Pattern-of-Life Modeling with Automatic Dependent Surveillance-Broadcast (ADS-B)

DISSERTATION

Sarah J. Bolton, Major, USAF

AFIT-ENG-DS-23-S-063

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-DS-23-S-063

Pattern-of-Life Modeling with Automatic Dependent Surveillance-Broadcast (ADS-B)

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Computer Science

Sarah J. Bolton, B.S., M.S.

Major, USAF

September 14, 2023

AFIT-ENG-DS-23-S-063

Pattern-of-Life Modeling with Automatic Dependent Surveillance-Broadcast (ADS-B)

DISSERTATION

Sarah J. Bolton, B.S., M.S.
Major, USAF

Committee Membership:

Maj Richard Dill, Ph.D., CISSP
Chair

Dr. Douglas D. Hodson, Ph.D.
Member

Dr. Michael R. Grimaila, Ph.D., CISM, CISSP
Member

AFIT-ENG-DS-23-S-063

Abstract

This dissertation outlines the plan to improve a portion of Intelligence, Surveillance and Reconnaissance (ISR) operations as mandated by the National Defense Authorization Act (NDAA) and Department of the Air Force (DAF) Posture Statements [1, 2, 3, 4, 5]. This research aims to improve pattern-of-life (POL) modeling for aircraft operations using Automatic Dependent Surveillance-Broadcast (ADS-B) data. The primary contribution of this dissertation is the development of models and techniques that allow for the identification and classification of aircraft using kinetic and unattributed data. While the ADS-B data set provides plenty of identifying information making it an excellent resource for research, during times of conflict, attributable data like ADS-B cannot be solely relied on. Therefore, this research uses ADS-B data as a surrogate for other, more available data sources.

The research is broken up into three studies. The first study looks at feature reduction and data minimization. The intent behind study 1 is to reduce the computation resources required to perform POL modeling on aircraft. This is important in remote locations where the resources may not be available. Study 1 extends previous research that used an Multivariate Long Short-Term Memory - Fully Convolutional Networks (MLSTM-FCNs) to predict aircraft engine types using ADS-B data. This study devises two experiments that alter the number of training data samples and input features to examine the impact on the ADS-B classification model's predictive power. The first experiment adjusts the number of training data observations from a limited feature set and achieves an accuracy of 83.9 % (within 10% of prior efforts with only 25% of the data). The results suggest that feature selection and data quality have a greater impact on classification accuracy than data quantity. The second

experiment accounts for all ADS-B feature combinations, discovering that airspeed, barometric pressure, and vertical speed have the greatest influence on aircraft engine type prediction.

The second study uses a larger data set and longer flight segments to perform more specific aircraft prediction. This is dissimilar to study 1 which predicts engine type using data from the first 10 minutes after takeoff. The utility of using only the first 10 minutes of flight is limited due to ADS-B sensors requiring line-of-sight to collect the data. Collecting just the takeoff segment of flight would require access to the airfield where it departs. Study 2 uses all phases of flight to make predictions. Additionally, since more data is available, study 2 predicts an aircraft's Wake Turbulence Category (WTC), description and type designator to provide more specific details on the aircraft.

In the third study, the classification accuracy was improved from study 2 with sensor fusion. Study 3 combines image and weather data with the ADS-B kinematic features using both pre- and post-processing methods to determine the best technique for classifying aircraft from multiple data sources. With this research, intelligence analysts would be able to use these techniques and ADS-B data as a surrogate for combining kinetic aircraft data with images and weather to predict aircraft.

In many ways, the previous approaches used to classify ADS-B data have attempted to do so without a clear understanding of how the information within the data is related. Early attempts did not account for the large class imbalance in the data set which caused overall accuracy to be falsely inflated, but caused accuracy between minority classes to be low. Later attempts used only a few subsets of the feature space which produced improved accuracy, but potentially less efficient models. This research takes a step back to better understand the ADS-B data set by employing various data engineering techniques and attempting to predict more specific information about the aircraft in the data set.

Acknowledgments

First, and most of all, I would like to thank my family for their patience and support over the last few years. Without them, none of this would be possible. I would also like to thank Major Dill, Dr Hodson, and Dr Grimaila for their advice and mentorship. Discussing problems and bouncing ideas helped to improve my studies, research, writing and coding. Thank you to the Air Force Research Laboratory for their prior research and financial support. Finally, thank you to all of the researchers whose papers helped me learn what I needed in order to continue my PhD studies.

Sarah J. Bolton

Table of Contents

Page

List of Figures

List of Tables

Pattern-of-Life Modeling with Automatic Dependent Surveillance-Broadcast (ADS-B)

I. Introduction

From 15.8 exabytes in 1993 to 6.8 zettabytes in 2020, storage on the internet has surged by more than 40,000%, according to [21]. As of May 2023, the size of the internet is roughly 24 zettabyes and doubles every two years [22]. To put this in perspective, if we assume the typical personal computer (PC) has a one-terabyte hard drive, 24 zettabytes is equivalent to 24 billion PCs, which is enough for every individual on the planet to have three PCs. While much of this information is personal data, a considerable percentage is considered publicly available information (PAI) and can be used by any internet user, business or government entity.

Due to the rise in available data, the study of recognizing trends (i.e., data analytics) has become more popular in a variety of areas of society, including business and government. Researchers and large organizations have developed a number of strategies for optimally utilizing this massive resource, fittingly termed Big Data. Potential fields include Internet of Things (IoT) analysis [23, 24, 25], traffic modeling [26], flight and maritime movement [27, 28, 29, 30, 31], image recognition [32], search engines [32] and natural language processing [32].

In recent years, multi-domain operations has become a critical, yet complex part of military defense operations. For that reason, military leaders are becoming aware of the rising utility of PAI and data analytics. By combining PAI with the multitude of other available sensor data, such as from intelligence, surveillance, and reconnaissance platforms, it is possible to enhance the predictive capability of those assets.

According to the FY22 Posture Statement, "real-time dissemination of actionable information" is essential for "joint warfighting across all domains at a pace faster than our competitors" [4]. Without recent technological, artificial intelligence, and machine learning advancements, this objective would be nearly unattainable. The good news is that there are new approaches that can be used to filter the noise in big data much faster than humans can, allowing military decision makers to make crucial decisions rapidly.

The Air Force Research Laboratory Layered Sensing Exploitation Branch (AFRL/RYA) has been tasked with the mission of improving military operations within the Distributed Common Ground System (DCGS) by providing enhanced capabilities for a major user of big data: intelligence analysts. The results of their work helps military leaders with the analysis of the immense data and improves military operations overall. Pattern-of-life (POL) modeling, understanding the habits, behaviors and movements of an entity or entities, is one of the goals of AFRL/RYA [33, 34, 35, 36, 37]. Recent research interests recommend analyzing ground-based and onboard aircraft sensors with deep learning in order to predict aircraft characteristics from only the kinematic data. This research interest intends to provide more tools for intelligence analysts and military leaders, and hopes to remove the sole reliance on secondary radar for aircraft identification purposes since it can be easily altered or unavailable.

