, and weather conditions.

To accomplish normal aircraft operation, FAA (2021b) has produced a solution that helps both the pilot and air traffic control with an automatic sending and receiving transponder. Automatic Dependent Surveillance-Broadcast (ADS-B) was required to be installed in any aircraft that operates in the controlled airspace after January 1, 2020. The ADS-B is a technology that is equipped with a 1090 MHz Mode-S transponder and utilizes satellite navigation technology. It can broadcast flight information unencrypted that can be received and decoded by any person or station on the ground with a simple set-up. (Sun & Hoekstra, 2016) The advantage of implementing ADS-B is obvious as aircraft equipped with ADS-B can enhance the awareness of air traffic control and make themselves easier to recognize and track in the airspace. The air traffic control will receive the updated information of each aircraft almost every second, which enables the controller to react to any emergency situations and prevent accidents from happening (FAA, 2021b). ADS-B is also helpful to pilots. It provides speed, altitude, distance, etc., other kinds of accurate and up-to-date information. Pilots can also receive the location information of surrounding aircraft directly on their displays. The environment and aircraft information helps pilots with their situational awareness and decision-making process. ADS-B enhanced pilot and passenger safety greatly and improved air mobility from a technical aspect.

### A. Affecting Factors

Although emergency situations can happen at any phase in flight, as analyzed in the previous section, the cause and reasoning of different emergency situations can be put in several large categories. Understanding different categories of emergency situations are helpful and essential to building an effective model. Bureau of Transportation Statistics (BTS), part of the United States Department of Transportation, collects, analyzes, and publishes accurate transportation data and information to the public. According to Bureau of Transportation Statistics report on airline on-time statistics and delay causes (Bureau of Transportation Statistics, 2021), there are five broad categories that affect air mobility: (1) Air carrier delay, (2) Extreme weather, (3) National aviation system (NAS), (4) Late-arriving aircraft, (5) Security.

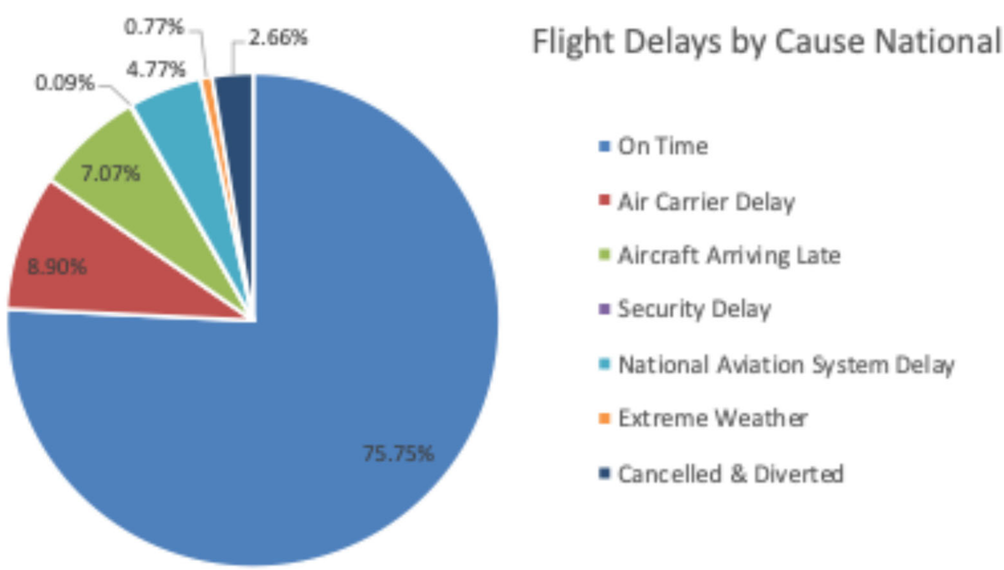



*Figure 2.* Flight Delays by Cause National (Bureau of Transportation Statistics, 2021)

### B. Air Carrier Delay

According to the report, the number one cause of flight delays is air carrier delay (Figure 2). Air carrier delay occurs within the airline control system, which includes abnormal situations caused by pilots or air traffic controllers, maintenance problems, fueling, baggage loading delay, etc., which are all defined as air carrier delay (Bureau of Transportation Statistics, 2021). Among all the problems, the abnormal situation caused by pilots and air traffic control has the most significant effect on air mobility and is also one of the most difficult factors to control because of the human factor. In the report, the challenge of aviation emergency and abnormal situations, published by NASA, Dr. Burian and Dr. Barshi (2005) addressed six issues that related to the crew's reaction to emergency and unusual situations in flight, including training for emergencies and unusual situations, as well as human performance and cognitive limitations under high workload and stress. Burian stated that, "the degree to which training truly reflects real-life emergency and abnormal situations, with all of their real-world demands, is often limited," and complex communication and cooperation are often simplified or even omitted in training (Burian & Barshi, 2005). In other words, real-life situations are much more complex and diverse than training can anticipate or simulate.

After reviewing 107 Aviation Safety Reporting System (ASRS) reports involving emergency or unusual situations, Burian and Barshi found that 19 of the 25 situations that were managed properly were "textbook" situations (Burian & Barshi, 2003). This proves that the closer the simulated situation is to the real-world emergency, the more quickly and correctly the trained person will react and choose when faced with the situation. However, the evidence also reported that 85 of the 107 situations are non-textbook situations, and 79 of these 85 situations were not handled well (Burian & Barshi, 2003). Obviously, most of the emergency or abnormal situations that occurred in the real world were not practiced during training. As a result, many flight crews did not respond well. But from another perspective, the data also confirmed that if more situations could be added to the training, most pilots can respond correctly and effectively when these emergencies occur.

Nevertheless, after considering economic cost and time cost, it is impossible to add all known emergency or abnormal situations into training (Burian, Barshi, & Dismukes, 2005). Moreover, Burian et al. also declared that most emergencies would cause high stress while increasing workload, and it is hard for a human to recall all these pieces of training under the enormous pressure of an emergency (Burian, Barshi, & Dismukes, 2005). Therefore, the combination of all the statements in Burian et al.'s report demonstrated the importance and urgency of a decision-making assistance system for overcoming the abnormal situation of the crewmember.

### C. Extreme weather

Extreme weather is the cause of only 0.82% of flight delays, which is insignificant compared to air carrier delays (7.53%) or aircraft arriving late (6.63%). However, the cost of damage, delays, and the evacuation and settlement for passengers due to extreme weather are unneglectable. Extreme weather, including hurricanes, blizzards, or tornados, can cause large-area delays, which affect the air mobility of multiple numbers of airports and airlines severely. Choi et al. (2016) pointed out in their paper that weather as one of the causes of the delays is closely related to other categories like National Aviation System and late-arriving aircraft. Although airlines report late-arriving aircraft as the delay cause instead of reporting weather, the weather is still one of the factors that could cause late-arriving aircraft. Therefore, extreme weather could account for 40% of total delay minutes, because it shares and causes the delays along with other factors (Choi et al., 2016). In the study, Choi et al. applied machine learning to build a prediction model based on historical data on traffic and weather. The supervised machine learning algorithm was implemented in the model. Although the predictive performance of the model is lower due to the uncertainties in the forecast, it still provides higher accuracy results with actual weather. Two possible prediction error sources are discussed: the limitation of the current model and the non-weather factors that caused flight delays which are not able to be captured by the model. The result indicates the randomness and uncertainty in predicting weather to keep the normal flight process.

In our case, we leverage Multi-agent Deep Reinforcement Learning (MADRL) instead of the supervised machine learning algorithm. Also, the uncertainty will be a consideration in our model, which will improve the performance of our model when compared to theirs. Another difference between our study and theirs is the object or prediction that the model is facing. Choi et al. (2016) focuses on the extreme weather factor while we also consider other factors, including the abnormal situations of the human and military operations. The difference in influencing factor consideration could lead to a universal result and help decision-making in more emergency situations.

### D. Issues and Challenges

When air mobility is affected by the five factors listed above, the cost of delay will become a challenging issue. For airlines, the cost consists of different components. Cook and Tanner (2009) considered delay cost management and divided distinct types of delay costs in accordance with each different phase of flight. Passenger costs, crewmember, maintenance costs, etc., are all considered airline delay costs and the fuel consumption while the aircraft is in the airborne phase should also be considered. Cook and Tanner provide the delay cost management from two large levels: strategic and tactical. The strategic level includes "resources committed at the planning stage," which means to predict and prevent delays or contingencies in advance. The tactical level was divided into three phases: pre-departure, airborne, and post-flight. Each phase has unique factors and solution.

## 1.2 Emergency Situations

Emergencies and unusual situations occur every day on flights and airports around the world. Whether it is an emergency serious enough to endanger people's lives and property or a small incident that is easily managed, it will have a negative impact on air mobility (Burian, Barshi, & Dismukes, 2005). By definition, emergencies are situations that people are not familiar with, which are unpredictable and need immediate attention and action according to the situation. These emergency situations can cause intense feelings of anxiety, uncertainty, and stress (Van de Walle & Turoff, 2008). Emergency situations can include a large variety of factors and vary in scale and magnitude, from small vehicle accidents to large area full-scale disasters.

However, no matter how large scale the emergency situations are, there will always be human and economic losses caused by emergency situations. Therefore, emergency management becomes an important approach to analyzing the emergency and reducing the impact. Using computer vision as a modern technology to prevent and mitigate emergency loss is one of the main research fields for emergency situations. There is a large variety of types of emergencies, and many of them have not been studied in computer vision.

Lopez-Fuentes et al. (2018) clarified in the study that emergencies like famine and pollution cannot be easily detected by visual sensors. Therefore, emergency situations are distinguished into two generic groups: natural and human-made emergency situations (Lopez-Fuentes et al., 2018). In our study, emergency situations are also distinguished in the same way: natural emergencies include extreme weather and human-made emergencies include abnormal situations and military operations. In this case, we will consider the situation as an emergency when one or more of the following unexpected situations is met (Lopez-Fuentes et al., 2018):

- a sudden situation which would or has caused risks to health or life of a person or a group of persons.
- a sudden situation which would or has caused potential loss of economy or reputation of a person, a group, or a company.
- a sudden situation which would or has caused damage to properties or environment.

In their study, Lopez-Fuentes et al. (2018) divide the life cycle of emergency management into four main phases: preparedness, response, recovery, and mitigation. Emergency situations occur between the preparedness and response phases, which are unpredictable most of the time. Because of the unpredictable character of emergency situations, the chances of emergencies could not be eliminated, only reduced. Thus, the more preventive actions or measures made in preparedness, the bigger chance to reduce the impact that could be brought by emergency situations.

In the aviation industry, the evacuation under emergency situations can be divided into two categories: aircraft evacuation and airport evacuation. According to the Federal Aviation Regulations (FAR) Part 25, "for airplane having a seating capacity of more than 44 passengers, it must be shown that the maximum seating capacity can be evacuated from the airplane to the ground under simulated emergency conditions within 90 s" (FAR, 2002). This is the "90 s evacuation rule", which is required for all the aircraft from manufacturers and airlines, has been proven effective for passengers' quick evacuation under emergency situations by numerous studies. For instance, in March 2006, there were 853 passengers, 18 crew members, and two pilots were evacuated successfully in 78 seconds in the A380 evacuation made by Airbus (Daly, 2006). Although the 90 seconds test has been proven as certification for aircraft, there are still some difficulties with the test. Passenger safety is not protected while evacuating, which means the test participants may get injured while doing the test. Data shows that about 6% of participants in the 90 seconds test are injured, ranging from cuts to broken bones (Congress, 1993). Another difficulty that exists in the evacuation test is the difference between real-world emergency situations and simulation. Sometimes a real emergency scenario is hard to achieve, which may cause a discrepancy between the test and evacuation (Fang et al., 2016).

Another category of evacuation in the aviation industry is airport evacuation. In the Code of Federal Regulations (CFR) title 14 part 139.325, FAA required that each certified airport must develop and maintain an airport emergency plan (AEP), which is designed "to minimize the possibility and extent of personal injury and property damage on the airport in an emergency." However, there is no rule for airport evacuation that points out a specific time frame for emergency situations, which is completely different from the aircraft evacuation's "90 s evacuation rule". In this case, time is not a variable nor a criteria to evaluate the effectiveness of an evacuation anymore. One of the most significant difficulties in airport evacuation and AEP is that it is impossible to apply a generic plan to all the varying designs and layouts of airports. To design and evaluate an effective airport evacuation, researchers will need to explore and build a model based on different airports because of the different layouts and locations, which is a time-consuming and complex task.

### A. Large-area Delay

When the degree of delay is measured by the number of flight delays and the delay time reaches a certain level, it is called large-area flight delay (LFD) (Gao et al., 2012). Other than aircraft accidents and incidents, large-area flight delays or flight cancellations are the last things passengers and aviation staff want to incur. There is no doubt that mass delay will cause a loss of time and money. For airlines, the important loss of credibility and reputation is also at stake. However, the most critical impact is that flight delays may compromise aviation safety. Due to the negative effects of LFD, Gao et al. focused on managing unexpected situations in response to LFD in their research paper, "Flight rescheduling responding to large-area flight delays" (Gao et al., 2012). Since LFD can have too many adverse consequences and will intensify over time, it is urgent for the aviation industry to manage those unexpected events.

Therefore, the research of Gao et al. is important, and it also proves the significance of our project to design a real-time flight rescheduling solution to resume normal operations in response to disruptions.

It has been stated that anything unexpected that causes the flight to deviate from the original plan is called "disruption," and Gao et al. claimed that both economic losses and non-quantifiable factors such as passenger satisfaction are considered when managing disruptions (Gao et al., 2012). Gao et al. emphasized the importance of the real-time status of flights, and according to the real-time status, the flights are divided into six classes to consider their priority (Gao et al., 2012). For example, a flight in the air is "prior to boarded" flight. Since a non-quantitative loss would also be considered, they designed priority indicators as well, such as flights with very important persons (VIP) and international flights, etc. (Gao et al., 2012). In other words, for all flights, their priority is decided by their statuses, while the flights in the same class are sorted by priority indicators.

After determining the priority of the flight, Gao et al. proposed two mathematic programming models to reschedule flights, and many rules were set in the inputs to get a credible output (Gao et al., 2012). Our research methods are similar in that they built mathematical models while we plan to use Multi-agent Deep Reinforcement Learning (MADRL) for flight rescheduling. Even though Dr. Gao et al. proved that their model is suitable to be applied to management disruptions (Gao et al., 2012), but it still has limitations when compared to our model. First, we consider the interests of more people, not only airlines and passengers but also airports, air traffic controllers, and other stakeholders. Moreover, while they only considered reordering the flights, we also considered rerouting and retiming the flights. Furthermore, they did not consider flight cancellation and other factors due to complexity (Gao et al., 2012). We have five real-world data sources such as ADS-B and AOTP to help our deep learning models to build an environment as close to the real world as possible. As a result, although we are both trying to reduce the negative effects of LFD, our model is more comprehensive and closer to reality and, therefore, can more accurately respond to LFD.

## B. Extreme Weather

Extreme weather is one of the most influential factors that can cause LFD. According to *Extreme Weather on Earth*, posted on National Geographic, Extreme weather includes hurricanes, tornadoes, floods, blizzards, hail, etc. (National Geographic Society, 2019).

Hurricanes and tornadoes are violent storms accompanied by heavy wind and rain, which not only cause large area delays for airlines but can pose risk to life and property. In 2017, the Florida Department of Transportation (FDOT) reported that Hurricane Irma caused roughly 6.8 million people, including Florida residents and tourists, to evacuate and head to the safe locations (FDOT, 2018). This evacuation was recorded as the largest evacuation in U.S. history (FEMA, 2018). Hurricane Irma caused many problems and affected people's living and evacuation process. Because of geographical factors and human issues, the fuel supply system in Florida is vulnerable (FDOT, 2018). Fuel shortage was the first problem to be solved in Florida during the hurricane. When the evacuation started, some gas stations in Florida, which started with a full supply of fuel, were empty in as little as two hours (FDOT, 2018). The congestion on the highway caused by the substantial number of people evacuating in vehicles also prevented quick evacuation from the hurricane. Thus, evacuating through an airline could be an option when fuel is limited, and cars are not able to transport people to a safe location.

No one living in the Atlantic Basin is unfamiliar with hurricanes, and everyone has heard and witnessed their devastation to some extent. Barry (2008), Doctor of Philosophy in Atmospheric Science, stated that the number of hurricanes in the Atlantic Basin each season, and the potential damage caused by these storms, have shown a cyclical pattern that changes on the scale of decades in *Hurricanes: A primer for the aviation professional*. Unfortunately, Barry also claimed that hurricane activity has recently increased, and this increase is likely to continue for up to 30 years (Barry, 2008). Of the 258 weather disasters in the U.S. since 1980, tropical cyclones have caused the most damage: $945.9 billion in total, with an average loss of nearly $21.5 billion per event. They also caused the most deaths: 6,593 between 1980 and 2020. Since the losses and casualties caused by hurricanes are so shocking, a model that can predict the trend of hurricanes and help people evacuate is very necessary. As a result, Davidson et al. (2020) built "a new integrated scenario-based evacuation (ISE) framework to support hurricane evacuation decision making".

Davidson et al. (2020) pointed out that as hurricane proceeds, its path, velocity, size, and intensity change over days and thus result in the spatial and temporal variation of wind, storm surge, and rainfall. A hurricane is like a fickle person, and it can change its mind any second. Maybe one second, its path shows that it will end up in the ocean; the next second, it looks like it is going to hit land. As a result, the dynamics, uncertainties of the hurricane activity, and interactions of human and natural systems are the three critical challenges to hurricane evacuation (Davidson et al., 2020). However, even the most widely used HURREVAC systems (FEMA, 2017) do not fully cover these three important challenges, and due to the complexity, rain and wind are not always considered alongside hurricane hazards (Davidson et al., 2020). However, the rain and winds are the most violent effects of a hurricane; rainfall might cause flooding while winds could result in extensive damage. Hurricane-induced flooding, for example, took days to recede from Louis Armstrong International Airport (MSY) in New Orleans after Hurricane Katrina passed (Hurricane Katrina and Aviation, 2005). This is exactly the meaning of Davidson et al.'s research and the advantages of the ISE system.

For a more accurate and up-to-date decision response to hurricanes, Davidson et al. (2020) integrated risk, people behavior, and transportation modeling by using a "multi-stage stochastic programming (MSP) model" and a "tree of evacuation order recommendation" would be built. In the ISE framework, for each hurricane scenario, s, every 15-minute time step time, t, all the latest information is recorded, and this procedure will keep repeated to help the emergency managers to make the correct decision (Davidson et al., 2020). A case study of Hurricane Isabel was introduced and demonstrated that the ISE model could minimize travel risk and time for people who need to be evacuated (Davidson et al., 2020).

The aviation industry also desperately needed a decision-making model in the face of such extreme weather, as Barry (2008) declared that "aircraft are particularly vulnerable to the destructive force of these storms, especially those in the general aviation class." As a result, while considering a hurricane, Davidson et al.'s research favors evacuations in the event of a hurricane, and our project favors the proper reaction and decision-making of the aviation industry, including airports, ATC, airlines, aircraft, and pilots.

In 2005, American Airlines substituted two Boeing 777s and two Boeing 767s for the narrower and smaller aircraft to accommodate the request of additional passengers due to Hurricane Rita (PR Newswire, 2005). The larger aircraft provides approximately 360 additional seats for evacuation compared to narrower body aircraft, which indicates the possibility and reliability of airline evacuation when extreme weather occurs. In 2017, United Airlines also provided “bonus miles, matching funds raised for relief, and added extra flights” for passengers evacuating from Hurricane Irma (Airline Industry Information, 2017). United Airlines flew a Boeing 777 aircraft to San Juan, Puerto Rico, as an extra way to move the people there to a safer location. In both cases, airlines and their flights played a key role in evacuation in natural disasters, which reiterates the importance of air mobility. It is essential to keep air mobility from being affected by extreme weather and disaster, which is one of our project’s goals. Predicting the effect that extreme weather may have on air mobility and acting accordingly would be beneficial for both passengers’ safety and airlines’ economy.

### C. Abnormal Situation

Even though pilots and flight crews are considered largely responsible for the main cognitive burden when managing emergencies, it is also widely believed that the handling of emergencies is influenced by communication and coordination with air traffic controllers (Burian et al., 2005). Air traffic controllers (ATC) play a critical part in all flights. They are responsible for scheduling flights, managing traffic, handling emergencies, and making decisions. In fact, compared to pilots, air traffic controllers’ jobs are more complex and essential to aviation safety since their decisions and instructions would affect multiple aircraft in the sector, not just one.

In the article, “Managing emergencies and abnormal situations in air traffic control (part 1): Taskwork strategies,” Malakis et al. (2010) repeatedly stressed the importance of ATC and the challenges of their work. They stated that ATC is a “complex safety-critical system” with five work characteristics, including “rapidly escalating situations,” “severe time pressure,” “severe error consequences,” “multi-component decisions,” and “conflicting/shifting goals” (Malakis et al., 2010). In other words, air traffic controllers must manage sudden changes in flight conditions with limited thinking and response time, and a small mistake could lead to serious undesired results.

The main goal of this paper is to generate a taxonomy of cognitive strategies in ATC taskwork. Since thinking strategies are closely related to making decisions and solving problems, different decision models were compared and analyzed and resulted in the T2EAM Model, which was based on five thinking strategies: anticipation, recognition, uncertainty management, planning, and workload management (Malakis et al., 2010). It has been claimed that controllers need to recognize signals of impending situations, anticipate outcomes as well as coordinate the plan while seeking available information and managing workload (Malakis et al., 2010). However, controllers are human, and no matter how trained they are, human beings make mistakes. According to *Models of man; social and rational*, (Herbert, 1957), stated that the deviation between human decision-making and rational decision-making models cannot be eliminated. When faced with a problem, people tend to use the first option that meets their needs rather than identifying and evaluating all available options and choosing the best one. However, computer systems do not have this problem; they can identify thousands of options from a database and evaluate them within a second. As a result, it is urgent for the aviation industry to have a decision-making assistance system to help ATC make decisions and thus resolve the emergency and reduce its impact.

Malakis's paper concentrated on en-route controllers, and to generate a taxonomy of cognitive strategies in ATC taskwork, two field studies took place in the paper. In the field studies, emergency and abnormal situations training were observed to collect data (Malakis et al., 2010). However, "the novice controllers did not achieve satisfactory score," and the results showed that there is a big gap between the mean score of novice controllers and the mean score of the expert controllers due to experiences (Malakis et al., 2010). The difference in performance in identifying emergencies and planning accordingly can lead to accidents in real life. This also proves the necessity of our project.

Although Burian and Barshi (2003) reviewed substantial amounts of ASRS and NTSB reports to produce their statements, the way they got their evidence is still somewhat monolithic. On the contrary, our research pays more attention to the diversity and comprehensiveness of the data, including ADS-B Data, Airline On-Time Performance Data

(AOTP), and Quality Controlled Local Climatological Data (QCLCD), Airport Capacity Profiles, and SWIM Records. All these data sources allow our deep learning models to build an environment as close to the real world as possible and thus predict the risks and influence of emergencies on air mobility systems. Moreover, the sample size and the number of people participating in the studies are limited, resulting in reduced reliability of the paper. Furthermore, since the data comes from observation and human performance may change while being observed, the validity of the paper is also reduced. On the contrary, our research pays more attention to the reliability and validity of the data. Millions of real-world data and cases from diverse sources will be examined and thus help our deep learning models to build an environment as close to the real world as possible to make the best decision for the emergency or abnormal situation.

### 1.3 Machine Learning

Machine learning (ML) is "the study of computer algorithms that can improve automatically through experience and data" (Mitchell, 1997). The definition of machine learning has pointed out one of the most significant advantages of machine learning, which is to learn through experience and data automatically and improve the results. Machine learning allows programmers not to have to specify all the steps for the machine learning program to achieve the goal; instead, the program itself can learn from the data provided and figure out the solution for certain tasks. The self-teaching characteristic is so important and helpful, especially in most advanced tasks, including speech recognition, computer vision, etc. In those advanced and complex tasks, it is more effective and realistic for the program to develop its own algorithm than the human designing a complete algorithm and steps for it. Under the wide title of machine learning, there are several types of approaches, and these approaches can be divided into three broad categories: supervised learning, unsupervised learning, and reinforcement learning. The following examples and applications will fall into these three big categories and indicate the wide use of machine learning in performing complex tasks in industries including aviation.

## A. Reinforcement Learning

Reinforcement learning is one of the powerful and effective machine learning training methods that let the agent choose what to do and how to achieve the goal by maximizing the reward at each different state in the environment (Sutton & Barto, 2018). The training agent directly interacts with the environment and improves its solution by learning from the environment. Reinforcement learning has a unique feature that does not require a complete model or a supervisor (Sutton & Barto, 2018). However, the training agent will need an environment to interact with and learn from the results, requiring a specific environment. In our case, the environment will be built using the deep learning technique, which will be discussed later in the chapter.

In the aviation industry, machine learning was widely used because of the self-learning discipline and its effectiveness. Prior research in air mobility using the reinforcement learning method includes solving the air traffic flow management problem. Air Traffic Flow Management (ATFM) generates the optical speed for aircraft. It could count the speed of aircraft and many other factors into airspace capacity consideration to optimize the enroute efficiency and decrease the delay rate (Bertsimas & Patterson, 1998).

With the continuously growing demand for flight and limited airspace capacity, airspace congestion has become an urgent problem that needs to be solved. ATFM is a helpful solution that was utilized to control and manage the air traffic flow to prevent and reduce congestion. With the help of ATFM, researchers have investigated more useful and effective solutions and models using the reinforcement learning approach and other machine learning algorithms. For instance, Crespo et al. (2011) presented a model with the Decision Evaluation and Support Module (MAAD), which is a computational agent that utilizes reinforcement learning. By utilizing MAAD and reinforcement learning, the module managed to provide adequate suggestions for traffic flow under the changing scenario. Thus, the traffic sectors between capacity and demand in the module will not reach an imbalance. In the research, the implementation of the computational agent was achieved by reinforcement learning, which uses Q-learning Algorithm. The author pointed out that there are five measures to control the applicable air capacity and traffic flow, which can be summarized as:

- reordering;
- rerouting;
- and retiming.

The Q-learning Algorithm was used in the decision-making process in ATFM and allowed the computational agent to deal with the flight departures and the related time delay, which holds the aircraft on the ground (Crespo et al., 2011). However, holding too many aircraft on the ground will decrease the efficiency of a certain airport, which will increase the workload for both air traffic controllers and pilots. Fuel consumption is also another concerning problem that could damage the economy of airlines and the environment. The traffic flow control measures are grouped by the amount of time needed for departure, which are 5, 7, 10, 15, 20, and 25 minutes. MAAD is used and suggests adequate traffic flow measures. Q-learning algorithm also successfully converged for the behavior of the computational agent in the complex air space scenario.

### B. Markov Decision Process (MDP)

The basic idea behind reinforcement learning is to model sequential decision-making problems by Markov Decision Process (MDP). MDP is also known as the Markov chain, which has Markovian states and transition probabilities with unobservable states (Mahboubi & Kochenderfer, 2017). In the paper, MDP was defined as "a stochastic process with Markovian states s∈S and transition probabilities sk+1∼T(·|sk), where k is a time index. In certain applications, the state of the system cannot be directly measured, but instead observations o∈ Ω are obtained from a distribution conditional on the current state, i.e., ok ∼ O(·|sk)." Such an MDP system with hidden states is called Hidden Markov Model (HMM). As discussed before, a learning agent will be trained to learn the best solution and action in an environment with policy and states in MDP. There will be multiple numbers of agents in our environment, including airports, aircraft, airlines, air traffic control, etc. The goal for each agent is to maximize the reward or gain.

Researchers have successfully applied HMM in many air traffic management problems to model the time series data—examples include intent prediction, target tracking, and speech recognition (Mahboubi & Kochenderfer, 2017). Lowe and How utilized the MDP approach to build an HMM model as the navigation model, thus predicting the action of human pilots to help the Unmanned Aerial Systems (UAS) to detect, sense, and avoid the intent collisions with other aircraft in the uncontrolled airspace (Lowe & How, 2015). In the paper, the process of predicting the flight path of other aircraft is achieved by following four steps: learn, estimate, predict, and plan, and the author claims that methods including MDP could successfully address the first three elements. The result indicated that the navigation model based on the HMM model could generate a concise and accurate representation of human-controlled aircraft, making the predictions more efficient. The method and simulation Lowe and How used in the research improves the result significantly compared to previous methods in predicting aircraft trajectory (Lowe & How, 2015). Utilizing the advantages of reinforcement learning and the MDP approach, we could avoid building a complex and precise model while keeping the concise and accurate feature in our approach to predict the flight path of aircraft and the potential collisions.

Another paper completed by Spatharis et al. (2019) also illustrated the effectiveness of applying reinforcement learning methods and Markov Decision Process in solving the Demand and Capacity Balance (DCB) problem. DCB describes the balance between capacity and demands, which was also mentioned in the previous machine learning module. DCB is one of the main problems related to air space congestion and could be affected by or cause emergency situations. Different from the MAAD and reinforcement learning approach (Crespo et al., 2011), the agent in this research model represents flights that aims to get the best performance while considering the operational constraints of airspace and preventing congestions.

## C. Artificial Neural Network

For most stakeholders in the aviation industry, time is money. Therefore, the more efficient the operation of airports and aircraft, the greater the benefits. Since operational efficiency is highly dependent on appropriate actions and reactions to the current situation (Schultz et al., 2021), there is an urgent need for a system to analyze current constraints and predict future conditions. Of all constraints, local weather is critical to aircraft and airport operations, and this influence could further impact the whole aviation network (Schultz et al., 2021). Frequent flyers know that if one airport is closed or affected by weather, flights at connecting airports can be delayed or canceled with a ripple effect. Thus, the systematization of prospective weather effects facilitates effective consideration at the tactical level in airport operations (Schultz et al., 2021). Due to the above factors, Schultz et al. (2021) used machine learning methods to quantify the correlation between airport performance degradation and the severity of local weather events, and created a new method that can predict future system states at airports.

By utilizing Artificial Neural Networks (ANNs), the study managed to classify different airport performances. ANN, as a machine learning approach, is a type of math model that simulates the operation and construction of neurons, like a biological brain. ANNs are comprised of one or more hidden layers between the input layer and output layer. Each artificial neuron connects to another neuron and has its associated weight and threshold (Stevens et al., 2020. The connections between artificial neurons will transmit signals to other neurons in the model. If the output of one of more neurons becomes above the threshold value, then the neuron will be activated and send data to the next layer of the model. In this study, Schultz et al. (2021) utilized ANN because of the large amount of data that needs to be analyzed. The ANN serves as an "adaptive intermediate model" to process the data of airport performance and the weather. The advantage of self-learning of ANN helped with the progress, provided more accurate data, and improved the results.

To analyze weather and airport performance, Schultz et al. (2021) utilized different credible data sources, including "flight plan and weather data from major European airports for the year 2013-2015". Moreover, in the analysis, London-Gatwick Airport (LGW) was considered a typical use case, as the airport has been operating at its declared capacity for many years (Schultz et al., 2021). Airport performance is mainly about its capacity under certain conditions, and Schultz et al. (2021) pointed out that LGW operates one of the busiest runways in the world, and only one runway in the airport has an instrument landing system. As a result, deviations from optimal conditions can immediately affect airport performance and make it difficult to recover from these conditions (Schultz et al., 2021). In other words, predicting expected airport performance based on local conditions can significantly improve airport air traffic flow and capacity management (Schultz et al., 2021).

To achieve the goal of providing a pre-alert and decision-making assistant system for passengers and airport staff when emergency situations occur, we must leverage the power of machine learning, which is one of the most essential and helpful tools.

**D. Deep Learning**

Deep learning is a branch of the broad machine learning family that is based on artificial neural networks (ANNs). ANNs, as mentioned in the previous machine learning chapter, are inspired and designed by the information processing of the biological brain. Deep learning can be supervised, unsupervised, or semi-supervised (LeCun et al., 2015). There are many examples and applications of deep learning, including deep neural networks (DNN), deep reinforcement learning, deep belief networks, etc. Deep learning has been applied to many industries and fields, including machine translation, speech recognition, climate, and computer vision.

Rahman and Hasan (2018) introduced a Long Short-Term Memory Neural Network (LSTM-NN) model based on a deep learning algorithm. In the paper, the focus point is on hurricane evacuation and the traffic speed prediction on the freeway. The congestion on the highway caused by large-scale evacuation by vehicle presents a challenge in achieving the goal of quick evacuation from a hurricane. Congestion on the freeway can cause slow and delayed evacuation, which will decrease the effectiveness of evacuation, putting people's lives in danger. Rahman and Hasan (2018) aimed to provide accurate traffic predictions for evacuation in a hurricane. LSTM-NN was applied to predict the short-term vehicle speeds on the highway during the hurricane evacuation period. The author pointed out that the LSTM neural network can be considered a specific type of Recurrent Neural Network (RNN – Figure 3). One of the biggest advantages of using RNN is that it can process sequential data precisely compared to other traditional classifications. Because of the internal memory, RNN can remember important input thus providing precise prediction.

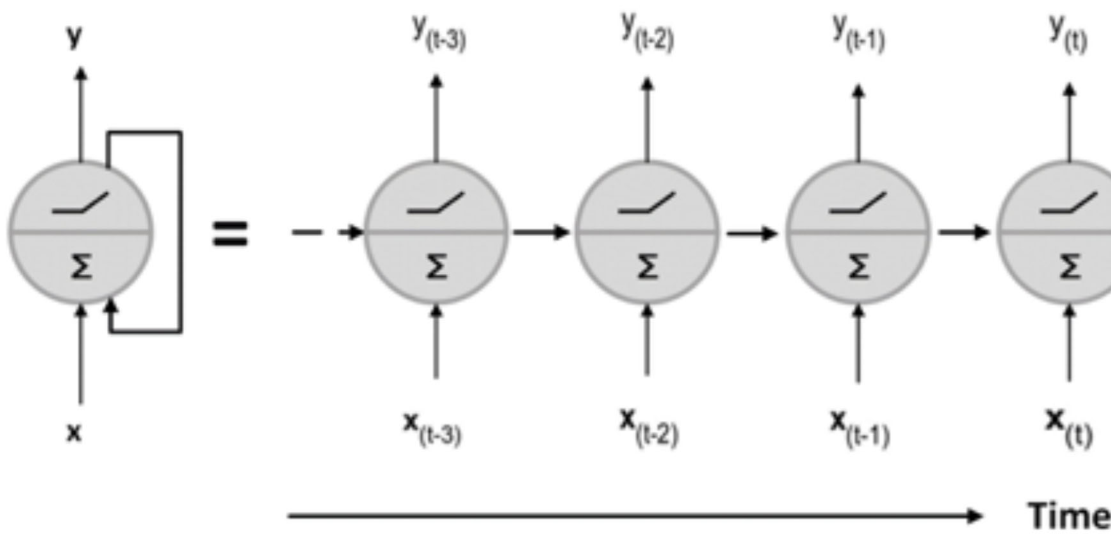



*Figure 3.* A recurrent neuron unrolled process (adopted from Géron, 2018)

Although RNN can capture the nonlinearity and provide precise prediction, it is weak in predicting and learning based on a lengthy period (Hochreiter & Schmidhuber, 1997). Thus, LSTM-NN was introduced as an improved approach that overcame the disadvantages of RNN. LSTM-NN capture long-term dependencies instead of being limited to short-term. Long-term dependencies make the prediction accurate and could determine the optical time lag and window for the time series problems.

**E. Random Forest**

Zhao et al. (2020) investigated a decision-making assistant system for pre-evacuation by using one of the machine learning techniques: random forest (RF). Random forest is an ensemble learning method that constructs a wide range of decision trees at training time for classification and regression (Zhao et al., 2020). Using the random forest method in the study has two significant advantages. RF can reduce the “overfitting problem of the individual decision tree by using bootstrapping.” Also, RF can enhance the accuracy of prediction by capturing the potential interactions in the complicated input data automatically, which is a substantial improvement compared to other traditional statistical models. RF, as a machine learning technique, has generated better predictive performance compared to logistic regression (Zhao et al., 2020). The result of the model and study also proved the advantages of using RF. It successfully captured the interactions and nonlinearities among independent variables and the outcome, which is achieved by its rich behavioral interpretations.

In the study, the authors pointed out that the evacuation time for buildings depends on the building occupants' behavior in the first stage of evacuation. Thus, controlling the behavior or assisting the building owner's decision-making processes becomes important for a successful and effective evacuation. Zhao et al. (2020) focus on the microscopic level of investigations, which is investigating how the building occupants respond to social and environmental factors. By using machine learning, several affecting factors were investigated and discussed. In the study, three modeling approaches were analyzed from the modeling point of view (Zhao et al., 2020). The first approach focused on the pre-defined time that was assigned to the agents at the pre-evacuation level. The second approach focused on the sequences that were assigned to the agents. The third approach predicted the decision-making of agents while considering both the external and internal factors that could affect the response (Zhao et al., 2020). Although the third approach could overcome the weakness of the first two approaches, the author pointed out that the third approach still needed the developers to select the relevant affecting factors and algorithm to simulate agent behavior. The study indicated the potential of using machine learning techniques in investigating the complex evacuation environment and affecting factors, which is helpful in building a decision-making assistant system for evacuation under emergency situations. And the author also pointed out that machine learning-based modeling can be combined and integrated with the agent-based evacuation model. The integration of both models will develop a more realistic and accurate simulation for emergency situations (Zhao et al., 2020).

**F. Multi-agent Deep Reinforcement Learning (MADRL)**

Tang and Xu (2021) utilized the multi-agent reinforcement learning (MARL) method to analyze the air traffic flow management and studied the demand-capacity balancing problem (DCB). By adopting the MARL technique, the intelligent agent in the DCB problem can learn a proper solution through numerous attempts and errors. MARL also allows the programmer to be free from formulating complex models by hand. Researchers observed that the multi-agent reinforcement algorithm perform better than other methods in terms of average aircraft delay time, average flight delay number, and the percentage of delayed flights. (Tang & Xu, 2021) Thus, the result indicated that using a reinforcement learning algorithm is appropriate and accurate for calculating the average delay information and data in our environment.