One branch of POL modeling research focuses on using kinematic Automatic Dependent Surveillance-Broadcast (ADS-B) data to create aircraft-related predictions [26, 28, 20, 31]. ADS-B transmissions can be collected when aircraft broadcast their ADS-B Out data via an onboard transponder. The broadcast includes kinematic and aircraft identifying information such as location, altitude, ground speed and International Civil Aviation Organization (ICAO) hex identifier. Due to the inclusion of identifying information, the data provides an abundance of truth data for classification purposes whereas primary radar does not. In many locations around the world, Aircraft are required to broadcast ADS-B which makes ADS-B an attractive alternative to other data sources for classification and model building [8, 38]. Since the data is openly broadcast without encryption, ADS-B data is collected at numerous locations around the world by enthusiasts and researchers that maintain a receiver to collect the data as it is broadcast publicly. Many ADS-B collectors upload their data to centralized repositories like the ADS-B Exchange [39], which aggregates the data and makes it available to the public. Statistical, kinematic and identifying information on the broadcasting aircraft are stored in these databases.

1.1 Benefit to Warfighter

Per the National Defense Authorization Act (NDAA) for Fiscal Year 2022, section 5201, the national defense science and technology strategy updated the NDAA from 2019 "to establish an integrated and enduring approach to the identification, prioritization, development, and fielding of emerging capabilities and technologies, including artificial intelligence-enabled applications" [1]. It is clear from this section that the United States government is focused on deep learning and machine learning in both the near and long term. The NDAA furthers this focus in section 5205 where the Secretary of Defense is required to submit a report detailing how the Department of Defense plans to compete in the global information environment [1]. One of most streamlined ways to meet both of these expectations is with POL modeling via deep learning. Additionally, a line of effort in the 2018 National Defense Strategy is to build a more lethal force [40]. Having a method to process large amounts of data faster than a human analyst ensures defense leaders have up-to-date information to improve the quality and speed of their direction.

The 2020-2023 Department of the Air Force (DAF) Posture Statements [2, 3, 4, 5] establish that the Intelligence, Surveillance and Reconnaissance (ISR) goal is to improve and optimize the Advanced Battle Management System (ABMS). The DAF would like to provide near real-time information to decision makers by modernizing the ABMS to enable forces to make decisions faster than the adversary is able to respond. Specifically, "we must achieve an operationally-optimized ABMS as the Department's primary contribution to Joint All-Domain Command and Control (JADC2)" [5]. Aircraft POL modeling would provide a way of detecting and predicting adversary aircraft characteristics and plans much much more efficiently than a human analyst could comb through the same amount of data. The approach suggested in this dissertation could be used to design models with other similar data sources.

1.2 Problem and Motivation

Several techniques and guidelines for POL modeling have been developed throughout the years [33, 34, 35, 36, 37]. Responsible for military actions during both peacetime and during conflict, military and defense personnel have a special interest in POL modeling that is more critical than civilian interests. With aircraft tracking, unattributed data from aircraft sensors can be problematic since the receiving entity can neither identify the aircraft nor its intent. To identify aircraft, secondary radar would be needed. Unfortunately, secondary radar is not passive and requires identity information to be provided by the aircraft. During conflict, an enemy aircraft willingly providing its identity is counterintuitive and unlikely. Despite this, with some analysis, it is possible to draw conclusions about the transmitting aircraft from the unattributed data. Prior studies have shown that simple kinematic flight data, like the data found in primary 3D radar which can detect azimuth, range and elevation, can be used to develop models that can predict details like engine type [16].

The problem with unattributed aircraft data such as Air Traffic Control (ATC) radar is that it lacks the associated truth data. ADS-B does not suffer from this limitation. For that reason, ADS-B data is useful for researchers since it can be used as a surrogate to build models for datasets without truth data. While aircraft intent continues to be a challenge, understanding more information about a particular unknown aircraft aids military and defense personnel in making tactical decisions during hostilities.

1.3 Research Objectives

This section discusses the goals and major research questions from each study in this dissertation. The questions are intended to support the overall objective of improving classification methods using aircraft based kinematic sensor data. Table 1 provides a summary of this section.

Table 1: Summary of Research Questions

ID	**Details**	**Chapter**
RQ-1	What is the minimum amount of data needed to effectively classify engine type within 10% accuracy of previous efforts?	III
RQ-2	What is the smallest feature subset that predicts an aircraft's engine type with an accuracy within 5% accuracy of previous efforts?	III
RQ-3	How does the identification of flight phase assist in aircraft prediction of description, designator and WTC?	IV
RQ-4	Which phases work best to predict aircraft by their WTC, aircraft description and type designator (as listed in ICAO Doc 8643)?	IV
RQ-5	Does fusing weather and image data improve the predictive power of a model that uses ADS-B data to predict description, designator and WTC?	V
RQ-6	Using pre- and/or post-classification fusion methods, how can a model be developed to predict description, designator and WTC using ADS-B, aircraft images and weather?	V

1.3.1 Study 1 - Feature Reduction and Data Minimization

There are up to 70 flight characteristics within the ADS-B dataset, but several features do not play a role in predicting aircraft characteristics. Determining which features are extraneous and can be removed is time consuming if done exhaustively. With a month of ADS-B data being approximately 300-400 GB of data, removing unnecessary data could reduce the processing time needed to classify the aircraft.

Training a model with restricted computing resources is challenging, if not impossible in some cases, due to the computational demands of deep learning classification. The quantity of the training data determines the amount of computational resources that are necessary to train a model. Therefore, it is critical to learn how to make efficient use of resources by reducing the quantity of information used to train the model without severely impacting the model's accuracy. To reduce the quantity of data, one can either use fewer training samples or fewer features. In this study, using aircraft kinematic data collected from the first 10 minutes after takeoff, we examine the impact of varying these factors when predicting engine type.

1.3.2 Study 2 - Aircraft Prediction

The differences in kinematic factors used to classify aircraft engine type are significant due to functional differences. Jet engine aircraft fly much faster and higher than a piston or turboprop. Therefore, the kinematic output of a jet engine compared to a turboprop or piston aircraft are linearly separable with few dimensions. The variation in kinematic features between aircraft models or types are less evident since they share many similarities. This study looks at techniques to predict WTC, aircraft description and type designator as listed in ICAO Doc 8643.

As part of this study, we predict flight phases. Previous research in engine type classification used only the first 10 minutes after takeoff to ignore the differences in aircraft capabilities during different flight phases. Since we use the entire flight path, the differences during each phase can no longer be ignored when comparing aircraft models. Adding the flight phase as a feature to the models allows for the models to account for differences such as altitude, pressure and speed during takeoff or descent when comparing flight during cruise or level flight. This research uses a fuzzy logic method to classify flight phases by ground, climb, descent, cruise and level flight.

1.3.3 Study 3 - Sensor Fusion

Previous sensor fusion research with ADS-B has provided use cases on how ADS-B could be combined with electro-optical, LiDAR and radar sensors to develop an unmanned aerial system traffic management system, but has not developed a functional system [41]. Hobbyists have developed tools that use ADS-B data to graphically track aircraft around the world [42]. However, these tools assume that the aircraft's ADS-B Out broadcast is transmitting the correct ICAO information. During wartime or covert operations, this assumption would be invalid, because military aircraft would disable the ADS-B broadcast. This experiment combines weather, images and the ADS-B kinematic features to predict the airframe. The hypothesis is that having additional sources of information would allow a model to make predictions with greater confidence.

There are two areas where sensor fusion can occur: pre-classification or post-classification. In pre-classification, features are combined before creating the model. In post-classification a voting/weight system is used to determine which model's output takes precedence. This study uses both techniques; pre-classification is used to fuse ADS-B and weather and post-classification is used to combine ADS-B, weather and images.

1.4 Assumptions and Limitations

Previous research completed by Basrawi et al. used the takeoff segment to make predictions [20]. In contrast, this research applies the methods developed by [43, 18, 44] to all segments of the flight. Making predictions from any flight segment is important since ADS-B requires line-of-sight to capture the transmission; only ADS-B receivers physically located at or near airports or satellite receivers could capture data during the takeoff segment. Having only the option to make predictions on the takeoff segment of the flight reduces the power of the model by limiting the input data to about 10 minutes of the total flight. The models developed in this research allow predictions to be made about the transmitting aircraft at any point during its flight which could reduce the accuracy, but provides more flexibility to use flight data from any point in time.

This research intends to classify aircraft with more specificity than previous experiments. The models in this research predict the airframe model instead of engine type [20, 28], trajectory [29], flight phase [43, 18, 44] or military vs commercial aircraft [45, 46]. Aircraft images have been shown to be more powerful predictors of airframe type than kinematic data [17, 47].

1.5 Dissertation Structure

This document is organized into various chapters based on each completed study. This chapter, Chapter I, discusses the research problem, motivation and objectives for this dissertation. Chapter II provides a literature review and background information on pertinent deep learning topics, various approaches to classifying ADS-B data, and a discussion of dataset size requirements for deep learning model building. Chapters III, IV and V detail papers related to the research questions outlined in this dissertation. Chapter VI summarizes the findings of this dissertation, discusses the contributions and suggests future research.

II. Background

This chapter covers the challenges and relevant literature review for aircraft classification using Automatic Dependent Surveillance-Broadcast (ADS-B) data. This chapter begins with a background on ADS-B and the International Civil Aviation Organization (ICAO) in section 2.1. The section reviews how ADS-B is collected, its format and relevant regulations. In section 2.2, information about pertinent deep learning techniques are discussed. The deep learning section focuses on algorithms that have been shown to be useful with time series and image data such as Recurrent Neural Networks (RNNs), Long Short-Term Memorys (LSTMs), and Convolutional Neural Networks (CNNs). After a brief overview of those algorithms, the section further discusses techniques and guidelines that have been combined with those deep learning algorithms to improve their resultant models. Section 2.3 discusses previous research efforts in classification with ADS-B and it's maritime variant, automatic identification system (AIS). This section reviews engine classification and discusses previous research in flight phase classification. The final section, section 2.4, discusses previous research in sensor fusion and appropriate fusion techniques. Table 5 outlines the papers discussed in this chapter.

2.1 Aircraft Standards

Each country across the globe is responsible for regulating their own airspace. The differences in priorities among nations can result in different rules depending where aircraft fly. Since aircraft will often fly over countries other than their departure point and destination, the ICAO and other organizations have standardized airspace rules as much as possible. One such way is ADS-B. While not required everywhere, most modern airspace has some form of ADS-B regulations. As a method of identifying aircraft, the ICAO developed a 24-bit address for each specific aircraft that is often expressed in hexadecimal notation. The ADS-B broadcast includes the ICAO 24-bit address and other information that can be found in the ICAO Doc 8643 [48]. In order to understand how ADS-B can be used to classify aircraft, it is important to be aware of ADS-B usage and the standards that guide airspace and aircraft regulations.

2.1.1 Automatic Dependent Surveillance-Broadcast

Used and deployed around the world, ADS-B is an important aircraft sensor that provides detailed information from the broadcasting aircraft. It is a form of secondary radar where the aircraft actively broadcasts the data. This is dissimilar to primary radar which does not require active interaction from the aircraft. Several studies have made use of ADS-B data to identify flight patterns [28, 31, 43, 18], improve aircraft operations [49, 50], increase ADS-B transmission security [51, 52] and classify targets [27, 20, 53]. The signals are broadcast publicly through an ADS-B Out transponder at 1090 MHz in several nations and at both 1090 MHz and 978 MHz within the US. Various methods are employed to receive and collect ADS-B data to include other aircraft, ground stations, and satellites. An ADS-B communications network can be seen in Figure 1.

This dissertation utilizes ADS-B transmissions received by ground stations. Ground stations are locations around the world where either commercial, government or even civilians setup antennas and receivers to collect the ADS-B broadcasts. Often, the broadcasts collected from these receivers are transmitted to various ADS-B repositories. These repositories host the data to the public for either free or paid usage. The data in this dissertation is obtained from one of those repositories, the ADS-B Exchange [39]. To contribute to the efforts of the ADS-B Exchange, the Air Force Institute of Technology (AFIT) and Air Force Research Laboratory (AFRL) main-

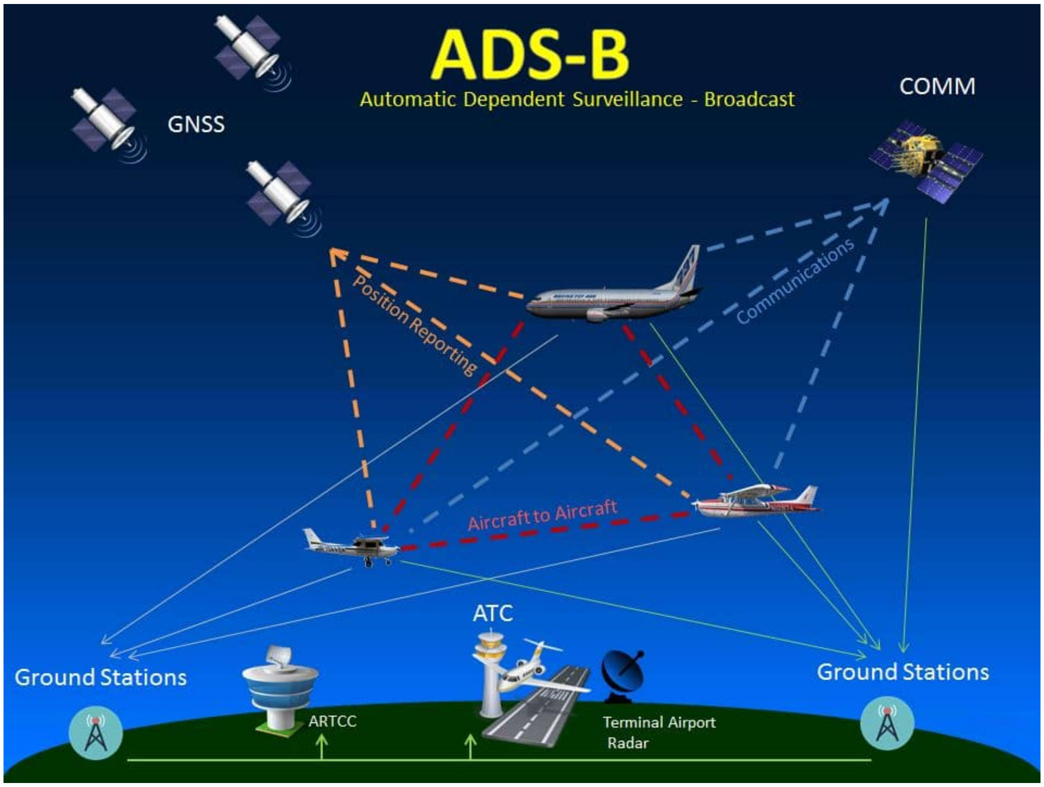


Figure 1: ADS-B High Level Operational Concept Graphic [6]

tain ground stations located near the National Center for Medical Readiness (Calamityville) in Fairborn, OH. The AFIT and AFRL ground station can be seen in Figure 2

ADS-B Exchange and other similar ADS-B hosting sites save the incoming data as JavaScript Object Notation (JSON) files at two to five second intervals. Each transmission includes up to seventy pieces of information about the transmitting aircraft, such as the ICAO hex identifier, altitude, airspeed, vertical speed, directional heading, position time, latitude, longitude, and ground status, in accordance with the United States' DO-260B standard and Europe's ED-102A standard [39, 54].