However, there are still problems that need to be addressed when using the multi-agent deep reinforcement learning method. In a cooperative environment of multiple agents, the policies and instructions for each different agent can be difficult to perform or optimize because of the curse of dimensionality in the environment. On the other hand, the delay between the correlated actions and rewards is unignorable and significant in the cooperative multi-agent environment. Because of the determined character of maximizing the gain for each different agent, there will be the agent who cannot receive their deserved reward and is affected by other agents' favorable actions. Thus, to address the problems of the curse of dimensionality and the unbalanced rewards between multiple agents, the agents should be trained from a macro aspect, which helps the agents have a bigger view of the whole environment and system, thus improving the results and solutions for the multi-agent model.

The research done by Menda et al. (2019) shows the effective method of using reinforcement learning in a cooperative multi-agent environment to solve the difficulties of unbalanced reward and policy optimization. In the study, researchers managed to train the agents through a set of multi-step macro actions. Instead of performing low-level primitive actions, the agents, after training, could perform high-level macro actions such as point-way tracking. Thus, the macro-actions and views performed by each agent in a multi-agent environment can reduce the impact of unbalanced rewards or actions between different agents. On the other hand, Menda et al. (2019) also implemented decentralized policies in their model and environment, which helps each agent make independent decisions based on the situation it is facing and the observation it has of the environment. Agents will follow the decentralized policies asynchronously while performing the macro-actions for a sharing goal.

The macro view and actions for multiple agents and decentralized policies are essential for our model and environment, where we implement macro-actions for each different agent, including airport, aircraft, airline, air traffic control, and passenger. Those five different agents will need to act asynchronously while having the common goal of macro-actions: to maintain and improve air mobility under emergency situations. By utilizing an event-driven method, we are able to implement both macro-actions and decentralized policies for the five different agents in our environment. When an agent achieves a primitive goal or specific event from the completion of macro-actions, it will be approved to choose a new macro-action. This way, the agents can work simultaneously without affecting others while all trying to achieve the macro goal, which is to improve the air mobility in emergency situations.

## 2. PROBLEM STATEMENT AND SOLUTION OVERVIEW

The advent of unforeseen emergencies, including natural disasters and other crisis situations, poses significant challenges to air mobility, impacting both flight operations and passenger safety. The primary problem addressed by this project is the need for robust, efficient, and adaptable air traffic management systems capable of minimizing disruptions and optimizing evacuations during emergencies. Traditional approaches fall short in predicting airport capacity constraints and effectively managing air resources under rapidly changing conditions. This project aims to fill these gaps by leveraging advanced machine learning and optimization techniques.

### 2.1 Problem Statement

Emergency situations necessitate swift and decisive action to ensure the safety and evacuation of affected populations. Air mobility systems, integral to emergency response efforts, must contend with numerous challenges:

- **Predicting Airport Capacity Constraints**: Accurately forecasting how weather and other emergency-related factors affect airport operations, including flight delays and cancellations.
- **Optimizing Resource Allocation**: Efficiently managing limited air resources, such as aircraft and crew, to facilitate the rapid evacuation of civilians from disaster zones.
- **Planning Evacuation Flight Destinations and Schedules**: Developing optimal flight plans that accommodate increased demand for evacuations while minimizing disruptions to the broader air traffic system.

### Solution Overview

The "Improving Air Mobility Under Emergency Situations" project employs a multi-faceted approach to enhance air traffic management during emergencies, leveraging the power of machine learning and optimization algorithms. In Phase I, the project introduces an explainable machine learning model utilizing a Gated Recurrent Unit (GRU) neural network. This model is designed to predict weather-related airport capacity constraints, specifically focusing on flights arriving with Estimated Departure Clearance Time (EDCT) delays. By integrating diverse datasets, including historical weather observations, flight schedules, and delay records, this phase provides critical insights into the factors influencing flight delays. The model's explainability ensures that its predictions are interpretable and actionable, enabling stakeholders to make informed decisions in real-time.

In the subsequent phase, the project shifts its focus towards optimizing the allocation of flight resources for emergency evacuations through a cost-aware approach. Utilizing the Particle Swarm Optimization (PSO) algorithm, Phase II identifies the most efficient strategies for diverting flights to facilitate rapid civilian evacuations while minimizing operational costs. This optimization considers several constraints, including the airport's operational capacity and the availability of flights, to ensure a balanced and effective evacuation strategy. The cost-effectiveness of the selected flights is paramount, ensuring that the evacuation process is not only swift but also economically viable, thereby maximizing the utility of limited resources during critical times.

The final phase of the project introduces a novel hybrid model that combines the strengths of a genetic algorithm (GA) with the speed of a neural network (NN) to plan evacuation flight paths efficiently. This Neural-Network Accelerated Genetic Algorithm approach significantly reduces the computational overhead typically associated with optimization tasks, allowing for the rapid generation of optimal evacuation plans. By training the NN on data from various airports, the model achieves a level of generalization that enables its application across different emergency scenarios and geographical locations. This phase exemplifies the project's commitment to developing scalable, efficient solutions for emergency air mobility management, demonstrating the potential of AI and machine learning to revolutionize how the aviation industry responds to crises.

## 2.2 APPROACH AND METHODOLOGY

This project employed a systematic approach and advanced methodologies to address the challenges of air mobility in emergency scenarios. This section outlines the comprehensive strategies, analytical frameworks, and technical tools utilized across the three phases of the project to achieve its objectives.

**Phase I: Explainable Machine Learning for Flight Delay Prediction**

- **Objective:** To develop a predictive model capable of forecasting weather-related airport capacity constraints, focusing on the number of flights arriving with Estimated Departure Clearance Time (EDCT) delays.
- **Data Collection and Preparation:** The methodology began with the collection and preparation of historical weather observations (METAR), weather forecasts (TAF), flight schedules, and delay data for O'Hare International Airport (ORD) spanning three years. Data cleaning and one-hot encoding techniques were applied to transform categorical attributes into a machine-readable format, combining all data sources on an hourly basis to form the dataset for model training.
- **Model Development:** A two-layer stacked Gated Recurrent Unit (GRU) neural network, augmented with a three-layer feedforward neural network, was proposed. This architecture was chosen for its ability to capture temporal dependencies in the data, essential for accurate delay predictions.
- **Training and Validation:** The model was trained on data from 2015 to 2017, validated on 2018 data, and tested using 2019 data. Performance metrics focused on the model's ability to predict the count of incoming EDCT-delayed flights accurately.

**Phase II: A Cost-Aware Approach for Flight Resources Aggregation During Pre-Disaster Evacuation**

- **Objective:** To optimize the allocation and utilization of flight resources for emergency evacuations, minimizing costs while ensuring efficient passenger evacuation.
- **Optimization Framework:** The Particle Swarm Optimization (PSO) algorithm was selected for its effectiveness in handling complex optimization problems. The algorithm was tasked with identifying the most cost-effective flight diversions to maximize evacuation efficiency from Daytona Beach International Airport (DAB).
- **Simulation and Evaluation:** A simulated environment was created to model the evacuation scenario, incorporating constraints such as airport operational capacity and the availability of flights. The PSO algorithm's performance was evaluated based on its ability to select optimal flights for evacuation within the specified constraints.

**Phase III: Neural-Network Accelerated Genetic Algorithm for Optimal Evacuation Flight Planning**

- **Objective**: To leverage a genetic algorithm (GA) accelerated by a neural network (NN) for developing efficient evacuation flight plans that minimally impact routine airspace operations.
- **Hybrid Model Development**: This phase introduced a novel hybrid model combining GA for optimization and NN for rapid convergence. The NN was employed to predict potential solutions' fitness, thereby accelerating the GA's search for optimal evacuation plans.
- **Training and Implementation:** The NN model was trained on data from various airports to ensure generalizability. The hybrid model was then applied to generate evacuation flight plans, demonstrating improved efficiency and reduced computational overhead compared to traditional GA.

# 3. PHASE-I EXPLAINABLE MACHINE LEARNING FOR FLIGHT DELAY PREDICTION

## 3.1 Introduction

With continuously growing air traffic, airspace administrators and airspace users are facing the increasingly complex challenge of safely maximizing the throughput of busy airports. The United States National Airspace System (NAS) is considered one of the busiest and most complex airspace systems in the world (MITRE, 2023). Many airports and sectors in the NAS already operate at their capacity limits (Airoldo, 2022). Weather is a common factor that slows down airport operations which temporarily decreases airport capacity (Dhal, et al., 2014). A majority of the NAS delays are caused by weather (Bureau of Transportation Statistics, 2023) Anticipating airport capacity constraints such as weather before they occur can help airlines and airports to alleviate the airport capacity problem at an operational level by allowing Air Traffic Management (ATM) experts as well as airline and airport staff to plan ahead and optimize the use of the available, albeit reduced, capacity.

Airport capacity is generally quantified as the Airport Arrival Rate (AAR) and Airport Departure Rate (ADR) or the sum of both values (Dhal, et al., 2014). The AAR and ADR are set by Air Traffic Control (ATC) as target values for the maximum number of incoming and outgoing flights per hour. Both AAR and ADR are adjusted based on traffic and weather conditions. When the demand for incoming and outgoing flights exceeds the available airport capacity, ATC uses Traffic Management Initiatives (TMIs) to reduce traffic to the target AAR. ATC then asks incoming flights to delay their departure by assigning an Expected Departure Clearance Time (EDCT) to each of these flights (Federal Aviation Administration, 2009). While AAR/ADR and EDCT delays are related, AAR/ADR is a target value that can involve human bias whereas the number of EDCT delays directly measures the capacity constraint's impact.

There is limited research on machine learning and deep learning approaches to predict weather-related airport capacity constraints. Nevertheless, there are related studies on predicting airport runway configurations, AAR, and ADR (Wang, 2019; Smith, 2008; Avery & Balakrishnan, 2015). Other researchers have developed simple models to predict an airport's EDCTs. However, a limited number of features, with respect to wind and weather forecast, has been used (Misra et al., 2022).

The goal of this part of work is to develop a predictive model to forecast airport capacity constraints. Based on a comprehensive feature set, this research can build the foundation for various decision-support tools used by ATM, airlines, and airports. Pertinent attributes of such a feature set include observations and forecasts of general weather phenomena, visibility, winds, temperatures, cloud types, cloud ceilings, as well as the current and expected traffic situation at the airport. These attributes are found in METeorological Aerodrome Reports (METARs), Terminal Aerodrome Forecasts (TAFs), and available traffic data. Possible target variables include the number of EDCT-delayed flights, the ratio of EDCT-delayed flights and scheduled flights, and the average EDCT-delay minutes. When deployed at airlines, the model shall provide ATM experts and airline employees concerned with the day-to-day operations with an early outlook on possible capacity constraints.

The research will demonstrate that commonly used aviation data can be used to create an early warning tool or decision support tool for ATM, airline, and airport staff to plan for weather-related airport capacity problems. This has the potential to reduce service delays by facilitating proactive planning of individual flights and traffic flows through the use of the expected capacity constraint predictions.

To achieve this goal, a comprehensive feature set will be prepared. This feature set includes weather observations (METAR), weather forecasts (TAF), and air traffic data on an hourly basis. The data will be prepared for one airport, Chicago O'Hare International Airport (ORD). It will consist of at least four years of observations. A Recurrent Neural Network (RNN) with Gated Recurrent Units (GRU) will be implemented to predict the number of EDCT flights per hour (or a similar metric). In addition, an explainable Artificial Intelligence (AI) method, occlusion sensitivity (Rojat, Puget, & Diaz-Rodriguez, 2021) is used to describe the behavior of the model.

### 3.2 Data Collection and Preparation

Four different data sources are used to prepare a data set for the airport capacity constraint model. These data sources are the FAA Aviation System Performance Metrics (ASPM) Delayed Flights report (Federal Aviation Administration, 2023a), the FAA ASPM EDCT report (Federal Aviation Administration, 2023b), METARs, and TAFs. All four data sources are publicly available. These data sources were selected for this project because of their real-time availability to airlines. While airlines already apply both METARs and TAFs in real-time, the data points from both ASPM reports can be approximated using FAA-provided streaming interfaces. The data covers the period from January 2015 to December 2019 on an hourly basis at Chicago O'Hare International Airport (ORD). This section gives an overview of the four different data sources followed by a description of applied data preparation steps.

The Delayed Flights report data has two categories, Airport and City Pair Analysis. The Airport Analysis category documents information on the "number and percentage of flight departures and arrivals delayed 15 or more minutes for a selected airport or group of airports" while the City Pair Analysis documents information on the "number and percentage of flight departures and arrivals delayed 15 or more minutes, between two selected airports or groups of airports" and each compares their respective data to the actual schedule and flight plan times (Federal Aviation Administration, 2023a & 2023c). It is important to note that this report uses the Airport Analysis data and that the ASPM Airport Analysis module tracks information for only 77 airports in the U.S. NAS. These airports are referred to as the ASPM 77 (Federal Aviation Administration, 2023d). The data features of the Delayed Flights Airport Analysis report are described in Table I.

The ASPM Airport Analysis EDCT report provides information on the number of flight arrivals and departures with EDCT delays at given airports that instruct the aircraft to "remain at the departure airport until a specified time is issued when weather, congestion, or other problems in route or at the arrival airport will impede flights from arriving at the destination airport as originally planned" (Federal Aviation Administration, 2023b). When EDCTs are issued, they are accompanied by traffic management initiatives that the FAA Command Center implements (Federal Aviation Administration, 2023b). This report uses multiple features from the EDCT report. These are described in Table I. Some of the key features from the EDCT data that determine the success of the prediction model investigated by this report include the Arrivals with EDCT which provides information on the total number of arrival flights with EDCT departure delay which is computed by comparing the EDCT Wheels Off Time to the Flight Plan Wheels Off Time (Federal Aviation Administration, 2023e). A key factor to note about this feature is that there are times when a "flight may have been issued an EDCT and not accumulated any EDCT delay if the Flight Plan Wheels Off time was after the EDCT Wheels Off time" (Federal Aviation Administration, 2023e). For example, if a Flight Plan Wheels Off time is planned for 1600 UTC and the EDCT Wheels Off time is issued for 1530 UTC since the EDCT Wheels Off time occurs before the EDCT Wheels Off time, the flight does not accrue an EDCT delay, therefore, making an issued EDCT Wheels Off time like this obsolete. However, if the EDCT Wheels Off time is scheduled after the Flight Plan Wheels Off time, an EDCT delay is accrued. Furthermore, the Arrivals For Metric Computation feature is another key attribute of the EDCT report data. It calculates metrics in ASPM that are aggregated based on the scheduled arrival time if available or the flight plan arrival time if the scheduled arrival time is not available. In addition, the Arrivals For Metrics Computation feature only includes flight information for itinerant flights to or from one of the ASPM 77 Airports or operated by one of the ASPM Carriers at any airport with flight plans or actual arrival and departure times. The Average EDCT for All Arrivals is also another important feature that provides information on the average EDCT delay at the departure airport for all the arrivals at any of the ASPM 77 airports. Finally, the Average EDCT for Arrivals Where EDCT greater than 0 is another key feature in the prediction model that documents the average EDCT delay in minutes at the departure airport for flights arriving with an EDCT delay.

A METAR is an aviation routine weather report that includes the type of report, "[station] identifier, time of observation, wind, visibility, Runway Visual Range (RVR), present weather phenomena, sky conditions, temperature, dewpoint, and altimeter setting" for a given aerodrome (Federal Aviation Administration, 2022). A METAR's format has two major sections, the body which has a maximum of eleven groups, and the remarks which consist of two categories: the first is Automated, Manual, and Plain Language Remark and the second is Additive and Maintenance Data (Federal Aviation Administration, 2022). In a

METAR, the airport identifier follows the International Civil Aviation Organization (ICAO) format to identify the airport that the METAR has been issued for, the time observation is in Coordinated Universal Time (UTC) time and includes the date and time that the METAR report is issued: the first two digits report the day of the month, followed by the hour and then the minutes, the wind category has the first three digits indicating the wind direction in degrees followed by two or three digits indicating wind velocity in knots, the visibility group is a measure of the atmosphere's opacity and uses Statue Miles (SM) to measure surface visibility, the RVR provides the runway and horizontal distance in Feet (FT) that a pilot can see down that runway, and lastly, the present weather group provides the current weather conditions at the given airport such as precipitation, sky conditions, etc. METARS are primarily issued by the FAAs automated weather observing systems and are routinely updated at the top of every hour hence they are at times referred to as "hourly" weather reports (Federal Aviation Administration, 2023f). The main target audience for METARs includes pilots, ATC, and meteorologists. Figure 4 is an image that provides a visual representation of the general layout of a METAR and what each of the categories corresponds to with respect to the aerodrome and its current weather.

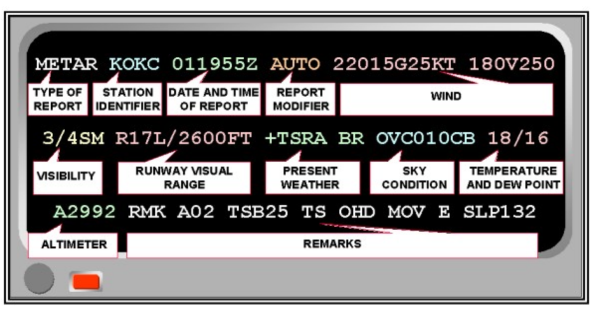



*Figure 4.* Image showing the Coding Format of a METAR obtained from the Aviation Weather Handbook (Federal Aviation Administration, 2022).


TAF
KORD 011729Z 0118/0224 23016G24KT P6SM SCT100 SCT250
FM020000 26012KT P6SM SCT060 BKN080 FM020800 31005KT P6SM
FEW030 SCT060 SCT250
FM021600 35003KT P6SM BKN250

*Figure 5.* Example of a TAF for Chicago O'Hare International Airport (KORD) January 01, 2015, at 1729 Zulu time.

Different than METARs, TAFs describe weather forecasts for a given airport during a specified time within 24 hours, as demonstrated in Figure 5. TAFs issue weather forecasts for active airports in the National Plan of Integrated Airport Systems (NPIAS) "including FAA-towered airports, Federal contract-towered airports, non-federal towered airports, and non-towered airports" (Federal Aviation Administration, 2023g). A TAF report includes the following elements: the type of report, the ICAO station identifier, the date and time the TAF is issued, the validity period date and time of the TAF, and the forecast weather meteorological conditions (Glazer, 2023). Since expected weather situations can change during a TAF's validity period, a TAF is split up into multiple forecasts. These forecasts are sometimes also referred to as lines in a TAF. Each line has a validity period. For example, FM011600 means that the following TAF line is valid from the first day of the month at 1600 UTC and valid until the time indicated in the next line break or the end of the TAF's validity period. To combine the four data sources into one data set, first, the ASPM Delayed Flights report and the ASPM EDCT report are merged. Both reports are on an hourly level, which allows for them to be merged based on matching dates and local time. The date column in both reports is based on local time. To allow for merging with METARs and TAFs based on UTC time, the date in both columns is converted to a UTC-based date. The result of these operations is a data set with delay and EDCT metrics in UTC time. A day-of-week attribute is created to allow the model to identify potential fluctuations throughout the week. In a separate step, METARs are converted into a tabular format such that every data point in a METAR is placed into a dedicated column. Similar operations are performed for the TAFs. A TAF contains several forecasts which are also referred to as lines. The forecasts share an identical structure. Each new forecast line in a TAF is placed into a new group of columns with each group of columns dedicated to the respective forecast line. Characters that are not categorical data but metadata such "SM" in the visibility column are removed, and where possible, text strings are converted to numerical data. For example, the string "1 1/2SM" which describes a visibility of 1.5 statute miles is converted to the value "1.5". Another example of this operation is the string "M21/M26" which describes a temperature of -21 degrees and a dew point of -26 degrees. This string is transformed into two separate columns for temperature and dew point with numerical values -21 and -26, respectively.