Figure 3 is a map of the countries that have some form of ADS-B usage mandate as of February 2022. Table 2 provides details on the countries that have an ADS-B mandate and the date when the mandate began or is projected to start. Table 2 illustrates that the majority of nations, including the United States, Australia, the

Figure 2: The AFIT and AFRL ground station

European Union Aviation Safety Agency (EASA), and a number of East Asian and Pacific island nations, began enforcing the installation of ADS-B Out equipment on aircraft flying within their borders on or before 2020. Since a large number of aircraft fly through these regions, ADS-B Out equipment is installed globally.

Within the US, the mandate is only required within controlled airspace. It is not required for low flying aircraft and aircraft operating in or around of smaller airports classified by Class D airspace. Figure 4 provides a more in-depth explanation of the US standards. Aircraft operating in the United States above 10,000 feet mean sea level (MSL) or in Class B or Class C airspace (close to major airports) are required to have ADS-B Out installed.

Table 2: ADS-B Mandate Timelines [19]

8. 1 Global ADS-B Out mandate summary

Mode S Transponders with Extended Squitter capability in accordance with RTCA DO-260 or DO-260B:	
Australia mandate for retrofit aircraft	12-Dec-2013
Singapore mandate for all aircraft	12-Dec-2013
Indonesia mandate for all aircraft	01-Jan-2018
Hong Kong mandate for all HKG-registered aircraft - Flying PBN routes L642 or M771 between FL290 and FL410 - Flying within HKG FIR between FL290 and FL410	 31-Dec-2013 31-Dec-2014
Australia mandate for SA Aware GNSS (new a/c)	08-Dec-2016
Mexico - SBAS GPS required as per FAA guidelines	January 1, 2020 (Pending)
Colombia - SA-ON acceptable	January 1, 2020 (Pending)
U.S. FAA mandate for 100% equipage - Class A, Class B and Class C airspace - 1090 ES (Extended Squitter) for FL180 and above - 1090 ES or UAT (Universal Access Transceiver) below FL180 - Must be compliant with TSO C166b (Mode S) or C154a (UAT)	01-Jan-2020
EASA mandate for *forward-fit and retrofit* aircraft	07-June-2020
China mandate for all aircraft	31-Dec-2022

2.1.2 International Civil Aviation Organization

The ICAO, an organization of the United Nations that manages international air navigation, represents 193 member states that agree on standardized and safe air operations. They do not hold authority over national governments and any regulations that conflict with those outlined by the ICAO supersede those created by the ICAO. It is the ICAO's responsibility to encourage diplomatic discussions between member states to resolve any discrepancies. Per their website, their mission is the following [55]:

> To serve as the global forum of States for international civil aviation. ICAO develops policies and Standards, undertakes compliance audits, per-

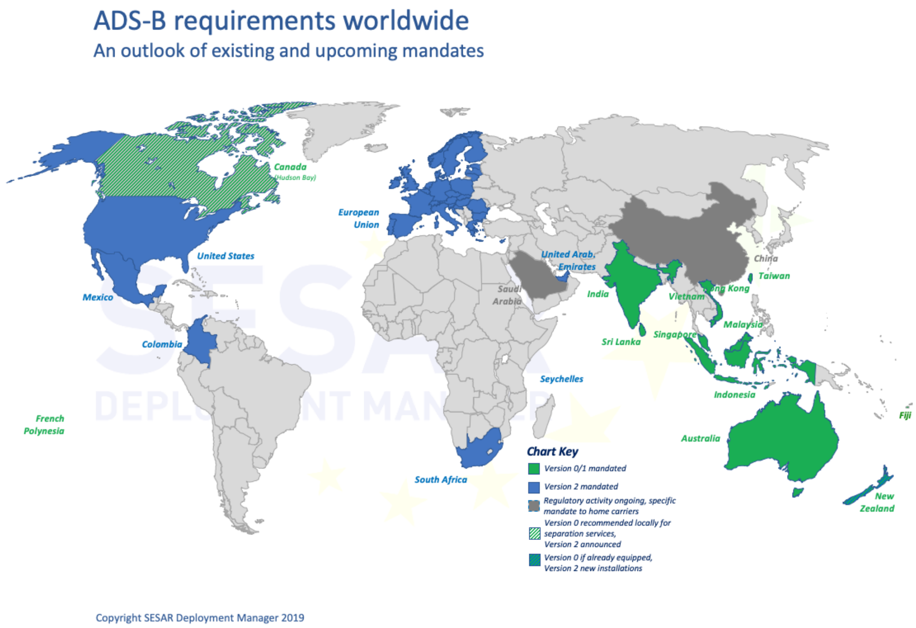


Figure 3: Countries that have mandated ADS-B usage [7]

> forms studies and analyses, provides assistance and builds aviation capacity through many other activities and the cooperation of its Member States and stakeholders.

The ICAO Doc 8643 outlines information about aircraft that are provided some type of air traffic service. Part 2 of the ICAO Doc 8643 provides a chart with aircraft designators, manufacturers, models, description and Wake Turbulence Category (WTC). In the ADS-B data set, the designator is referred to as the "type." Within the ICAO documentation, the designator is similar or the same as the aircraft model. For example, for all C-130J model aircraft, the designator is "C30J." All other C-130 models use the more complete designator "C130." Some designators are less obvious, such as the MC-12 Liberty which is "B350" due to the model it is based on: the Hawker Beechcraft Super King Air 350ER.

The description column in the ICAO Doc 8643 does not contain a written description of each aircraft, instead it provides each aircraft with its 3-letter categorization.

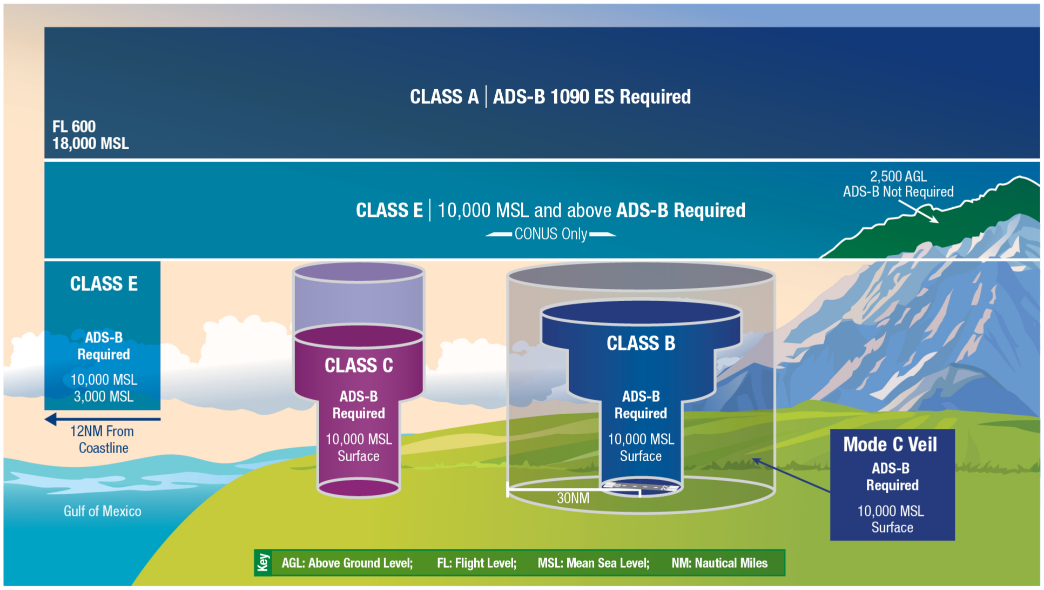


Figure 4: US ADS-B Airspace Requirements [8]