Both METAR and TAF data are then merged with the ASPM reports such that each METAR and TAF is linked to the corresponding hour in which they were published [1]. Some hours will exhibit multiple METAR and/or TAF publications, resulting in more than one row for such hours. Only the latest METAR and TAF are kept in these cases.

The model is trained to predict EDCTs in the upcoming one hour. Corresponding target attributes are created by shifting arrivals_with_edct, arrivals_with_edct_ind, avg_edct_for_all_arrivals, and avg_edct_for_arrivals_where_edct_greater_than_0 by one hour. This creates, for example columns arrivals_with_edct_1, as possible target variables for predicting the number of arrivals in the next one to four hours. The same shifts are applied to the number of scheduled departures and scheduled arrivals which can be used as additional feature attributes. An overview of the resulting data set can be seen in Table I.

In addition to these steps, one-hot encoding is applied for categorical attributes. This includes the cloud descriptions, such as BKN (broken clouds) and SKC (sky clear), the weekday, and distinct types of TAF forecast line breaks. All timestamps are transformed to the difference in minutes to the baseline hour timestamp. For instance, a TAF that was published between 9 and 10 am UTC with the line break FM011200 would have a validity time of 180 (minutes) for the particular forecast line. All continuous attributes are standardized to a distribution with a mean of 0. The target attribute is scaled to values between 0 and 1.

The data set covers five years from January 2015 to December 2019 with a total number of 43,800 observations. The sizes of the training, testing, and validation set are 60%, 20%, and 20%, respectively. Consequently, the years 2015, 2016, and 2017 are used for training. The year 2018 is used for validation and the year 2019 is used for testing.

Due to the availability of EDCT data in the feature set, there is a risk that the trained model makes a prediction for the next hours of the EDCT values solely based on the current EDCT and delay situation. To test the model's ability to predict the EDCT situation without knowledge of EDCTs in prior observations, the model is trained on two different feature sets.

[1] In SQL lingo, this operation would be described as a join of ASPM reports and METAR on airport, year, month, day, and hour with a second join of the resulting table and TAF on airport, year, month, day, and hour.

The first feature set entails the number of scheduled departures and arrivals as well as METAR and TAF attributes. The second feature set entails all features from the first feature set and attributes that describe the current traffic and EDCT situation, for example, delayed gate arrivals and the number of arrivals with EDCT.

### 3.3 Descriptive Data Analysis

The following section provides a detailed breakdown of the feature sets used to run the model predictions with data collected between January 1, 2015, to December 31, 2019. Two feature sets were created to train the model: the first feature set includes the schedule and weather data while the second includes the schedule, weather, and current traffic situation data. Table I below shows the list of features per column that were used and their respective descriptions. Each feature set has a total of 8,730 data entries (rows). As for the columns, the first feature set has a total of 308 columns and the second has a total of 323 columns. Over the five years, a total of 3,950 TAFs were issued with METARs being issued approximately every hour of each day.

TABLE I. Description of input variables. Metadata that is not part of the feature set is marked in italic. The variables in this table are presented in their original form for improved interpretability, while categorical variables that are later encoded using one-hot encoding are denoted with an asterisk (*). Variables from TAF lines are marked with the suffix {n} with n being a number between 1 and 10, representing the nth forecast in the TAF. Variables with the suffix {m} represent the corresponding value in m hours after the baseline observation, with m being a number between 1 and 4.

| Source | Variable | Description |
|---|---|---|
| | *airport* | Airport code |
| | *year* | Year of the row's baseline time |
| | *month* | Month of the row's baseline time |
| | *day* | Day of the row's baseline time |
| | hour | Hour of the row's baseline time with values between 0 and 23 |
| | day_of_week | Weekday from 0 to 6 extracted from the baseline time |
| ASPM Delayed Flights | scheduled_departures | Number of scheduled departures |
| ASPM Delayed Flights | departures_for_metric_computation | Number of departures used by ASPM for metric computation |
| ASPM Delayed Flights | delayed_gate_departures | Number of delayed gate departures (more than 15 minutes) |
| ASPM Delayed Flights | percent_delayed_gate_departures | Percent of delayed gate departures (more than 15 minutes) |
| ASPM Delayed Flights | average_minutes_of_delay_per_delayed_gate_departure | Average minutes of delay per delayed gate departure |
| ASPM Delayed Flights | scheduled_arrivals | Number of scheduled arrivals |
| ASPM Delayed Flights | arrivals_for_metric_computation | Number of arrivals used by ASPM for metric computation |
| ASPM Delayed Flights | delayed_gate_arrivals | Number of delayed gate arrivals (more than 15 minutes) |
| ASPM Delayed Flights | percent_delayed_gate_arrivals | Percent of delayed gate arrivals (more than 15 minutes) |
| ASPM Delayed Flights | average_minutes_of_delay_per_delayed_gate_arrival | Average minutes of delay per delayed gate departure |
| ASPM EDCT report | arrivals_with_edct | Number of arrivals with EDCT delay. Flight that were issued an EDCT but arrived on time are excluded. |
| ASPM EDCT report | arrivals_with_edct_ind | 1 if arrivals_with_edct is greater than 0, 0 otherwise. |
| ASPM EDCT report | percent_of_arrivals_with_edct | Percent of arrivals with EDCT delay. Flight that were issued an EDCT but arrived on time are excluded. |
| ASPM EDCT report | avg_edct_for_all_arrivals | Average EDCT delay for all arrivals at their departure airport. |
| ASPM EDCT report | avg_edct_for_arrivals_where_edct_greater_than_0 | Average EDCT delay of flights that accumulated actual delays at their departure airport. |
| ASPM EDCT report | departures_with_edct | Number of departures with EDCT delay. Flight that were issued an EDCT but departed on time are excluded. |
| ASPM EDCT report | percent_of_departures_with_edct | Percent of departures with EDCT delay. Flight that were issued an EDCT but departed on time are excluded. |
| ASPM EDCT report | avg_edct_for_all_departures | Average EDCT delay for all departures at their departure airport. |
| ASPM EDCT report | avg_edct_for_departures_where_edct_greater_than_0 | Average EDCT delay of flights that accumulated actual delays at their departure airport. |
| METAR | m_issue_time | METAR issue time from the datetime element. Encoded as difference in minutes to the row's baseline time. |

*continued on next page*

continued from previous page

| Source | Variable | Description |
| --- | --- | --- |
| METAR | m_modifier | 1 if COR, 0 if AUTO |
| METAR | m_wind_direction | Wind direction with values between 1 and 360, 0 if not available |
| METAR | m_wind_speed | Wind speed, 0 if not available |
| METAR | m_wind_gusts | Wind gusts, 0 if not available |
| METAR | m_wind_var_1 | First value of variable wind direction with values between 1 and 360, 0 if not available |
| METAR | m_wind_var_2 | Second value of variable wind direction with values between 1 and 360, 0 if not available |
| METAR | m_visibility | Visibility string converted to a floating point number such as 1.5 |
| METAR | * m_weather | Weather qualifier and weather phenomena (see https://www.faa.gov/sites/faa.gov/files/FAA-H-8083-28_FAA_Web.pdf for a list of possible values) |
| METAR | * m_clouds_conditionA | Cloud contraction of the first cloud layer |
| METAR | m_clouds_aglA | Height of the first cloud layer |
| METAR | m_clouds_typeA | 1 if first cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| METAR | * m_clouds_conditionB | Cloud contraction of the second cloud layer |
| METAR | m_clouds_aglB | Height of the second cloud layer |
| METAR | m_clouds_typeB | 1 if second cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| METAR | * m_clouds_conditionC | Cloud contraction of the third cloud layer |
| METAR | m_clouds_aglC | Height of the third cloud layer |
| METAR | m_clouds_typeC | 1 if third cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| METAR | * m_clouds_conditionD | Cloud contraction of the fourth cloud layer |
| METAR | m_clouds_aglD | Height of the fourth cloud layer |
| METAR | m_clouds_typeD | 1 if fourth cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| METAR | * m_clouds_conditionE | Cloud contraction of the fifth cloud layer |
| METAR | m_clouds_aglE | Height of the fifth cloud layer |
| METAR | * m_clouds_typeE | 1 if fifth cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| METAR | m_temperature | Temperature in degrees Celsius, which can be a positive or negative number. Value from preceding observation is used for missing entries |
| METAR | m_dew_point | Dew point in degrees Celsius, which can be a positive or negative number. Value from preceding observation is used for missing entries |

continued on next page

| *continued from previous page* | | |
|---|---|---|
| Source | Variable | Description |
| TAF | t_issue_time | TAF issue time from the datetime element. Encoded as difference in minutes to the row's baseline time. |
| TAF | t_amd_ind | 1 if AMD, 0 otherwise |
| TAF | t_taf_valid_from | Start time of the TAF's validity period. Encoded as difference in minutes to the row's baseline time. |
| TAF | t_taf_valid_to | End time of the TAF's validity period. Encoded as difference in minutes to the row's baseline time. |
| TAF | t_wind_direction_{n} | Wind direction with values between 1 and 360, 0 if not available |
| TAF | t_wind_speed_{n} | Wind speed, 0 if not available |
| TAF | t_wind_gusts_{n} | Wind gusts, 0 if not available |
| TAF | t_visibility_{n} | Visibility string converted to a floating point number such as 1.5 |
| TAF | * t_weather_{n} | Weather qualifier and weather phenomena (see https://www.faa.gov/sites/faa.gov/files/FAA-H-8083-28_FAA_Web.pdf for a list of possible values) |
| TAF | * t_clouds_conditionA_{n} | Cloud contraction of the first cloud layer |
| TAF | t_clouds_aglA_{n} | Height of the first cloud layer |
| TAF | t_clouds_typeA_{n} | 1 if first cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| TAF | * t_clouds_conditionB_{n} | Cloud contraction of the second cloud layer |
| TAF | t_clouds_aglB_{n} | Height of the second cloud layer |
| TAF | t_clouds_typeB_{n} | 1 if second cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| TAF | * t_clouds_conditionC_{n} | Cloud contraction of the third cloud layer |
| TAF | t_clouds_aglC_{n} | Height of the third cloud layer |
| TAF | t_clouds_typeC_{n} | 1 if third cloud layer has cumulonimbus (CB) or towering cumulus (TCU) clouds, 0 otherwise |
| TAF | t_wind_shear_agl_{n} | Wind shear altitude |
| TAF | t_wind_shear_direction_{n} | Wind shear direction with values between 1 and 360, 0 if not available |
| TAF | t_wind_shear_speed_{n} | Wind speed, 0 if not available |
| TAF | * t_line_{n+1} | TAF line break that starts a new forecast line (possible values are PROB, FM, BECMG, and TEMPO) |
| TAF | t_line_valid_from_{n+1} | Start time of the forecast's validity period. Encoded as difference in minutes to the row's baseline time. |
| TAF | t_line_valid_to_{n+1} | End time of the forecast's validity period. Encoded as difference in minutes to the row's baseline time. 0 for FM line breaks |
| | | *continued on next page* |

*continued from previous page*

| Source | Variable | Description |
|---|---|---|
| TAF | ... | ... |
| ASPM Delayed Flights | scheduled_departures_{m} | Scheduled departures in m hours |
| ASPM Delayed Flights | scheduled_arrivals_{m} | Scheduled arrivals in m hours |
| ASPM EDCT report | avg_edct_for_arrivals_where_edct_greater_than_0_-{m} | Target Variable: Average EDCT delay of flights that accumulated actual delays at their departure airport |
| ASPM ... | ... | ... |
| ASPM EDCT report | arrivals_with_edct_{m} | Target Variable: Number of arrivals with EDCT delay in m hours |

A descriptive data analysis of the collected data with respect to EDCTs was conducted to better understand the correlation between features as well as their contribution to the model's predictions. Tables and graphs were applied to provide a visual representation of this analysis. To begin with, the total number of EDCT days per year was determined. Figure 6 shows these results. It is noted that 2019 had a total of 120 EDCT days which was the highest number for all five years.

| Year | Number of EDCT Days per Year |
|---|---|
| 2015 | 95 |
| 2016 | 99 |
| 2017 | 92 |
| 2018 | 119 |
| 2019 | 120 |
| **Average** | **105** |

*Figure 6.* Total Number of EDCT Days per Year

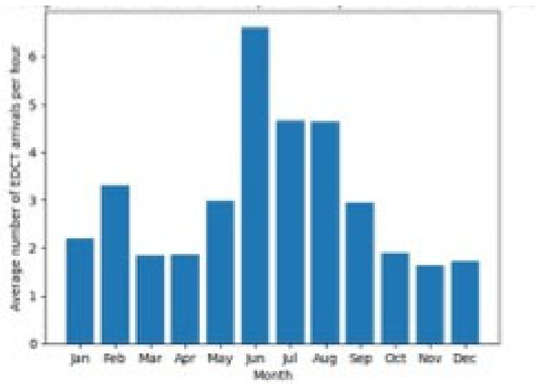



*Figure 7.* Average Number of EDCT Arrivals per Hour by Month from 2015 to 2019

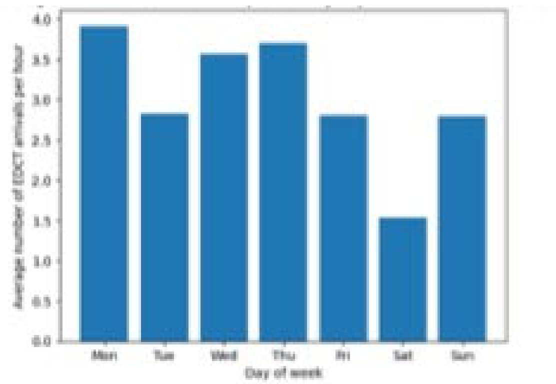



*Figure 8.* Average Number of EDCT Arrivals per Hour by Day of the Week from 2015 to 2019

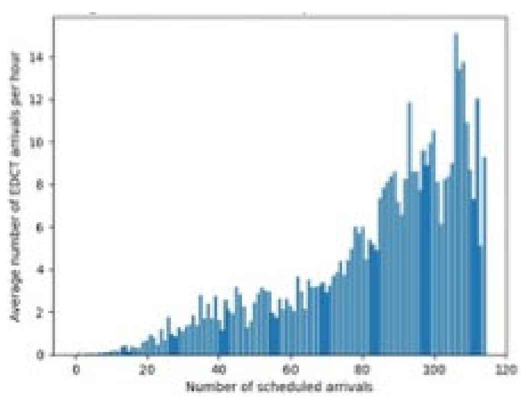



*Figure 9.* Average Number of EDCT Arrivals per Hour by Scheduled Departures from 2015 to 2019

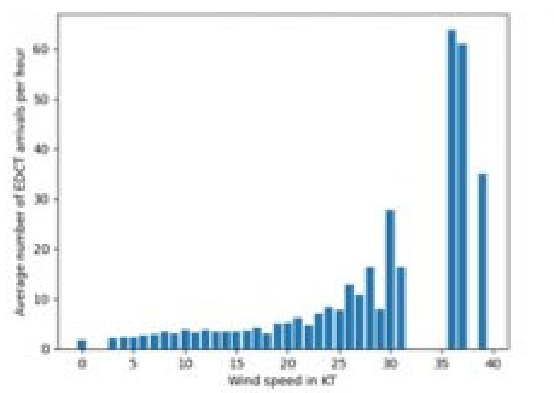


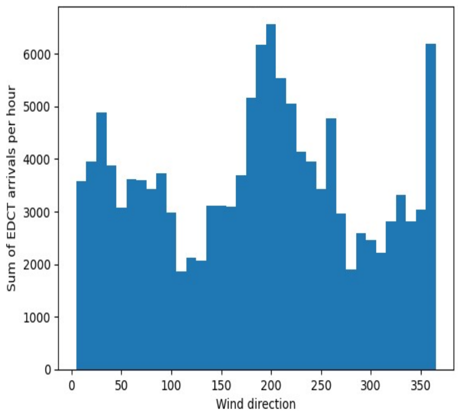

*Figure 10.* Average Number of EDCT Arrivals per Hour based on Wind Speed 2015 to 2019

Figure 7 shows the average number of EDCT arrivals per hour for each month. Per year, June consistently had the highest average of approximately 6.8 EDCT arrivals per hour which correlates to the fact that it is typically the peak air travel month of the year in the U.S. The average number of EDCT arrivals per hour by day of the week was determined to understand which day of the week was typically the busiest in terms of EDCT arrivals. Figure 8 shows that Mondays had the highest average of more than 3.5 EDCT arrivals per hour.

In addition, the average number of EDCT arrivals per hour with respect to the number of scheduled departures was also investigated. As expected, a positive correlation can be observed. As the number of scheduled departures increases, the average number of EDCT arrivals per hour increases as in Figure 9. There were approximately 118 scheduled departures with 15 EDCT arrivals per hour. Furthermore, the average number of EDCT arrivals per hour based on weather elements such as wind speed (KT), wind direction relative to true north, and visibility (SM) was explored. Based on the results, it was observed that wind speed positively correlates with the average number of EDCT arrivals. Figure 10 shows that an increase in wind speed led to an increase in the average number of EDCT arrivals per hour. At 36 knots, there was an average of 65 EDCT arrivals per hour at ORD.

Given that the runways at airports are built to align with the direction of wind patterns specific to an airport and its location, wind direction plays a huge role in how it affects flight operations and EDCTs. During landing, headwinds are conducive because they help reduce the speed of the aircraft which in turn reduces the landing distance as well as the use of fuel when landing. Figure 11 shows the sum of EDCT arrivals per hour with respect to the wind direction. The lowest sum of EDCTs per hour occurs at a wind direction of 270 or 280 degrees which aligns with the direction of most ORD runaways, as in Figure 12. Figure 13 shows the layout of the ORD airport airfield as of 2021.

While this data analysis was conducted solely based on METARs, i.e., weather observations, and the current number of EDCT flights, it demonstrates how different weather descriptors can be used to predict airport capacity. To further boost predictability, the data set used to train the model contains current weather observations, weather forecasts, schedule data, and current delay and EDCT data.