The first symbol indicates the aircraft type:

- L - Landplane
- S - Seaplane - only lands on water
- A - Amphibian - lands on both water and land
- G - Gyrocopter - rotors powered by airflow
- H - Helicopter - rotors powered by engine
- T - Tiltrotor - both horizontal and vertical landing, i.e. V-22 Osprey

The second symbol is the number of engines or the letter “C.” The “C” symbol is only used for fixed-wing aircraft where two engines are coupled to drive a singular propeller. The third symbol specifies the engine type.

- J - jet
- T - turboprop/turboshaft

- P - piston
- E - electric
- R - rocket

The WTC is based on the aircraft's maximum certified takeoff mass and is categorized by the following [56]:

- H (Heavy) aircraft types of 136 000 kg (300 000 lb) or more
- M (Medium) aircraft types less than 136 000 kg (300 000 lb) and more than 7 000 kg (15 500 lb)
- L (Light) aircraft types of 7 000 kg (15 500 lb) or less.

Within this research, we separate "Medium" into two categories. The first is 15,500 to 75,000 lbs and the second is 75,000 to 300,000 lbs. We also put high performance aircraft into their own category. High performance aircraft are those that can maneuver in excess of 5Gs and maintain an airspeed above 400 knots.

2.2 Deep Learning

Artificial Neural Networks (ANNs) have been shown to be successful at classifying a variety of data types. Many ANN variations have been developed over the years to classify the infinite number of data types researchers encounter. Since the classification of data is paramount to pattern-of-life (POL) modeling, it is important to understand how each of the variations are best used. A few of the more common ANN types include Feed-Forward Neural Networks (FNNs), RNNs, LSTM networks and CNNs. FNNs are the most basic type of ANN, RNNs were developed to classify time series data, LSTMs were an improvement upon the RNN, and CNNs were developed to classify images. A model of each of these ANN types can be seen in Figure 5. This section discusses the basics of these ANN types, reviews the variations researchers have made to improve these ANNs, and describes current literature in this area.

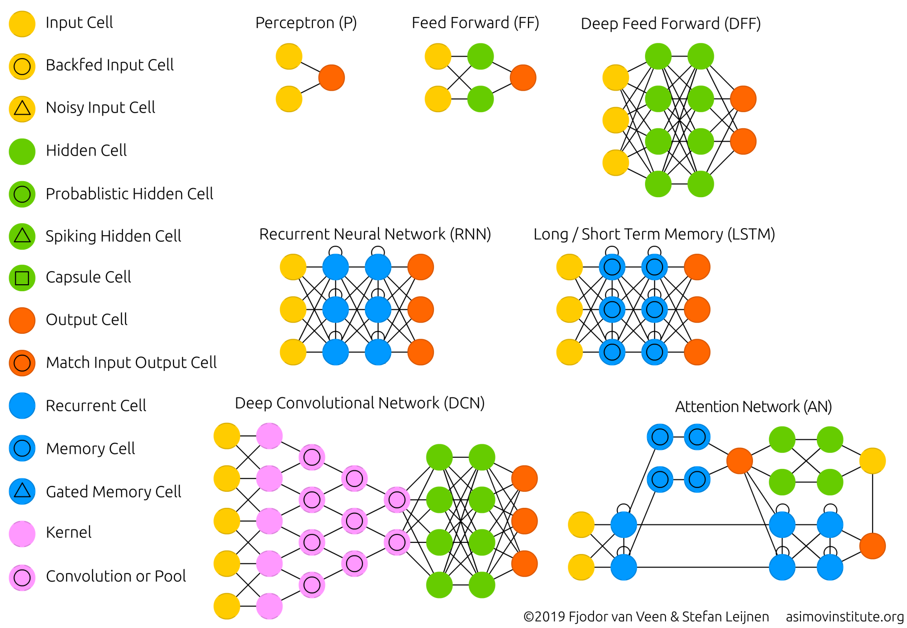


Figure 5: Neural Networks [9]

2.2.1 Feed-Forward Neural Network

An FNN is the simplest type of ANN and functions much like logistic regression. Unlike more advanced ANNs, information moves in one direction: forward. At minimum, the FNN has input nodes and output nodes. This type of FNN is called a perceptron. Data that is classifiable by a perceptron must be linearly separable. FNNs with input nodes, a hidden layer and an output node can handle more complex data.

When building a neural network model, each node is provided a starting weight. The training of an ANN involves optimizing the weights of each node to more accurately reflect the data that is being processed. The first step requires computing the weighted sum of the inputs, adding on the bias and then feeding that value into an activation function. If we let x be the vector of input values, w be the vector of weights, and b be the bias, the weighted sum of the inputs would be expressed in equation 1.

$$z = x \cdot w + b \tag{1}$$

Then, we feed the result from equation 1 into an activation function like the sigmoid activation function as seen in equation 2 to produce the output after propagating all of the data.

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

The next step, backpropagation or *backward propagation of errors*, trains the network by computing the gradient, or rate of change, of the loss function with respect to the weights to determine the direction the weight needs to move (positive or negative) resulting in a model with less error. There are various loss functions; the most common of these functions are mean square error for regression problems and cross entropy loss for classification problems. The loss function is calculated for the entire training dataset and the average loss (C) is computed. To determine the best weights for the next or final iteration of an ANN model, the gradient of the cost function with respect to the weights and biases must be found. This is done with partial derivatives and the chain rule as shown in equation 3.

$$\frac{\partial C}{\partial w_i} = \frac{\partial C}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w_i} \tag{3}$$

In the final step of training a neural network the algorithm updates the weights in the neural network to minimize error. It performs this algorithm repetitively using gradient descent as the model searches for the local or global minima. A graphic depiction of gradient descent can be seen in Figure 6.

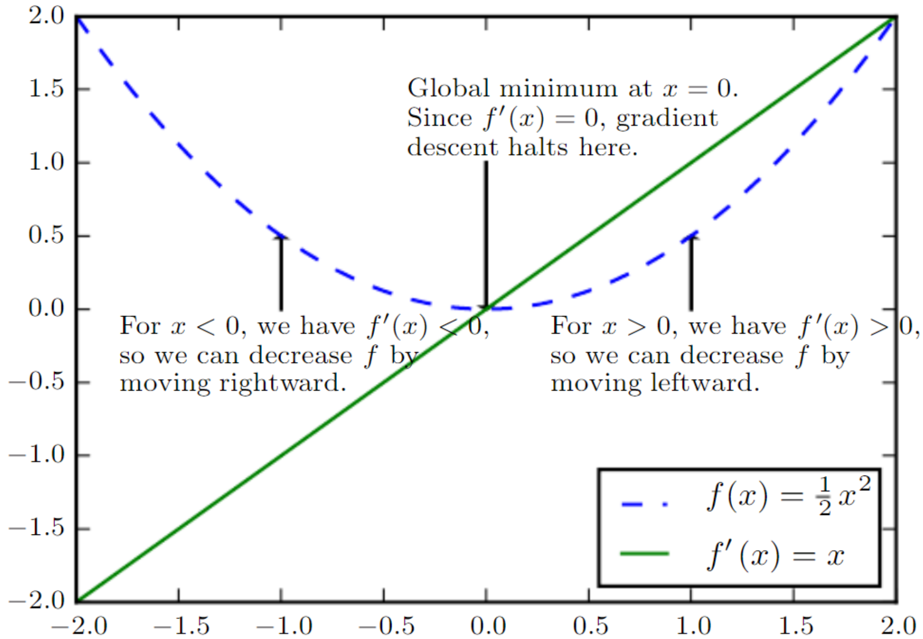


Figure 6: Gradient descent. An illustration of how the gradient descent algorithm uses the derivatives of a function to follow the function downhill to a minimum. [10]

2.2.2 Recurrent Neural Network

A simple type of network that allows for the prediction of time series or sequential data is the RNN. RNNs get their name by the way they move information: recurrently. In other words, RNNs process data with a feedback loop that allows the neural network to look back at data that preceded the data that is currently being processed. A matrix of weights are then updated for data within the RNN's recent past giving RNNs the ability to have a short term memory. Figure 7 depicts how RNNs work in comparison to the simpler, FNN.

2.2.3 Long Short-Term Memory Network

RNNs struggle to process longer sequences of data since longer sequences require more gradients to be multiplied during backpropogation. This problem causes vanishing and exploding gradients. In the case of vanishing gradients, during backpro-

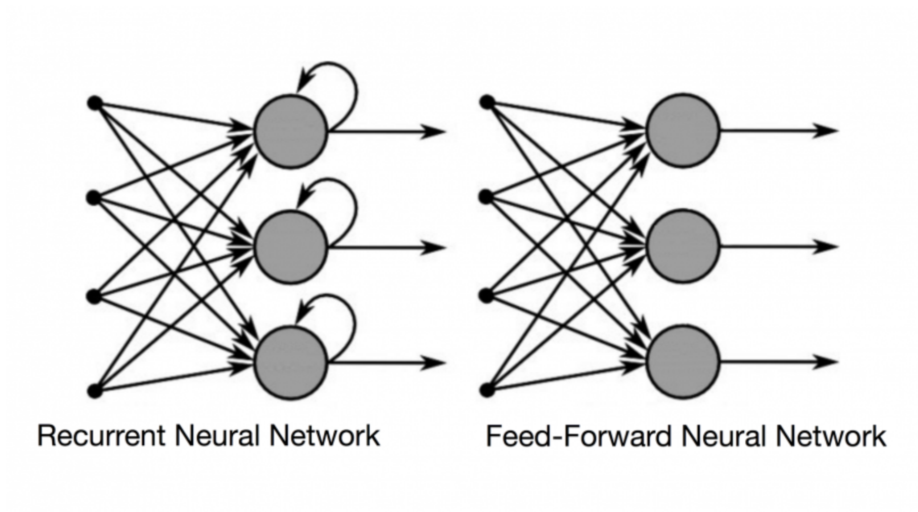


Figure 7: Recurrent Neural Networks and Feed Forward Neural Networks

pogation many small numbers (i.e. numbers less than one) are multiplied resulting in an even smaller number. When this small number is used to update the weights, the new weights do not change with statistical significance. This reduction in weight change stalls the improvement of the model. Exploding gradients cause the opposite problem.

LSTMs solve gradient problems by increasing the neural network's memory. LSTMs have feedback connections and gates that allow the neural network to decide what data is saved. Specifically, an LSTM has a forget, input, and output gate; each gate functions like its name implies. Figure 8 shows how an LSTM is setup. LSTMs solve the gradient problem by ensuring that the derivatives of the weights and activation functions stay relatively constant. It does that by allowing the gates to help decide when to increase or decrease the gradient [10, 57, 58]. This mechanism ensures the gradient never vanishes or explodes.

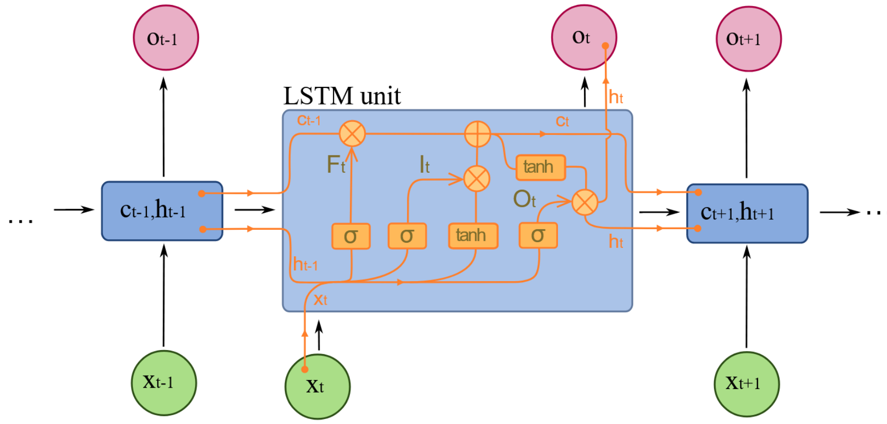


Figure 8: Long Short-Term Memory Network Diagram [11]

2.2.4 Transformer Network

One major issue with both RNNs and LSTMs is the amount of time required to train due to the fact they take input sequentially. In fact, while LSTMs can process longer sequences than RNNs, LSTMs are slower than RNNs due to their increased complexity. Since most modern computers are designed to parallelize data, obtaining input sequentially is an ineffective use of a computer's processing power. If the entire sequence could be read at once instead of one input at a time, this would allow the data to be processed in parallel and be a more effective use of the graphics processing unit (GPU).

Designed in 2017 with research from Google Brain, parallelization is where the transformer network excels [12]. The transformer network removes instances of recurrence from the RNNs and replaces them with attention mechanisms. The benefit in this design is that with the removal of recurrence, the current input in a sequence is not dependent on its predecessors. Figure 9 shows the architecture design for the network. One of the main components of the transformer architecture is the multihead self-attention mechanism where each element in the sequence is compared to every other element in the sequence. To do this, each input in a sequence is encoded into a $n \times 1$ vector with values based on meaning and importance. Then, the vectors are processed simultaneously by the network to provide an output. This provides a faster, albeit data hungry, method for processing sequential data [12].