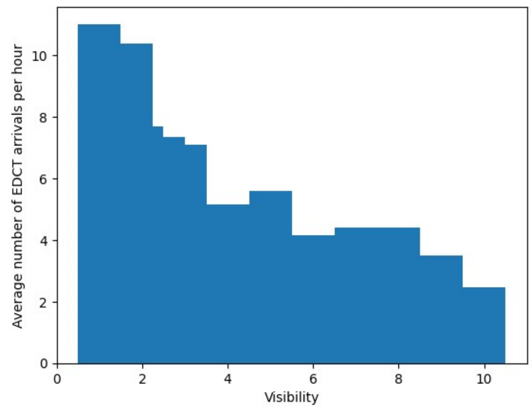




*Figure 11.* Sum of EDCT Arrivals per Hour as per Wind Direction from 2015 to 2019

*Figure 12.* Average Number of EDCT Arrivals per Hour based on Visibility from 2015 to 2019

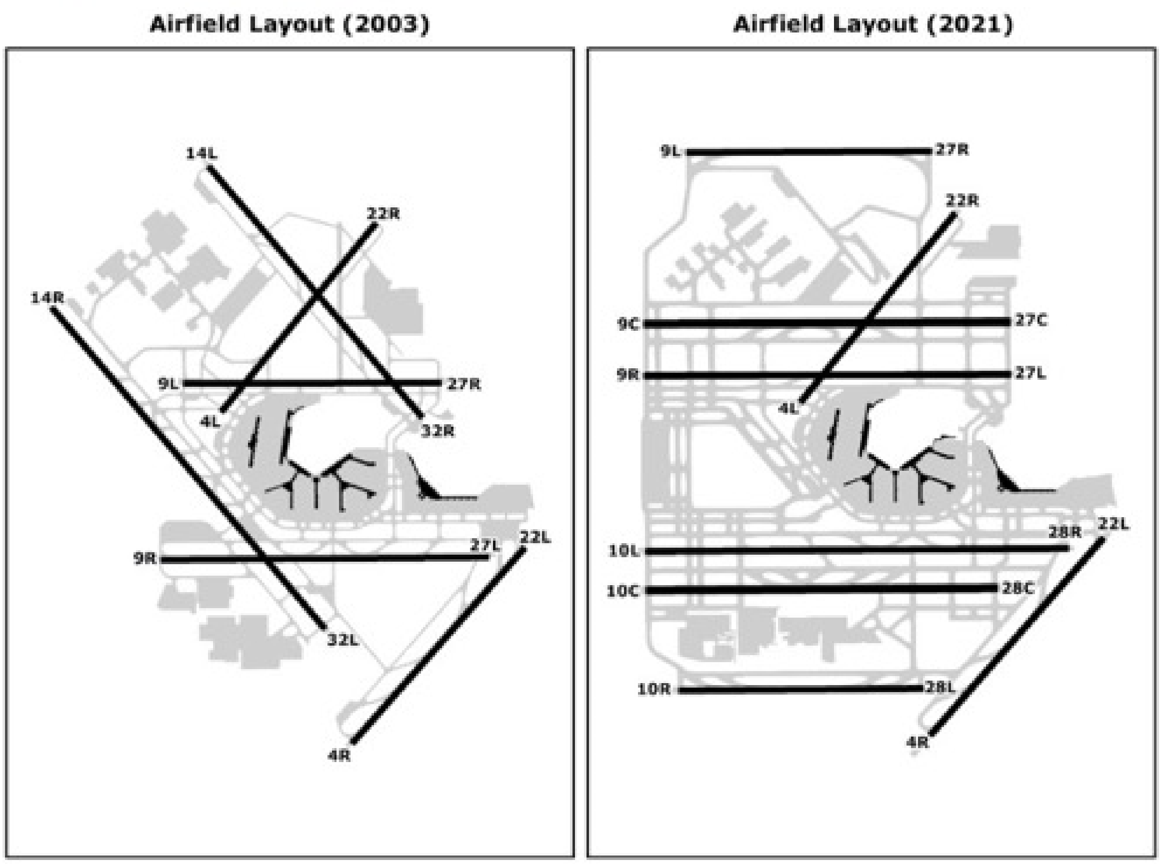



*Figure 13.* Chicago O'Hare International Airport Airfield Layout as of 2021 (Airport Operations 101, 2024).

## 3.4 Base LSTM Model and Performance Analysis

Pertinent characteristics in time-series data can be described by single observations but also by changes in values between multiple observations. For example, sudden changes in weather or wind from one hour to the next may have direct implications on airport operations different from gradual changes over a longer time. This implies that observations and predictions from previous hours can entail crucial information for the next hour's EDCT prediction.

To predict EDCTs for the next hours, a multi-layer model is used. First, the input consisting of 308 or 323 input variables is passed through a layer of two stacked GRUs. The output of the stacked-GRU layer, also referred to as the hidden state, is of dimension 128. The values from the hidden state are then passed through three fully-connected layers with output dimensions of 64, 32, and 1 for the first, second, and third fully-connected layers, respectively. A Rectified Linear Unit (ReLU) is used following every fully-connected layer to convert possible negative values to zero. For both the stacked-GRU layer and the fully connected layers, a dropout probability of 0.8 is used. This randomly removes 80% of the outputs during training to prevent overfitting and allow for a more robust model. Figure 14 is the learning curve of the model. It provides the model's performance on learning the data going through it and its ability to make predictions from the given data. To work with the model's best performance, only eight epochs were applied. Although the validation loss graph decreases gradually and then begins to increase toward the end of the epoch cycle, the model's performance improves on the training data over time.

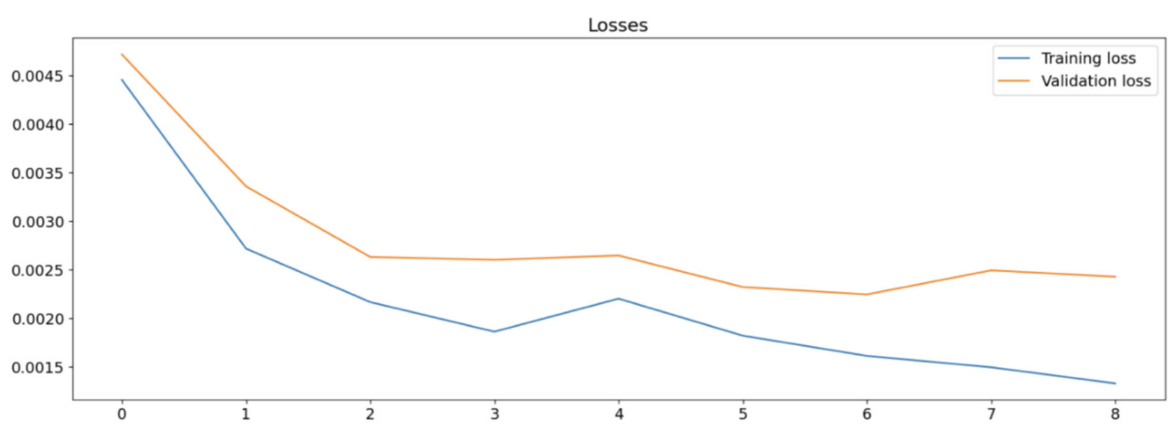



*Figure 14.* Model Training and Validation Loss

Both the ASPM source data sets show that EDCTs and a high number of delays can often be observed over the course of several hours. EDCTs are likely to be issued and delays are likely to occur in hour $h$ when the airport was already congested in hour $h$-$1$. To test the model's ability to predict EDCTs solely based on weather data, two feature sets are created. The base feature set only contains TAF and METAR data as well as data on scheduled departures and arrivals, i.e., the columns scheduled_departures, scheduled_arrivals, scheduled_departures_{m}, and scheduled_arrivals_{m}, (see also Table I). The full feature set contains all TAFs, METARs, and describes the current delay and EDCT situation. The model is trained and tested on both feature sets to predict the next 1 hours. For each of these scenarios, the model is trained on data from 2015 to 2017, validated on data from 2018, and tested on data from 2019.

The results for each of these tests can be seen in Figures 12 and 13. It shows that the model performs better when data on current delays and EDCTs is known. However, the results also show that the model does predict EDCT trends without knowing the current EDCT or delay situation. For example, Figure 11 shows that the model predicted EDCTs during times of congestion and predicted no EDCTs for calmer periods. The mean absolute error (MAE) for a look-ahead horizon of one hour is 3.78 EDCT-delayed arrivals when working with only METAR, TAF, and schedule data. The MAE for the same time horizon is 2.24 when data on the current EDCT and delay situation is added to the feature set (see Figure 15 and Figure 16).

While the time-series graphs show that the model can predict upcoming EDCT situations in all tested settings, they also reveal that the model lacks the ability to predict the exact number of EDCT-delayed flight arrivals, i.e., the actual severity of the airport capacity constraint. While the actual number of EDCT-delayed flights can reach a number higher than 150, the model's predictions vary between 0 and a value of around 50. This may be caused by the fact that the model is trained using mean-squared error (MSE) as a loss function. The MSE penalizes high deviations from the target values, thereby encouraging more cautious predictions. Addressing this problem through a custom loss function or other approaches can be the focus of future works.

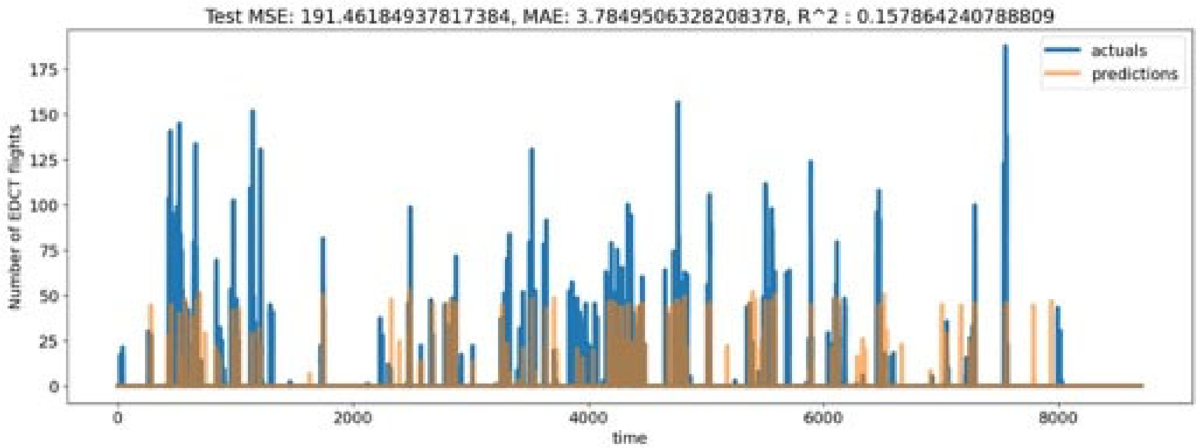



*Figure 15.* Model performance - base feature set, prediction for h+1

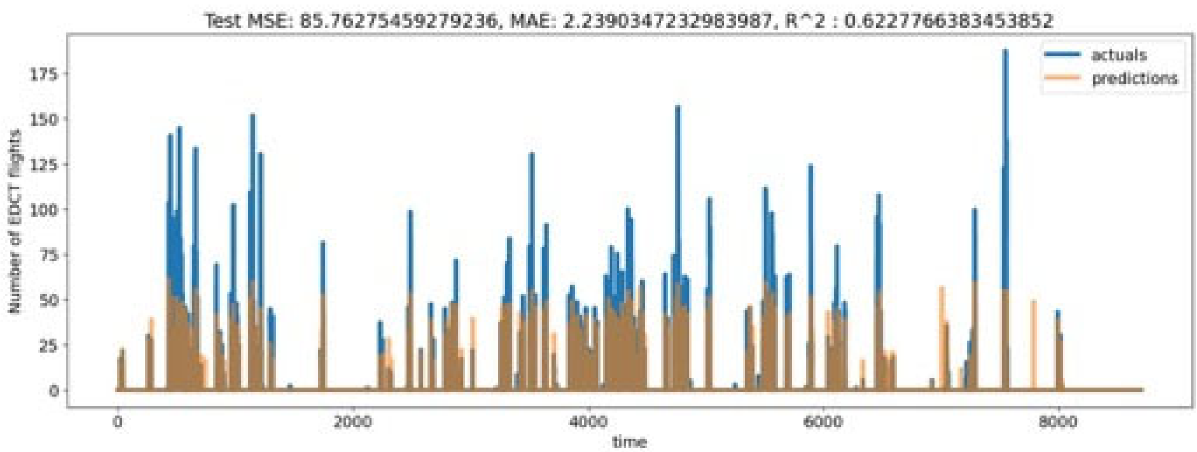



*Figure 16.* Model performance - full feature set, prediction for h+1

### 3.5 Model Explainability Analysis

XAI models are increasingly being used to build explainability and trust in AI-driven decisions. Various techniques and frameworks have been proposed to address the need for interpretability in complex time-series tasks. There is currently a diverse range of XAI models being applied. A study conducted by Fouladgar et al. evaluated three XAI models which included the Local Interpretable Model-Agnostic Explanations (LIME), Integrated Gradient (IG), and SmoothGrad (SG) on Convolutional Network Waterfall (Fouladgar, Alirezaie, & Främling. 2022). Each of these models is used to explain feature importance in a model. LIME works on the principle of perturbation to generate explanations by approximating a given model locally. The input data is perturbed, and the change in the output is observed to determine feature importance. IG is gradient-based. It calculates the gradient of the output with respect to the input data and integrates the gradient along a path of interest. SG also uses gradient. It reduces the noise in a model by averaging the gradients of multiple noisy samples. Although these models are not further discussed in this paper, they provide a general idea of other XAI options in the field in addition to Bi-LSTMs.

Single feature occlusion is then investigated to determine the impact that each feature in the dataset has when contributing to the EDCT predictions. The feature with the highest impact is TAF weather data indicating showers. It has an impact of 0.72 while the feature with the least impact is the METAR cloud type with an impact of 0.016. Figure 17 shows a stem graph of the single feature impact when occlusion is applied.

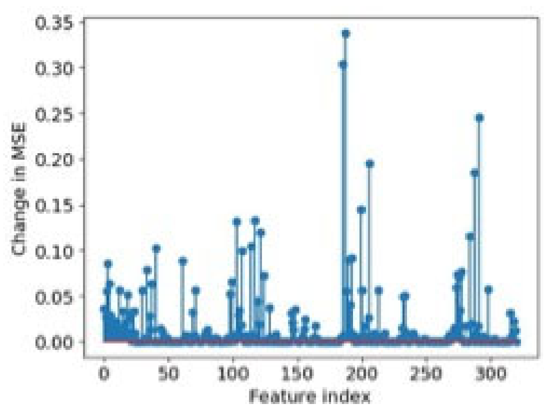



*Figure 17.* Single Feature Importance

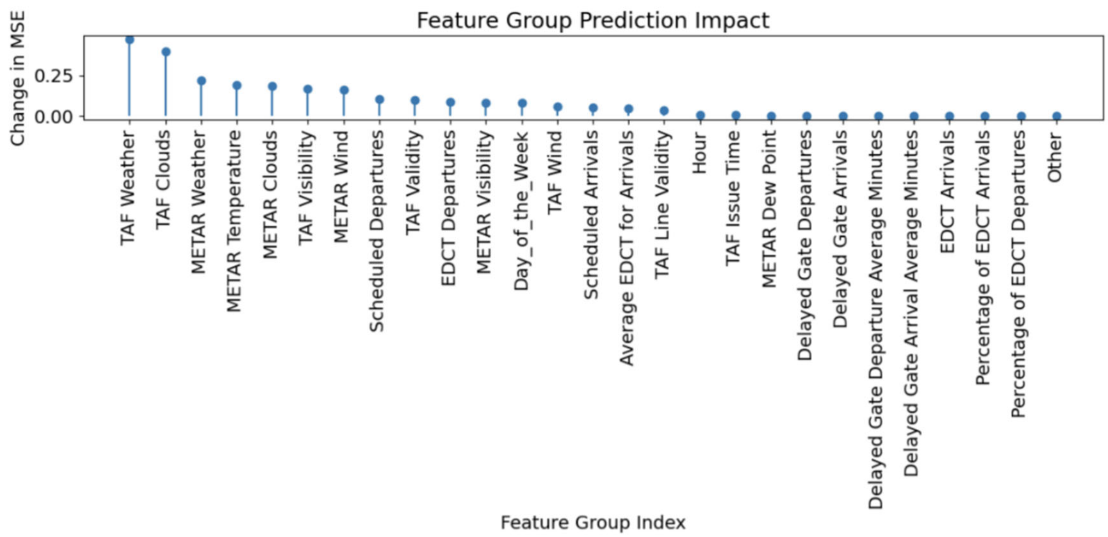



*Figure 18.* Multiple Feature Occlusion Impact

Multiple feature occlusion is then investigated to determine the impact that features with similar characteristics have on the models' predictions when grouped together. There were a total of 27 feature groups and the group with the highest impact was the TAF weather data with an impact of 0.47 as seen in Figure 18.

### 3.6 Summary of Phase-I

Or the model demonstrates the capability to predict the timings of EDCT delays, there is room for improvement concerning the predicted number of EDCT-delayed flights, i.e., the severity of the airport capacity constraint. This problem could be addressed by introducing a loss function other than MSE. Alternatively, a two-step approach can be used by first classifying if any EDCTs are to be expected in a given hour, and if so, predicting the number of EDCT flight delays in a second step.

In addition, this work assumes that the number of EDCT-delayed arriving flights can be used to approximate the severity of airport capacity constraints. It should be noted that not all traffic management initiatives assign EDCTs to flights (Federal Aviation Administration., 2009). More work is needed to incorporate the impact of these traffic management initiatives.

Furthermore, METARs and TAFs are data in form of a text string. Various data preparation steps are needed to convert this data into a tabular format. In a future version, a model could be expected to identify the pertinent features from a text string without relying on heavy data preparation. A model that uses text input to predict airport capacity constraints could also be trained to work with other textual data that cannot be easily converted into a tabular format, for example ATC advisories. Alternatively, other real-time data sources with a predefined (tabular) structure can be explored.

Moreover, as the interpretability of the predictions may increase the trust of decision-makers in air traffic management, adding explainable AI approaches is expected to boost the acceptability of the model and its output.

In summary, this part of the project research builds a foundation for feature-rich, NN-based airport capacity constraint prediction. When combined with the above-mentioned improvements and implemented for major airports in the NAS, this work can help to contribute to better predictability in air traffic management, thereby decreasing overall delay impacts and increasing airspace efficiency.

# 4. PHASE-II A COST-AWARE APPROACH FOR FLIGHT RESOURCES AGGREGATION DURING PRE-DISASTER EVACUATION

## 4.1 Introduction

In an emergency, such as a hurricane or other natural disaster, logistics become the main priority for emergency services. While helping citizens in dangerous situations is important, removing as many citizens as possible from the situation itself makes the emergency significantly more manageable. As a result, prompt and cost-effective evacuation of civilians is vital. In the modern world, air travel is a fast and flexible transportation option when given minimal time to prepare (Ahmed & Dey, 2020; Alahi Kawsar, et al., 2019; Guo, Zhang, & Wu, 2023).