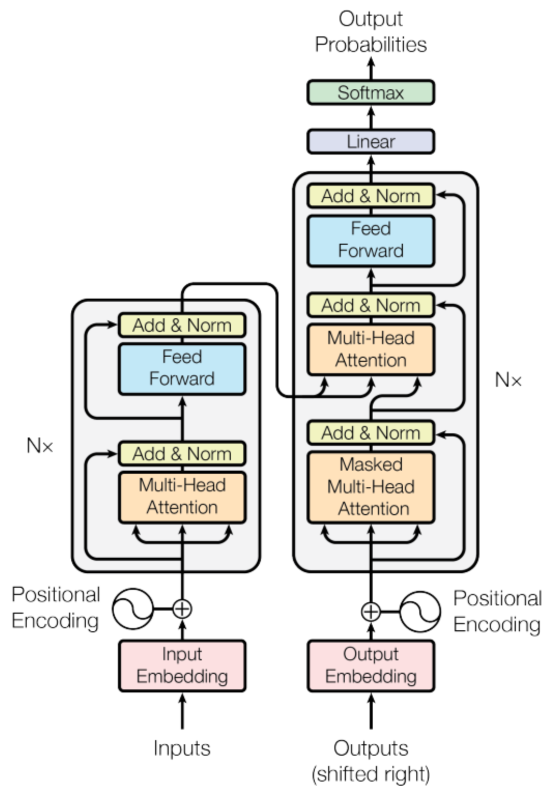


Figure 9: The Transformer architecture [12]

2.2.5 Convolutional Neural Networks and Fully Convolutional Networks

CNNs specialize in processing data with a grid-like topology such as time-series data and images [10]. Like RNNs, CNNs get their name from how they function. CNNs use mathematical convolutions, a form of matrix multiplication, to process data. CNNs were modeled after how brains evaluate visual input by using convolutions to extract high-level features from images such as edges. The formal definition of a convolution is shown in formula 4.

$$(x * w)(t) := \int_{-\infty}^{\infty} x(a)w(t-a)da \tag{4}$$

The input, x, is a multidimensional array or tensor of the grid-like input data. The filter or kernel, w, is another tensor that extracts features from the data. The output, or feature map, is the convolution of x and w. A visual example of how the convolutions work can be seen in Figure 10. Essentially, the kernel is a sliding window where matrix algebra is performed on the input to create a new matrix.

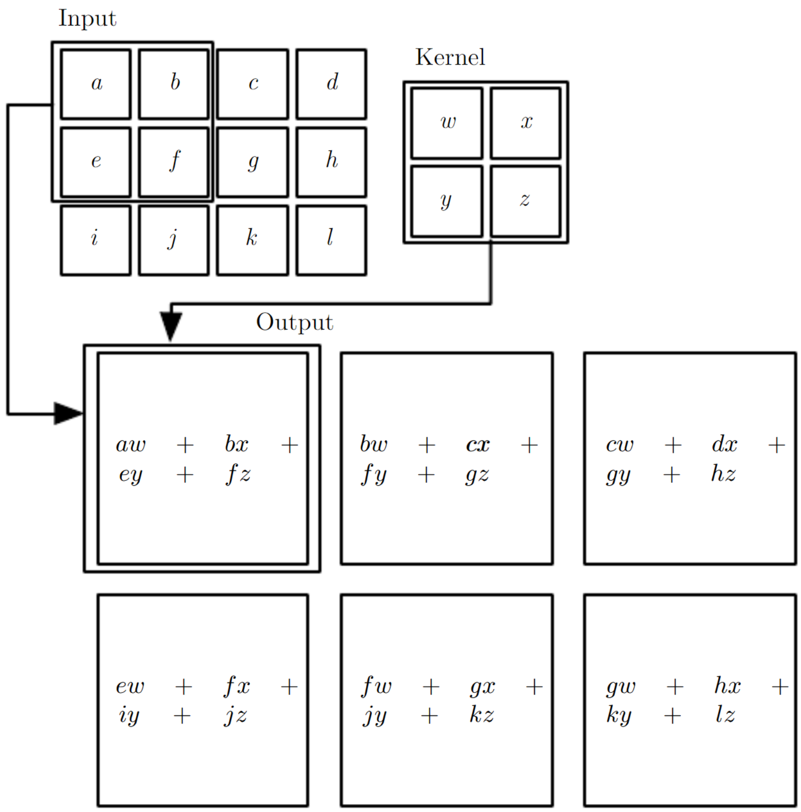


Figure 10: Example of a 2-D Convolution [10]

Another important component of CNNs is pooling. Simply, the pooling function provides a summary of the output by merging adjoining values. This helps to reduce the dimensionality of the feature map. The areas that are combined are adjacent or in the same neighborhood. There are a few types of pooling, but the two most common are max pooling and average pooling. Max pooling reports the maximum output of the neighborhood as the summary and average pooling reports the average value. Typically, max pooling is suggested over average pooling since it works as a noise suppressant for values that are unrelated to the rest of the neighborhood. A graphical representation of pooling is shown in Figure 11.

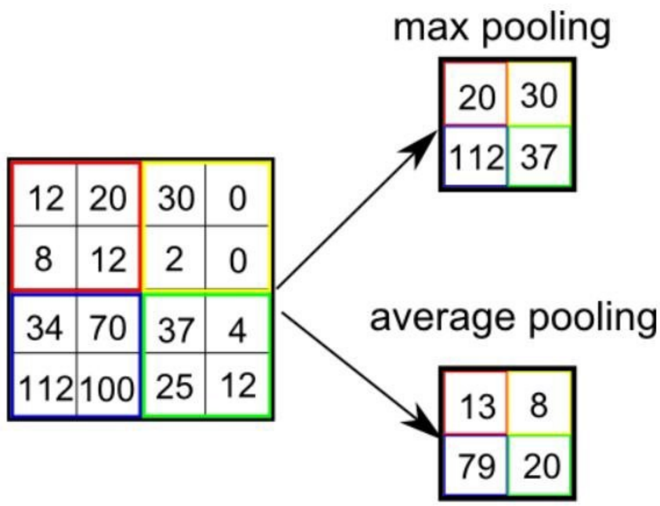


Figure 11: Example of Pooling [13]

CNNs are made up of multiple convolutional layers and pooling layers depending on the complexity of the network and input data. To ready it for classification purposes, it must be flattened into a vector and fed into a FNN. This FNN is referred to as the fully connected or dense layer.

An Fully Convolutional Network (FCN) is a type of CNN without any fully connected layers. Instead, it uses convolutional layers to format the data into an appropriate format for classification. There are a few advantages to FCNs. One benefit to the FCN is that it allows arbitrary input sizes [59]. The dense layers in a traditional CNN expect a static input size, but without dense layers, a static input size is not a requirement. Another advantage of the FCN is that it has been shown to require less preprocessing or feature engineering [60, 61].

2.2.6 Squeeze and Excite Blocks

While LSTMs and FCNs are well established, the use of squeeze-and-excite blocks is a newer technique. Squeeze-and-excite blocks were introduced by Hu et al. in 2018 to improve the representational capacity of a neural network [14]. They are used with convolutional networks to help model the dependencies between the channels. As the name implies, it consists of a squeeze mechanism followed by an excitation mechanism. In the squeeze step, the classifier uses global average pooling to aggregate the spatial information of each channel. Using the example of an image, where an image has dimensions of $H \times W \times C$ for height, width and (normally 3 color) channels, the squeeze step would pass the image through a global average pooling operator where it would become a shape of $(1 \times 1 \times C)$. Then, the excite step uses a gating mechanism to capture channel-wise dependencies [15]. The excitation step uses a multi-layer neural network with one hidden layer. The input and output layer are the same shape $(1 \times 1 \times C)$, but the hidden layer reduces the space by a reduction factor, r, making the number of neurons in the hidden layer C/r. Karim et al. and Hu et al. use a reduction factor of 16 [15, 14]. This $(1 \times 1 \times C)$ value is multiplied element-wise with the original $(H \times W \times C)$ input. The graphical representation of the Squeeze-and-Excite block developed by Hu et al. can be seen in Figure 12. Implementing it with a residual network can be seen in Figure 13.