Historically, air travel tends to be used for either long distance evacuations or evacuations from non-natural disasters (Baharnemati & Lim, 2012; Bi et al., 2019). As a result, current solutions for evacuation from natural disasters in Florida focus more on sheltering in place and evacuation by land or sea. However, in cases of serious disasters, land and sea evacuations have historically resulted in serious traffic jams, traffic accidents, and gas and supply shortages. These circumstances extend and exacerbate the emergency situation, as well as limit the outflow of citizens and inflow of supplies. Although sheltering in place is a common emergency strategy for many citizens, in larger and more destructive natural disasters, it quickly becomes a detriment instead of a benefit as valuable resources are being used to rescue people from compromised shelters.

Rather than explore the optimization of traffic flow for land evacuations (as many current solutions for emergency evacuations do), this paper seeks to optimize the use of air travel for such evacuations. Aircraft can be diverted from their scheduled flights and rerouted to the evacuation airport for use. The flights to be used in the evacuation must be selected based on the cost of diverting the flight and the flight's usefulness (e.g., the number of passengers that can be evacuated on the aircraft). Aircraft evacuations also avoid the main difficulties of traditional evacuations, such as traffic accidents and jams, while moving large batches of citizens faster and farther away from the inclement danger.

To optimize this air-based evacuation, this part of the research utilizes a Particle Swarm Optimization algorithm to select flights to divert for evacuations from the Daytona Beach International Airport (DAB). The algorithm seeks to minimize the cost of diverting aircraft from their scheduled flights while evacuating a specified number of passengers without overwhelming the airport's operational capacity.

## 4.2 Methodology

### A. Data Collection

Data for this paper was gathered from public databases on the Bureau of Transportation Statistics website and the Flight Aware aircraft registration website. Some supplementary reference data was gathered from Wikipedia after checking the reliability of the website's cited sources. The process of web scraping was employed to gather data from these sources. The web scraping involved writing programs in Python using packages such as BeautifulSoup and Selenium to automate the extraction of data by navigating the web pages, locating the desired content, and retrieving it for analysis. The use of web scraping allowed for systematic and efficient retrieval of data from various online sources, ensuring a comprehensive and up-to-date dataset for this paper. Care was taken to adhere to ethical guidelines and legal considerations during the web scraping process, respecting the terms of use of the targeted websites. For data that was more difficult to scrape, an available .csv file was directly downloaded from the web page. This file would need to be re-downloaded if the data was needed in future usages and applications of this research.

### B. Data Preparation

To preprocess the collected data, some assumptions were made to reduce the number of variables and unknowns in the problem. These assumptions can be easily adjusted to alter the situation being evaluated in the optimization problem.

It was assumed that evacuations would take place for seven days leading up to the hurricane's arrival. The hurricane referenced here was Hurricane Matthew, which hit Daytona

Beach on October 7, 2016. Therefore, the dataset of flights was restricted to flights occurring on or between October 1 and October 7, 2016.

Next, it was assumed that flights for the evacuations would only be diverted from six hub airports since these hub airports would more likely be able to handle the cancellation and diversion of aircraft. The six airports considered were: ATL (Atlanta, GA), CLT (Charlotte, NC), HOU (Houston, TX), IAH (Houston, TX), DFW (Dallas, TX), and DAL (Dallas, TX). This assumption allowed the dataset to be further reduced to only include flights originating from these six airports.

It was also assumed that only the legacy airlines (American Airlines, Delta Airlines, and United Airlines), Southwest Airlines, and Jet Blue would be able to effectively and efficiently handle diversions due to their size. Smaller airlines were not considered since they would not be as well equipped to manage the diversion of their aircraft from their scheduled flights. Therefore, the dataset was further reduced to only include these five airlines' flights.

The tail numbers of the aircraft in this reduced dataset were then used to identify the type of aircraft used in each flight and how many passengers each aircraft would hold. Any aircraft which could not be identified was removed from the dataset. Finally, any aircraft with 50 or fewer seats was removed from the dataset since smaller aircraft would not be useful in mass evacuations and would not be worth the cost of diverting them. This resulted in the final dataset used by the Particle Swarm Optimization algorithm to select the optimal flights. This final dataset had 10,478 rows with the following quantities as features: day of the month, date, airline code, tail number, origin airport, scheduled destination airport, departure time, taxi time out, taxi time in, actual elapsed time, air time, distance to DAB, and number of seats.

To calculate the fuel cost for diverting a flight, distance, fuel cost (per gallon per seat), number of seats, and fuel consumption rates (mpg) were used. An average fuel cost over the last five years (excluding 2020 due to the biased effect of COVID-19 on the airline industry) was found using the Bureau of Transportation Statistics' Airline Fuel Cost and Consumption (U.S. Carriers - Scheduled) dataset, and an average fuel consumption rate was determined by cross-referencing various sources cited on Wikipedia. Dividing the average value of fuel cost by the average fuel consumption gave an average value used as a constant multiplier for all flights investigated. The constant was multiplied by the distance a particular flight would travel and the number of seats on the aircraft to obtain the estimated fuel cost of diverting the flight.

To determine the delay of diverting a flight from its original destination to DAB, the average airtime of all flights in the reduced dataset per mile was calculated. This average value was multiplied by the distance from the origin to DAB. The delay was calculated by multiplying the calculated airtime by two and adding two hours to account for aircraft turnover time.

To estimate the customer compensation payments for the delay caused by the relocation of a flight, the individual policies of each airline considered were analyzed to determine under which conditions delayed passengers would be given meal or overnight compensations. Meal compensation has been estimated at 15 dollars per person, while hotel fees were taken from the relative surrounding area of each origin airport. The conditions and costs of passenger compensations were used to estimate the cost of diverting any flight more accurately for the evacuations.

**C. Particle Swarm Optimization**

The Particle Swarm Optimization is a stochastic algorithm that uses a random initialization of particles with a position and velocity in the solution space. The objective function is evaluated for each particle based on its current position in the solution space. Afterward, each particle's position and velocity is updated as each particle adjusts its position and velocity based on both its own historical best location and the historical best position of its neighboring particles. This process is repeated until a defined termination criterion is met.

In this paper, a PSO algorithm, as in Figure 19, was implemented by loading the preprocessed data into Python, defining the objective function and constraints of the problem, and optimizing the global best position using the Python Pyswarm package. The objective function considers the cost of fuel, the number of passengers able to evacuate in the aircraft, and the meal and overnight delay compensation given to passengers of delayed flights. The problem is constrained by a maximum number of aircraft that can be hosted at DAB at any instance and the number of people who need to be evacuated from Daytona Beach.

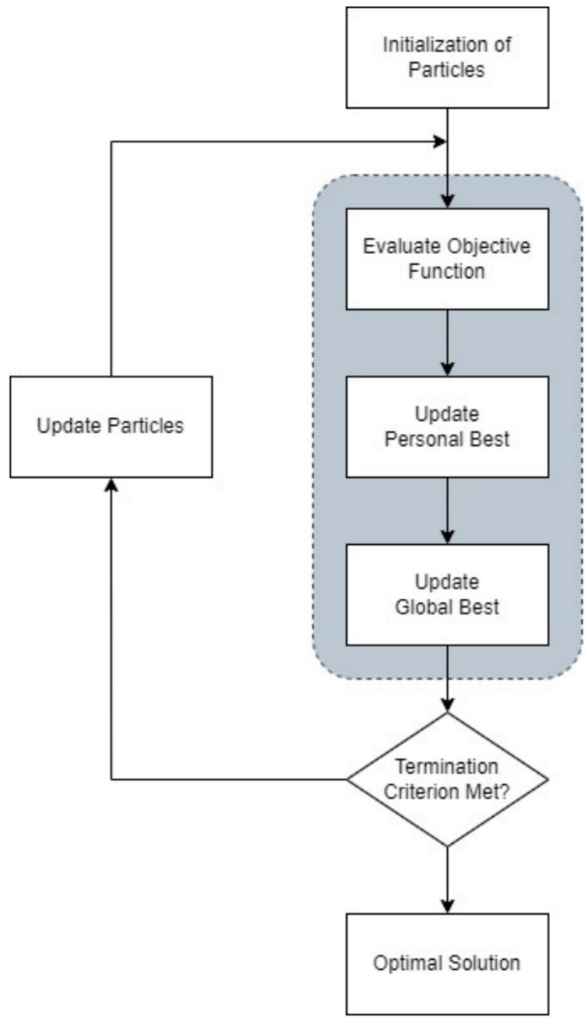



*Figure 19.* General structure of the Particle Swarm Algorithm

### 4.3 Evaluation and Discussion

A test set was run to validate the proposed algorithm simulating a small evacuation from DAB. In this simulation, ten candidate flights are considered by the algorithm, which aims to evacuate a total of 1000 people from DAB at a minimum total cost. The algorithm input data for the ten candidate flights is shown in Table II.

TABLE II: SAMPLE DATA SET USED TO SHOW ALGORITHM RESULTS

| *AIRLINE* | *ELAPSED TIME* | *DISTANCE* | *NUMBER SEATS* |
| --- | --- | --- | --- |
| *AA* | 1911.0 | 416.0 | 379.0 |
| *AA* | 1727.0 | 416.0 | 422.0 |
| *DL* | 1410.0 | 366.0 | 189.0 |
| *DL* | 2313.0 | 366.0 | 222.0 |
| *DL* | 2.0 | 366.0 | 142.0 |
| *DL* | 1616.0 | 366.0 | 142.0 |
| *DL* | 904.0 | 366.0 | 100.0 |
| *UA* | 2305.0 | 861.0 | 149.0 |
| *UA* | 820.0 | 861.0 | 191.0 |
| *WN* | 1025.0 | 969.0 | 149.0 |

The algorithm was then allowed to run and gave a result with outputs including airline and flight number to identify which flights are considered, the total price to divert the aircraft for the evacuation, and a binary value noting whether the algorithm has selected the flight for its optimal solution. The solution found for the test simulation is shown in Table III.

TABLE III: TABLE OF SELECTED AND NON-SELECTED FLIGHTS

| **AIRLINE** | **FLIGHT NUMBER** | **PRICE (USD)** | **REDIRECTED** |
| --- | --- | --- | --- |
| **AA** | N127AA | 39914.4608 | 1 |
| **AA** | N833AA | 44443.0144 | 1 |
| **DL** | N3739P | 24372.5328 | 0 |
| **DL** | N804DN | 28628.0544 | 1 |
| **DL** | N912DL | 18311.6384 | 0 |
| **DL** | N932DN | 18311.6384 | 0 |
| **DL** | N961AT | 12895.5200 | 0 |
| **UA** | N14242 | 21220.4608 | 0 |
| **UA** | N71411 | 27202.0672 | 0 |
| **WN** | N340LV | 34323.1632 | 0 |

The algorithm found that a minimum cost solution to evacuate 1,000 people from DAB would require $112,985.52 to be paid to relocate a total of 3 flights. This price will account for the delayed passengers' compensation (meals and lodging) and fuel costs for the aircraft to be diverted and used.

### 4.4 Summary of Phase-II

In our research, PSO algorithm was able to successfully optimize the selection of flights for the simulated evacuation, though the number of candidate flights was severely limited in this test run. By its nature, the PSO algorithm operates by searching through all combinations of candidate flights while searching for an optimum. This results in long computation times if the list of candidates gets large; therefore, the algorithm would take a long time to locate a global minimum if it is asked to search through the entirety of the dataset of candidate flights.

Potential future works would include working to reduce the computation time and power required to find a global solution and exploring a method of accounting for all the available flights rather than a small set, like the test simulation discussed in IV. This could be achieved by reducing the dataset to search for solutions for smaller time blocks instead of searching a full week of potential flights in a single simulation. Another potential method to account for large sets of flight data would be to apportion the data into evenly sized subgroups, each placed through the particle swarm algorithm to produce the optimal flights for that respective subgroup. Once all the local optimal flights are obtained, a set of data containing only the optimal values would be run through the particle swarm algorithm to obtain a true minimum expense with the flights needed to meet the constraints. The use of such batch testing would allow for several local minima to be discovered independently of each other, narrowing the set of potential global solutions for the algorithm to search through.

# 5. PHASE-III NEURAL-NETWORK ACCELERATED GENETIC ALGORITHM FOR OPTIMAL EVACUATION FLIGHT PLANNING

## 5.1 Introduction

Emergency situations are an unavoidable phenomenon and their impact on the aviation industry cannot be neglected. Every year, many people are forced to move from one place to another due to the impact of emergencies, such as fires and hurricanes (Smith and Katz 2013; Tri et al., 2020). Hurricanes are the most widespread problem in the southeastern part of America where "hurricane season" occurs yearly. During an emergency period, there is mass movement of people to safer places, which increases traffic on all methods of transportation from roadways to airways. Due to the limited capability of airports to support the landing and takeoff of planes, not all evacuated focused aircraft can be used for transporting people. There is an essential requirement for integrating a meticulously calibrated and validated traffic incident response module into the process of modeling and simulating evacuation scenarios (Zhu et al. 2019).

A well-devised air mobility evacuation plan is characterized by its ability to not only sustain ongoing aerospace traffic but also its ability to seamlessly accommodate the surge in demand for outgoing flights. However, existing studies commonly formulate evacuation plans that utilize an airport's entire capability, often at the expense of disrupting ongoing air mobility

This research focuses on efficiently evacuating people without compromising the ongoing airport schedule. Specifically, we leverage the selected airport's existing capabilities that is used for military (MIL) and General Aviation (GAV) operations. By temporarily redirecting management capabilities from these less critical tasks, we enable evacuation flights to seamlessly coexist with regular air traffic. The primary contributions of this study are:

- We propose a novel approach for creating evacuation plans during emergencies, which maximizes the emergency-impacted airport's outgoing capacity without interrupting existing airspace operations.

- Enhancement of our model's efficiency through the integration of the Genetic algorithm(GA) with a Neural Network(NN) to reduce the required number of iteration towards optimal solution.
- We evaluate our solution extensively with real flight operation data.

## 5.2 Methodology

### A. Problem Formulation

We analyzed the capabilities of diverse aircraft types at nine major airports in Florida. We also analyzed the operation records of diverse types of aircraft, including Air Carrier (AC), Air Taxi (AT), GAV, and MIL aircraft. In which an AC is termed as an aircraft with a seating capacity exceeding 60 seats. An AT aircraft can accommodate a maximum of 60 seats. GAV encompasses all civil aircraft movements involving takeoffs and landings, excluding those classified as AC or AT aircraft. Furthermore, MIL aircraft activities, encompassing the full spectrum of MIL takeoffs and landings provided Federal Aviation Administration (FAA) and the Federal Test Center (FTC).

We utilize the GAV and MIL capabilities of the airports during evacuation to mitigate the impact on regular AC and AT traffic. In the meantime, we want to minimize potential delays at the destination airports. To achieve this, we introduced two key metrics: the mean of combined capability ($c$) and its standard deviation ($s$). The mean of combined capability ($c$) represents the average value of the combined capability of GAV and MIL operations at the airports according to historical data. This metric serves as a reference point for evaluating the effectiveness of airport capabilities in emergency scenarios. The standard deviation ($s$) measures the extent of variation in the mean value of the combined capability across the airports. Mathematically, we aim to derive an evacuation flight schedule in an hourly manner that can maximize $c$ and minimize $s$ simultaneously.

## B. Data Processing

The dataset used for our study was collected from the FAA data repository, specifically focusing on the first two months of 2023. The data was sourced from various datasets to comprehensively address our research objectives which are:

- Traffic Flow Management System Counts (TFMSC (https://aspm.faa.gov/tfms/sys/OPSNET.asp): This dataset is derived from the Air Traffic Airspace Lab's Traffic Flow Management System. It aids in assessing the combined capabilities of the airports which are essential for gauging the potential of GAV and MIL operations during evacuation scenarios.
- Aviation System Performance Metrics (ASPM. (https://aspm.faa.gov/apm/sys/AnalysisCP.asp): The ASPM dataset is accessible through an online access system provided by the FAA. It delivers comprehensive data regarding flights to and from the ASPM airports, as well as all flights operated by ASPM carriers. This dataset is integral to performing city pair analysis, crucial for determining flight durations and finding the top ten destination airports.

We organized the TFMSC dataset into an hourly basis. We computed two key statistical metrics for each hour slot of the day: the mean of combined capability ($c$) and its corresponding standard deviation ($s$). These metrics provided valuable insights into the average and variability of the airports' incoming flight acceptance capabilities hourly. We also incorporated data from an additional 24 airports outside Florida, representing the top ten destinations from the nine airports in Florida. In total, we include 33 airports modeling the emergency evacuation scenario.

From these nine Florida airports, we identified a set of 24 unique domestic airports that served as the top ten destinations. When ranking these destinations, we prioritized them based on the number of flights connecting each destination. Higher flight frequencies resulted in higher ranks. Figure 20 displayed the top ten ranked destinations for each of the nine major airports in Florida.

To facilitate neural network training, an additional data synthesis process was introduced. During this procedure, the target airport for which the evacuation strategy was being developed was excluded from consideration. Data from all other airports was amalgamated, encompassing attributes such as popularity ($p$), the mean combined capability ($c$), and the standard deviation ($s$) of $c$. For each of the top ten destinations, a set of three values was recorded, culminating in a total of 30 columns in the dataset. These values were collected on an hourly basis, offering a comprehensive temporal perspective. In a subsequent phase, supplementary columns were introduced to the dataset. These additional columns were designed to include eleven entries for the best fitness score attained and ten entries for indicating the sequential selection or non-selection of the top ten destination airports. The complete algorithm for synthesizing the data is given in Algorithm 1.

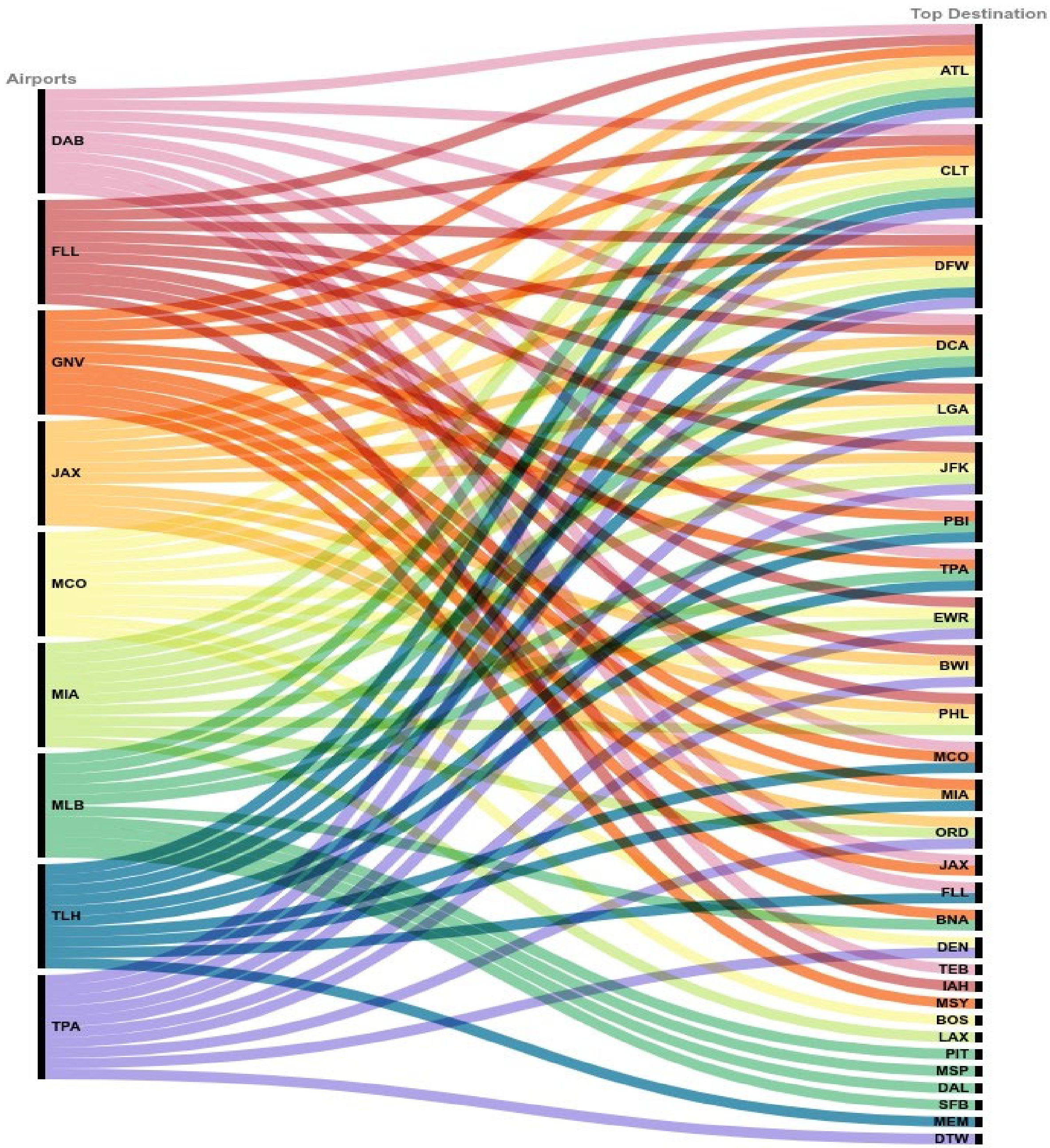



*Figure 20:* Top ten destination airports from the nine major airports in Florida.

Algorithm 1: Algorithm for generating dataset for Neural Network

**Input**:Historical flight data from TFMSC and ASPM having N entries

**Output**: Synthesized dataset containing containing 42 columns, 30 for each $p$,$c$ and $s$ values of top ten destination, 1 for mean Capability of evacuating airport, 1 for fitness score and 10 for best combination listed individually per column.

```
Let i = 0.
while i < N do
    Create all possible 1024 combination of selection list
    [S1, S2, S3, S4, S5, S6, S7, S8, S9, S10]
    Calculate fitness score for each of them
    Let j = 0, bestFitness = 0, bestCombination = []
    while j < 1024 do
        if fitness > bestFitness then
            bestFitness = fitness, bestCombination = combination
        end if
    end while
    ADD mean Capability of evacuating airport, bestFitness and bestCombination to the corresponding row of the dataset
end while
return synthesized dataset
```

## C. Genetic Algorithm

The GA operates by identifying optimal chromosomes from a population pool, employing the principles of crossover and mutation over multiple generations. In the context of our specific problem, a chromosome is symbolized as a list containing ten elements, with each element assigned a value of either 1 or 0. These ten elements correspond to the decision of whether to select or reject flights to the top ten destination airports from the evacuating airport. The effectiveness of a chromosome is evaluated through a fitness function. This fitness function is formulated using the equation specified below:

$$FitnessScore = 0.5*p+0.2*c-0.3*s-penalty \quad (1)$$

- $p$ is the popularity value of the airport
- $c$ is the mean of the combined capability
- $s$ is the standard deviation of the combined capability
- *penalty* equals to 1 when choosing more number of flights in destination airports beyond the capability of evacuating airport

To enhance the effectiveness of our GA, several strategies were incorporated to optimize the evacuation strategy selection process. These strategies were fine-tuned to prioritize certain factors and balance trade-offs during decision-making.

- **Positive Influence of Popularity and Capability:** In the pursuit of creating more effective evacuation strategies, the algorithm was configured to place a positive emphasis on \$p\$ and \$c\$. This incentives the algorithm to prioritize airports with higher popularity and greater capabilities to maximize the utilization of well-equipped and frequently used airports.
- **Minimization of Standard Deviation:** To mitigate potential delays arising from disparities in the c across different hours of the day, s was introduced with a negative weight which prompts the algorithm to favor airports with lower standard deviation values, thus reducing the variability in evacuation efficiency.
- **Penalties for Over Selection**: To avoid an unrealistic scenario where the algorithm selects too many airports, a penalty mechanism was introduced which comes into play when the selection of airports surpasses the combined capability of the evacuating airport. By penalizing such instances, the algorithm is encouraged to strike a balance between maximizing the fitness score and adhering to practical constraints.

The complete GA implementation, encompassing these strategies, is presented in Algorithm 2.

**Algorithm 2: GeneticAlgorithm**

**Input**:Population size($population\_size$), number of generations($num\_generation$), crossover rate, mutation rate, mean of combined capability of evacuating airport(c), dataset(d) containing details of destination airports
**Output**: The best selection list found in the last generation.

1: Initialize the population of ($population\_size$)
2: Evaluate the fitness score of each individual in population
3: Let $gen = 1$
4: **while** $gen < num\_generation$ **do**
5:   Create an empty new population
6:   **while** $size(newpopulation) < population\_size$ **do**
7:     Select two parents, parent1 and parent2, from population based on their fitness score(higher fitness score has higher chance of selection)
8:     Perform crossover operation to create two children, child1 and child2, from parent1 and parent2
9:     Perform mutation operation on child1 and child2 to introduce variation in their genetic makeup
10:     Add child1 and child2 to new population
11:   **end while**
12:   Replace population with new population
13:   Evaluate the fitness of each individual in pop
14: **end while**
15: best selection list $\leftarrow$ individual in population with fitness equal to best fitness score
16: **return** best selection list

**Algorithm 3: Algorithm for training and testing for Neural Network**

**Input**:Synthesized dataset for neural network
**Output**:Trained model and result

1: Separate 31 features and 10 labels in the dataset
2: Split data into training(0.8) and testing(0.2) set
3: Model that have 31 input features, utilizes ReLU activation in its two hidden layers with 64 units each, and generates multiple single-value outputs through ten parallel output layers is created
4: loss=MSEloss(), optimizer=Adam(), batchSize=32
5: Let $i = 0, j = 0, numberOfBatches = len(X\_train)/batch\_size$
6: **while** $i < epochs$ **do**
7:   **while** $j < numberOfBatches$ **do**
8:     Create mini batch of data
9:     Calculate the output by passing minibatch of data to the model
10:     Calculate loss by comparing output with the labels
11:     Loss is back propagated
12:     Perform gradient based optimization to update model parameters
13:   **end while**
14:   Print loss for the epoch
15: **end while**
16: Save the model
17: Evaluate the model on the test dataset
18: **return** result

## D. Neural Network

We employed a NN model to accelerate the optimal selection of destination airports for evacuation planning. This model is trained using a curated dataset derived from the integration of TFMSC and ASPM datasets mentioned in Algorithm 1. The NN model predicts the optimal airport selection for evacuation strategies on other airports. By training on the synthesized dataset, the model learned patterns and relationships between different attributes, enabling it to make accurate predictions of the selection of the destination airports. The overall implementation of the NN is given in the Algorithm 3.

## E. Combining Genetic Algorithm and Neural Network

In this phase, we integrated the GA with the trained NN model to facilitate a quicker convergence of the algorithm. We aimed to harness the predictive power of the model to inform the initial population for the GA. The process of generating the parent population involved two approaches:

- **Approach 1:** Randomly inserting the population generated by the model in to the population list.

- **Approach 2:** Sorting the population list based on fitness score in descending and removing the lower order population with the generated population.

### 5.3 Evaluation and Discussion

This section focuses in evaluating the performance, accuracy, and efficiency of the developed methodologies.

**A. Genetic Algorithm**

We used only GA to find the optimal evacuation flight schedules originating from DAB airport. The algorithm was initiated with a population size of 15 and executed for 5 generations. The population size and number of generations were systematically increased by factors of two and five to explore their impact on the algorithm's performance. During this iterative process, it was observed that the algorithm exhibited a substantial convergence trend when the population size was set to 75 and the number of generations was increased to 25, as in Figure 21. The observed convergence trend suggests that a population size of 75 with 25 generations struck a balance between exploration and exploitation, leading to optimal results in terms of identifying the most effective evacuation flight schedules from the selected airport.

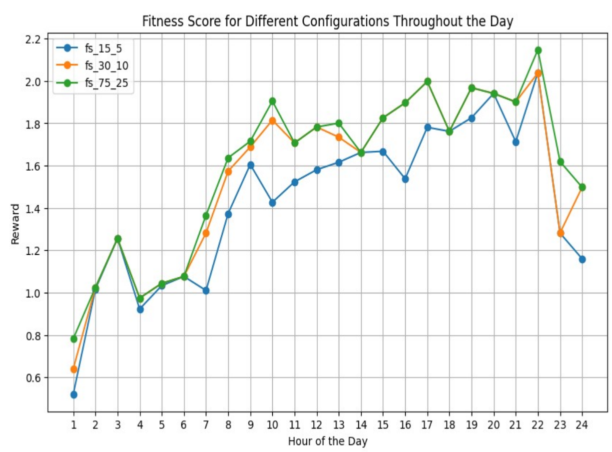



*Figure 21:* Fitness Score for various combination of population size and number of generation where fs p g represents the fitness score for model using population size of p and number of generation g.

### B. Neural Network

The NN was trained using the synthesized dataset outlined in Algorithm 1. During training, the NN was exposed to varying numbers of epochs, 5, 15, and 25, and it was observed that convergence commenced at approximately 25 epochs. Prior to convergence, there was a notable performance fluctuation as depicted in Figure 22.

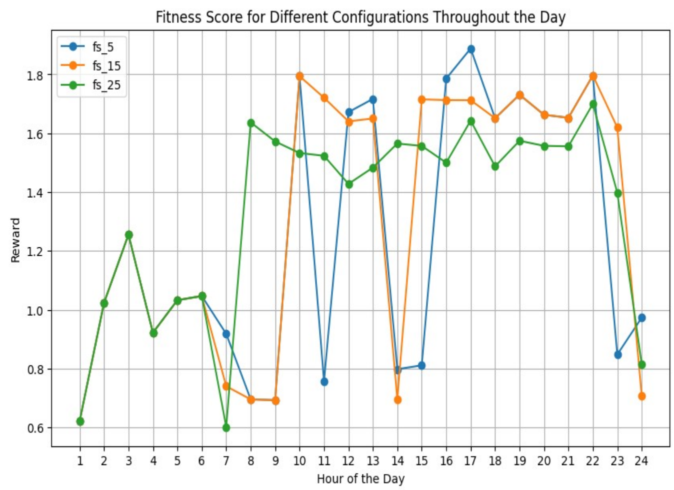



*Figure 22:* Fitness Score for various neural networks trained under different number epochs where fs n is the model trained for n epochs.

### C. Combination of Genetic Algorithm and Neural Network}

To demonstrate the effectiveness of the Neural Network-accelerated Genetic algorithm. We select suboptimal scenarios for each method individually and merge them to find an optimal scenario. Specifically, the GA with a population size of 15 and 5 generations and the NN trained for only 5 epochs, respectively. The combined approach, referred to as the "NN-accelerated GA," leveraged the predictive capabilities of the NN to enhance the parent population generation for the GA. The strategy entailed using the NN to produce parent candidates, which were then integrated into the parent pool. The GA was then employed to refine the parent pool further. Parents were added based on the fitness score, considering the NN-generated parents alongside those generated by the GA itself.

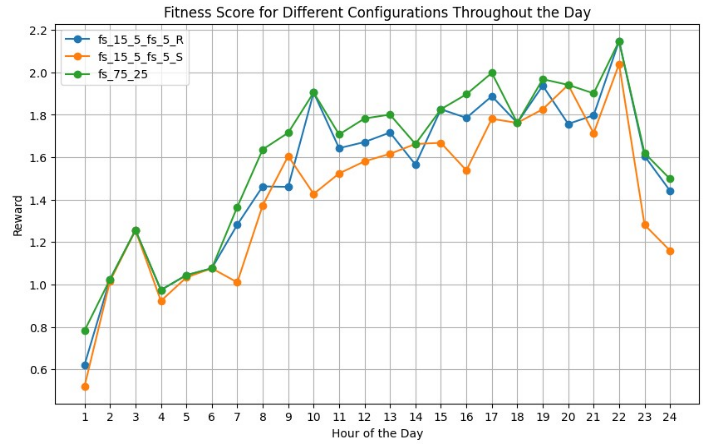



*Figure 23:* Fitness Score of different models where fs 15 5 fs 5 R is fitness score for NN-accelerated GA where fs 15 5 GA model is combined with fs 5 NN model , also the population is randomly replaced by population generated by NN and in case of fs 15 5 fs 5 S least fit population are replaced by sorting them.

We compared two methods: random addition of NN-generated parents and the selective removal of the least-fit population. In the latter approach, we ranked the entire population pool in descending order based on their fitness scores. The least-fit populations were then replaced by the parents generated by the NN. The results from this experiment, as depicted in Figure 23 revealed that the strategy of random addition of parents from the NN yields better outcomes than the approach of selectively removing the least-fit populations.

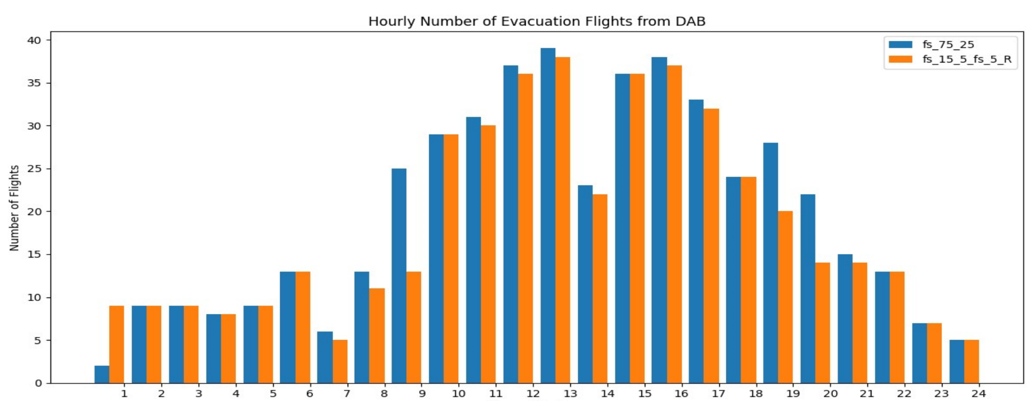



*Figure 24:* Number of evacuation flights scheduled by the different models.

Figure 24 shows the number of evacuation flights that can be originated on hourly basis by models fs_75_25 and fs_15_5_fs_5_R. We found that the prediction of the number of flights by both the model is almost identical except in the midnight scenario. The reason might be due to the fact the number of flights in the DAB airport in that time duration is very random sometimes they have few flights but most of the time number of flights during those time is almost none. Figure 25 provided detail about the selection of top ten destination airports by our top two models which are fs_75_25 and fs_15_5_fs_5_R. As shown in the figure, we found that most of the time airport selection by both are identical.

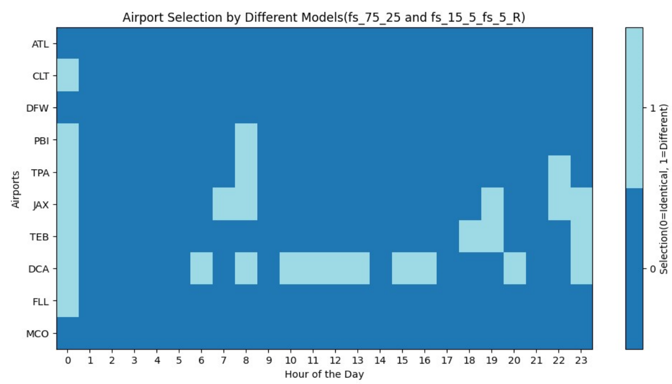



*Figure 25:* Selection of the airports for evacuation by different models

### 5.4 Summary of Phase-III

In this research, we delved into the utilization of GA to develop efficient evacuation plans during emergencies. We leverage the non-commercial flight capability of impacted airports to ensure that evacuation plans minimally affect routine airspace operation. We proposed a neural network accelerated genetic algorithm to derive our solution with smaller computational overhead. We found that GA exhibited slower convergence when operated independently with higher population pool and generations and that its performance significantly improved when assisted with a NN model trained for a mere 5 epochs. This integration yielded comparable results in terms of fitness function, flight numbers, and airport selection. This intriguing finding underscores the NN's ability to expedite GA's convergence, even when trained on data from different airports, displaying its generalization capabilities.

As a prospect for future exploration, we are planning to incorporate another NN to predict fitness scores. This endeavor aims to ascertain whether the fusion of these two NN models could yield even faster convergence rates when combined with our current hybrid model.

# 6. FINDINGS, CONCLUSIONS, RECOMMENDATIONS

## 6.1 Findings and Conclusions

**Phase I:** Explainable Machine Learning for Flight Delay Prediction

- **Findings**: The study introduced a GRU neural network architecture to predict weather-related airport capacity constraints, quantified as the number of flights arriving with EDCT delays. The model was trained on a comprehensive dataset including weather observations, weather forecasts, and flight schedules and then inspected using Occlusion Sensitivity method. The analysis demonstrated that certain weather conditions, notably misalignment of wind direction with runway direction, significantly impact EDCTs. The proposed model reliably predicted EDCT situations using the dataset, with potential for improved accuracy by incorporating additional data on delays and EDCTs.
- **Conclusions**: While the model showed promise in predicting the timing of EDCT delays, there's room for enhancement in predicting the exact number of delayed flights. The study also highlighted the potential for models that directly interpret textual weather data, improving the prediction of airport capacity constraints without extensive data preprocessing.

**Phase II:** A Cost-Aware Approach for Flight Resources Aggregation During Pre-Disaster Evacuation

- **Findings**: This phase focused on optimizing the use of air travel for emergency evacuations, specifically from natural disasters, using a Particle Swarm Optimization algorithm. The algorithm aimed to select flights for diversion to evacuation airports, minimizing costs while considering the airport's operational capacity. The approach demonstrated success in selecting optimal flights for evacuation scenarios, albeit with a limited number of candidate flights in the simulation.
- **Conclusions**: The PSO algorithm effectively optimized flight selection for evacuations, suggesting air travel as a viable and efficient evacuation method.

However, the computational intensity of the algorithm poses challenges for scalability. Future directions include reducing computation time and exploring methods to accommodate larger datasets of available flights, through data segmentation or batch testing for more efficient global minimum searches.

**Phase III:** Neural-Network Accelerated Genetic Algorithm for Optimal Evacuation Flight Planning

- **Findings**: The study explored the use of a genetic algorithm (GA), accelerated by a neural network (NN), to develop efficient evacuation plans that minimally impact routine airspace operations. The hybrid model showed that even a briefly trained NN could significantly improve the GA's convergence speed, yielding effective evacuation plans without extensive computational resources.
- **Conclusions**: The integration of NNs with GAs presents a promising approach for optimizing evacuation flight planning, enhancing the efficiency of emergency responses in the aviation sector. The study suggests further research into the use of additional NN models for predicting fitness scores, potentially achieving faster convergence rates and more effective evacuation strategies.

## 6.2 Recommendations

### A. For Policy Makers:

- **Invest in Data Infrastructure:** Enhance the collection and accessibility of real-time weather, flight schedule, and air traffic management data to support advanced predictive analytics for air mobility under emergency situations.
- **Encourage Collaboration**: Foster partnerships between governmental agencies, airlines, and research institutions to share data and insights, facilitating the development of more efficient and effective emergency response strategies.
- **Support Technological Innovation:** Allocate resources towards the research and development of AI and machine learning applications in air traffic

management, emphasizing the importance of explainable and interpretable models to ensure trust and reliability.

- **Update Regulatory Frameworks:** Revise current regulations to incorporate the use of advanced predictive and optimization tools in emergency planning and response, ensuring that these tools can be swiftly and effectively integrated into operational practices.
- **Prioritize Training and Preparedness:** Implement training programs for emergency management personnel on the use of AI and optimization tools in evacuation planning and execution, ensuring readiness to deploy these technologies in real-world scenarios.

**B. For Airline Operators:**

- **Adopt Advanced Predictive Tools**: Integrate machine learning models for predicting flight delays and airport capacity constraints into operational decision-making processes, enhancing the ability to adapt to emergency situations.
- **Participate in Data Sharing Initiatives:** Engage in industry-wide efforts to share relevant data, contributing to the collective ability to respond more effectively to emergencies and improve air mobility.
- **Invest in Flexible Resource Management:** Develop capabilities for rapid reallocation of flight resources during emergencies, utilizing cost-aware approaches to minimize disruptions and optimize evacuation efforts.
- **Enhance Customer Communication:** Utilize predictive models to provide passengers with real-time updates on flight statuses and potential delays, improving transparency and customer service during emergency situations.
- **Implement Dynamic Planning Systems:** Leverage optimization algorithms to adapt flight schedules and routes dynamically in response to emergency conditions, maximizing safety and efficiency.

C. For Researchers:

- **Focus on Model Generalizability:** Work towards developing models that can be easily adapted to different airports, emergency scenarios, and data conditions, increasing the applicability of research findings across the aviation industry.
- **Explore Hybrid Models:** Investigate the integration of different AI techniques, such as combining genetic algorithms with neural networks, to enhance the efficiency and effectiveness of predictive and optimization models.
- **Advance Explainable AI:** Prioritize the development of explainable and interpretable machine learning models to increase their acceptance among decision-makers and stakeholders in air traffic management.
- **Conduct Impact Studies:** Evaluate the real-world impact of implementing AI and optimization tools in emergency air mobility scenarios, providing evidence-based recommendations for policy and practice.
- **Engage with Stakeholders:** Collaborate closely with airline operators, air traffic controllers, and emergency management personnel to ensure that research is aligned with operational needs and practical constraints.

## REFERENCES


1. Ahmed, S., & Dey, K. (2020). Resilience modeling concepts in transportation systems: A comprehensive review based on mode, and modeling techniques. Journal of Infrastructure Preservation and Resilience, 1(1), 1–20.
2. Airbus. (2021). Accidents by flight phase | Airbus Accident Statistics. https://accidentstats.airbus.com/statistics/accident-by-flight-phase'
3. Airoldo, D. M. (2022). United airlines to pull out of New York's JFK airport in October. Retrieved from https://www.businesstravelnewseurope.com/Air-Travel/United-Airlines-to-pull-out-of-New-York-s-JFK-airport-in-October
4. Airport Operations 101 (2024). Retrieved from: https://www.flychicago.com/community/ORDnoise/AirportOperations/pages/default.aspx
5. Airline Industry Information (2017) United Airlines provides update on operations with hurricane Irma. Retrieved from: https://united.mediaroom.com/2017-09-09-United-Airlines-Provides-Latest-Update-on-Operations-and-Relief-Efforts-with-Hurricane-Irma
6. Alahi Kawsar, L., Abdul Ghani, N., Abdulbasah Kamil, A., & Mustafa, A. (2019). Optimization based controlled evacuation. Journal of Intelligent Transportation Systems, 23(5), 477–498.
7. Avery, J., & Balakrishnan, H. (2015). Predicting airport runway configuration: A discrete-choice modeling approach.
8. Baharnemati, M. R., & Lim, G. (2012). An Information-based Rerouting Decision Making Tool for Regional Evacuation. In *IIE Annual Conference. Proceedings* (p. 1). Institute of Industrial and Systems Engineers (IISE).
9. Barry, R. J. (2008). Hurricanes: A Primer for the Aviation Professional. Journal of Aviation/Aerospace Education & Research, 17(3). DOI: https://doi.org/10.15394/jaaer.2008.1455

10. Bertsimas, D., & Patterson, S. S. (1998). The Air Traffic Flow Management Problem with Enroute Capacities. Operations Research, 46(3), 406–422. http://www.jstor.org/stable/222834
11. Bi, C., Pan, G., Yang, L., Lin, C. C., Hou, M., & Huang, Y. (2019). Evacuation route recommendation using auto-encoder and Markov decision process. *Applied Soft Computing*, *84*, 105741.
12. Bureau of Transportation Statistics (2021, 2022, 2023). Airline On-Time Statistics and Delay Causes. Retrieved from https://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?6B2r=FE&20=E
13. Burian, B. K., & Barshi, I. (2003). Emergency and abnormal situations: A review of ASRS reports. In R. Jensen (Ed.), Proceedings of the 12th International Symposium on Aviation Psychology. Wright State University Press. https://human-factors.arc.nasa.gov/publications/20051028111241_Review_of_ASRS_Reports.pdf
14. Burian, B. K., Barshi, I., & Dismukes, K. (2005, June 1). The challenge of aviation emergency and abnormal situations. NASA Technical Reports Server (NTRS). https://ntrs.nasa.gov/citations/20060023295
15. Choi, S., Kim, Y. J., Briceno, S., & Mavris, D. (2016). Prediction of weather-induced airline delays based on machine learning algorithms. In Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference (DASC) (pp. 1-6). https://doi.org/10.1109/DASC.2016.7777956
16. Congress, U. S. (1993). Biopolymers: Making Materials Nature's Way-Background Paper. Office of Technology Assessment, Washington, DC.
17. Cook, A. J., & Tanner, G. (2009). The challenge of managing airline delay costs. In Proceedings of the German Aviation Research Society and University of Belgrade Conference on Air Traffic Management Economics.
18. Crespo, A. M. F., Weigang, L., & Timoszczuk, A. P. (2011). ATFM computational agent based on reinforcement learning aggregating human expert experience. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence in Scheduling (SCIS) (pp. 323-328). https://doi.org/10.1109/FISTS.2011.5973614
19. Daly, K. (2006). A380: everyone off in time. Flight International, 169(5030).

20. Davidson, R. A., Nozick, L. K., Wachtendorf, T., & Blanton, B. (2020). An Integrated Scenario Ensemble‐Based Framework for Hurricane Evacuation Modeling: Part 1—Decision Support System. Risk Analysis, 40(1), 97‐116. https://doi.org/10.1111/risa.12990
21. Dhal, R., Roy, S., Tien, S.-L., Taylor, C. P., & Wanke, C. R. (2014). An operations-structured model for strategic prediction of airport arrival rate and departure rate futures. In 14th AIAA Aviation Technology, Integration, and Operations Conference (p. 2160).
22. Fang, Z., Lv, W., Jiang, L., Xu, Q., & Song, W. (2016). Modeling and assessment of civil aircraft evacuation based on finer-grid. Physica A, 448, 102-112. https://doi.org/10.1016/j.physa.2015.12.092
23. FAR (2002). Part 25: Airworthiness standards: Transport category airplanes. Federal Aviation Administration, Washington, DC, 7.
24. Federal Aviation Administration. (2009). Traffic flow management in the national airspace system. Retrieved from https://www.fly.faa.gov/Products/Training/Traffic_Management_for_Pilots/TFM_in_the_NAS_Booklet_ca10.pdf
25. Federal Aviation Administration. (2021a, November 2). Air Traffic By The Numbers. https://www.faa.gov/air_traffic/by_the_numbers/
26. Federal Aviation Administration. (2021b, November 30). Equip ADS-B. https://www.faa.gov/nextgen/equipadsb/capabilities/benefits/
27. Federal Aviation Administration. (2022). Aviation weather handbook (No. 24.4, pp. 291–312). Retrieved from https://www.faa.gov/regulationspolicies/handbooksmanuals/aviation/faa-h-8083-28-aviation-weather-handbook
28. Federal Aviation Administration. (2023a). ASPM: Analysis: Delayed flights. Retrieved from https://aspm.faa.gov/aspmhelp/index/ASPM__Analysis__Delayed_Flights.html
29. Federal Aviation Administration. (2023b). ASPM: Airport analysis: EDCT report. Retrieved from

https://aspm.faa.gov/aspmhelp/index/ASPM__Airport_Analysis__EDCT_Report.html

30. Federal Aviation Administration. (2023c). ASPM airport analysis manual. Retrieved from https://aspm.faa.gov/aspmhelp/index/ASPM_Airport_Analysis_Manual.html
31. Federal Aviation Administration. (2023d). ASPM 77. Retrieved from https://aspm.faa.gov/aspmhelp/index/ASPM_77.html
32. Federal Aviation Administration. (2023e). ASPM airport analysis: Definitions of variables. Retrieved from https://aspmhelp.faa.gov/index/ASPM_Airport_Analysis__Definitions_of_Variables.html
33. Federal Aviation Administration. (2023f). Meteorology. Retrieved from https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap7_section_1.html
34. Federal Aviation Administration. (2023g). Terminal area forecast (TAF). Retrieved from https://www.faa.gov/data_research/aviation/taf
35. FEMA. (2017). HURREVAC user's manual. https://www.hurrevac.com
36. FEMA. (2018). 2017 hurricane season FEMA after-action report. https://www.fema.gov/sites/default/files/2020-08/fema_hurricane-season-after-action-report_2017.pdf
37. Florida Department of Transportation. (2018). Hurricane Irma's Effect on Florida's Fuel Distribution System and Recommended Improvements. https://www.fdot.gov/docs/default-source/info/CO/news/newsreleases/020118-FDOT-Fuel-Report.pdf
38. Fouladgar, N., Alirezaie, M., & Främling, K. (2022). Metrics and Evaluations of Time Series Explanations: An Application in Affect Computing. IEEE Access, 10, 23995-24009.
39. Gao, M., Hong, C., Xu, B., & Ding, R. (2012). Flight rescheduling responding to large-area flight delays. Kybernetes, 41(10), 1483-1496. https://doi.org/10.1108/03684921211276693

40. Géron, A. (2018). Neural networks and deep learning. O'Reilly Online Learning. https://www.oreilly.com/library/view/neural-networks-and/9781492037354/ch04.html
41. Glazer, N. (2023). How to Read a TAF (Terminal Aerodrome Forecast) [online]. Retrieved from: https://www.pilotmall.com/blogs/news/how-to-read-a-taf-terminal-aerodrome-forecast?srsltid=AfmBOookdG3Mg8DOdMehV4_MEMc_3rLZwtAz_54YtYKHV55jw_NS_KqD
42. Guo, K., Zhang, L., & Wu, M. (2023). Simulation-based multi-objective optimization towards proactive evacuation planning at metro stations. Engineering Applications of Artificial Intelligence, 120, 105858.
43. Herbert, A. S. (1957). *Models of Man. Social and Rational*.
44. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
45. Hurricane Katrina and Aviation. (2005). Air Safety Week, 19(35), 1. https://www.proquest.com/trade-journals/hurricane-katrina-aviation/docview/204680353/se-2
46. LeCun, Y., & Bengio, Y., Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
47. Lopez-Fuentes, L., van de Weijer, J., González-Hidalgo, M., et al. (2018). Review on computer vision techniques in emergency situations. Multimedia Tools and Applications, 77, 17069–17107. https://doi.org/10.1007/s11042-017-5276-7
48. Lowe, C., & How, J. P. (2015). Learning and predicting pilot behavior in uncontrolled airspace. In AIAA Infotech @ Aerospace. https://doi.org/10.2514/6.2015-1199
49. Mahboubi, Z., & Kochenderfer, M. J. (2017). Learning Traffic Patterns at Small Airports From Flight Tracks. IEEE Transactions on Intelligent Transportation Systems, 18(4), 917-926. https://doi.org/10.1109/TITS.2016.2598064

50. Malakis, S., Kontogiannis, T., & Kirwan, B. (2010). Managing emergencies and abnormal situations in air traffic control (part II): teamwork strategies. *Applied Ergonomics*, *41*(4), 628-635.
51. Menda, J., Chen, Y.-C., Grana, J., Bono, J. W., Tracey, B. D., Kochenderfer, M. J., & Wolpert, D. (2019). Deep Reinforcement Learning for Event-Driven Multi-Agent Decision Processes. IEEE Transactions on Intelligent Transportation Systems, 20(4), 1259–1268. https://doi.org/10.1109/TITS.2018.2848264
52. Misra, S., Dsouza, G. & Truong, D. (2022). A machine learning approach towards analyzing impact of surface weather on expect departure clearance times in aviation. Collegiate Aviation Review International, 40(2), 79-102. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9356/8433
53. Mitchell, T. (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7.
54. MITRE. (2023). Digital-Age Innovations Shape the Future of Aviation . Retrieved from https://www.mitre.org/news-insights/impact-story/digital-age-innovations-shape-future-aviation
55. National Geographic Society (2019, February 7). Extreme Weather on Earth. https://www.nationalgeographic.org/activity/extreme-weather-on-earth/
56. PR Newswire. (2005, September 22).-American Airlines Ramps Up Evacuation Efforts in Anticipation of Hurricane Rita; Airline Flies Larger Aircraft Out of Houston. Retrive from: https://www.proquest.com/wire-feeds/american-airlines-ramps-up-evacuation-efforts/docview/451367072/se-2
57. Rahman, R., & Hasan, S. (2018). Short-term traffic speed prediction for freeways during hurricane evacuation: A deep learning approach. In Proceedings of the IEEE 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 1291-1296). https://doi.org/10.1109/ITSC.2018.8569443
58. Ranter, H. (n.d.). ASN Aircraft accident BAC One-Eleven 528FL G-BJRT Didcot. Aviation Safety Network. Retrieved August 28, 2020, from https://asn.flightsafety.org/asndb/326018

59. Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., & Díaz-Rodríguez, N. (2021). Explainable artificial intelligence (xai) on timeseries data: A survey. *arXiv preprint arXiv:2104.00950*.

60. Schultz, M., Reitmann, S., & Alam, S. (2021). Predictive classification and understanding of weather impact on airport performance through machine learning. Transportation Research Part C: Emerging Technologies, 131, 103119. https://doi.org/10.1016/j.trc.2021.103119

61. Smith, D. A., & Sherry, L. (2008, May). Decision support tool for predicting aircraft arrival rates from weather forecasts. In 2008 Integrated Communications, Navigation and Surveillance Conference (pp. 1-12). IEEE.

62. Smith, A. B., & Katz, R. W. (2013). US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural Hazards*, *67*(2), 387-410.

63. Spatharis, C., Blekas, K., Bastas, A., Kravaris, T., & Vouros, G. A. (2019). Collaborative multiagent reinforcement learning schemes for air traffic management. In Proceedings of the IEEE International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-8). https://doi.org/10.1109/IISA.2019.8900719

64. Stevens, E., Antiga, L., & Viehmann, T. (2020). Deep learning with PyTorch. Manning Publications.

65. Sun, J., Ellerbroek, J., & Hoekstra, J. (2016, June). Large-scale flight phase identification from ADS-B data using machine learning methods. In Proceedings of the 7th International Conference on Research in Air Transportation (pp. 1-8).

66. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT Press.
67. Tang, Y., & Xu, Y. (2021). Multi-agent deep reinforcement learning for solving large-scale air traffic flow management problem: A time-step sequential decision approach. In Proceedings of the IEEE/AIAA 40th Digital Avionics Systems Conference (DASC) (pp. 1-10). https://doi.org/10.1109/DASC52595.2021.9594329
68. Tolson, J. J. (1973). Airmobility, 1961-1971 (Vol. 90, No. 4). Department of the Army.
69. Tri, N. K., Wright, S. J., Betts, J. M., & Vu, H. L. (2020). Studying the impact of public transport on disaster evacuation. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1–6).
70. Van de Walle, B., & Turoff, M. (2008). Decision support for emergency situations. Information Systems and e-Business Management, 6, 295-316. https://doi.org/10.1007/s10257-008-0087-z
71. Wang, Y. (2019). "Dynamic Prediction of Runway Configuration and Airport Acceptance Rate." USF Tampa Graduate Theses and Dissertations. https://digitalcommons.usf.edu/etd/8693
72. Wang, G., & Zhao, W. (2020). The principles of integrated technology in avionics systems. Academic Press.
73. Zhao, X., Lovreglio, R., & Nilsson, D. (2020). Modelling and interpreting pre-evacuation decision-making using machine learning. Automation in Construction, 113, 103140. https://doi.org/10.1016/j.autcon.2020.103140
74. Zhu, Y., Ozbay, K., Xie, K., & Yang, H. (2019, October). Modeling of incident-induced capacity loss for hurricane evacuation simulation. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 613-618). IEEE.

## ACKNOWLEDGEMENTS

The PI and Co-PI(s) of this project want to acknowledge the contributions from the following participants from Embry-Riddle Aeronautical University: Wairimu Mwangi, Thomas Fiello, Jose Gonzalez, Marc Jacquet, Isaac LaRosee, Jordan Sanders, and Carina Shanahan, as well as Kamal Acharya from University of Maryland, Baltimore County.

## APPENDIX

Publications, presentations, posters resulting from this project:

### Conference Papers

- Y. Zhou, D. Liu and H. Song, "A Survey of Machine Learning Algorithms and Techniques for Air Mobility Under Emergency Situations," 2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), 2022, pp. 582-588, doi: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics55523.2022.00111.
- Feng, K., Liu, D., Liu, Y., Liu, H., & Song, H. (2023, August). GraphDAC: A Graph-Analytic Approach to Dynamic Airspace Configuration. In *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 235-241). IEEE.
- Acharya K., Velasquez A., Liu Y., Liu D., Sun L, Song H. (2023 December), Improving Air Mobility in Emergency Situation with Neural Network-Accelerated Genetic Algorithm. Accepted for Publication in IEEE ITS 2024,
- Thomas Fiello, Jose Gonzalez, Marc Jacquet, Isaac LaRosee, Yongxin Liu, Jordan Sanders, Carina Shanahan. (2023 December), A Cost-Aware Approach for Flight Resources Optimization During Pre-Disaster Evacuation. Under review by AIAA 2024.