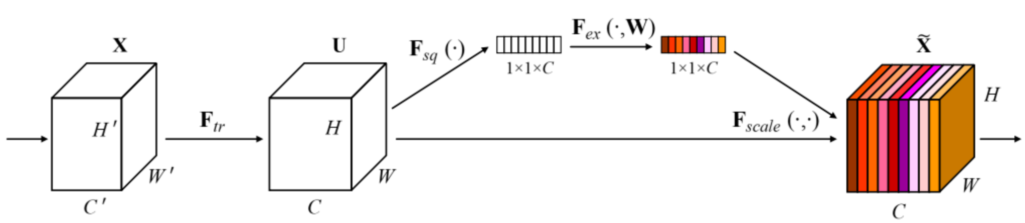


Figure 12: Squeeze-and-Excitation block [14]

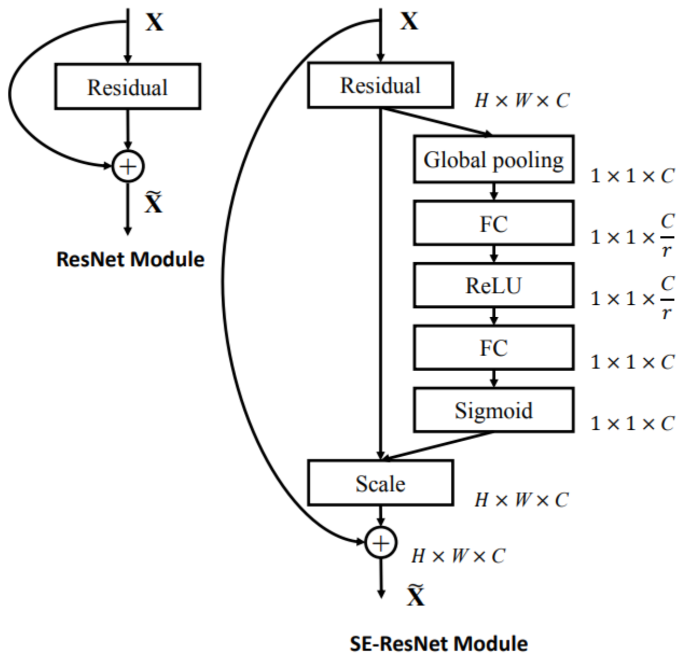


Figure 13: The schema of the squeeze and excite model with a Residual Network [14]

2.2.7 Multivariate Long Short-Term Memory - Fully Convolutional Network

The Long Short-Term Memory - Fully Convolutional Network (LSTM-FCN) developed by Karim et al., combines FCNs, LSTM networks, and squeeze-and-excite blocks. The algorithm improves upon previous methods to classify time series data [61]. FCNs were utilized to allow for the CNN benefit of class action maps without the requirement of extensive hyperparameter preprocessing that is normally required by a CNN. LSTMs were selected to detect the importance of sequences of observations in time series data. Squeeze-and-excite blocks were added to the algorithm to improve the classification power of multivariate data sets by ensuring feature maps have a similar impact to subsequent layers. In their study, Karim et al. tested four variations of the algorithm against 35 different data sets to include voice, human signal monitoring, and other time series data. Their algorithms outperformed the current state-of-the-art algorithm in 27 out of the 35 data sets [61]. Two variations of the algorithm, the Multivariate Attention Long Short-Term Memory - Fully Convolutional Network (MALSTM-FCN) and Multivariate Long Short-Term Memory - Fully Convolutional Network (MLSTM-FCN), were created a few years later to handle multivariate data. This algorithm was modified further and used to predict aircraft engine type using ADS-B data [16, 20]. Figure 14 shows the architecture Karim et al. designed for their research.

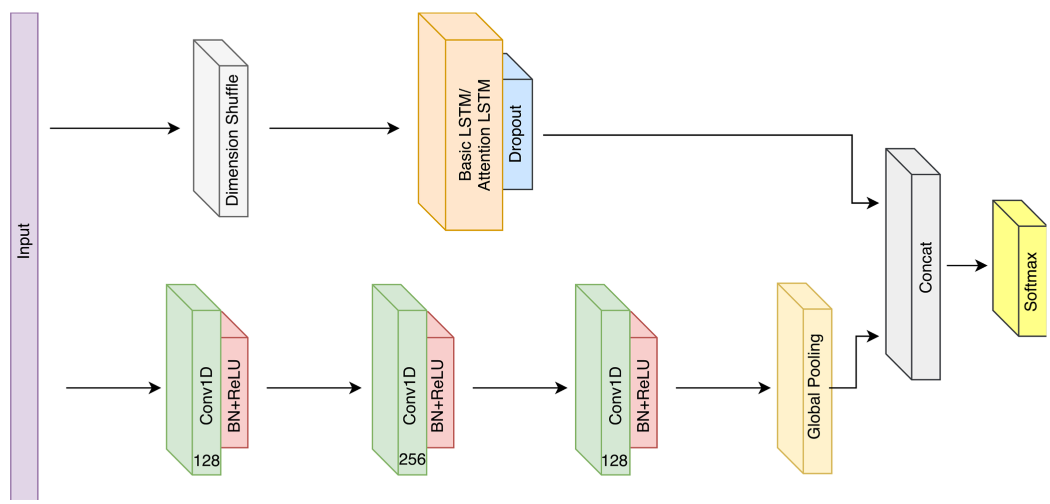


Figure 14: MLSTM-FCN architecture [15]

2.2.8 Data set Size Guidelines for Deep Learning

Knowing the amount of data needed to train and test an ANN model is a difficult problem. Thus far there is no specific formula to provide the number of features and observations needed to make the most effective predictive model.

Determining the exact data set size needed to build an ANN classifier is an open research problem that may never have a complete solution due to the infinite number of ANN combinations. However, experts recommend following a few rules of thumb to assure an adequate amount of data with a high level of accuracy [62, 63, 64, 65, 66]. There are three key variables when it comes to selecting the data set size: prediction classes, features and weights. The prediction classes are defined as the number of choices for prediction. In the case of engine type, we have three: jet, piston and turboprop. Features are the number of things you know about each observation. In the case of ADS-B, each observation has up to 70 features. A few of the common ones are speed, altitude and location. Weights are one of trainable parameters in a neural network. They are, essentially, a value attached to each feature that expresses the importance of that feature. The guidelines for each of these key variables is summed up as follows.

1. **Prediction Classes**: There should be 50-1000 times as many observations as prediction classes [63].

2. **Features**: There should be 10-100 times as many observations as features [64]

3. **Network Weights**: The sample size should be 10 times the number of weights in the network [65, 66]

   (a) A more rigid restriction, requiring 50 times as many observations as network weights, was proposed in a different paper [62]

2.3 Aircraft Classification Research

Multiple researchers have focused on using the ADS-B data set to predict engine type with ADS-B kinematic data [16, 20, 27, 28]. Earlier attempts used Gradient Boosting [27], Support Vector Machines (SVMs) [28] and Random Forests (RFs) [28]. A more recent study developed the Dual-Stage Deep Engine Classifier (DSDEC) using a modified MALSTM-FCN to improve classification accuracy [16, 20].

2.3.1 Statistical Features Extraction and Boosting

One of the first studies to classify aircraft using kinematic data was completed by Ginoulhac et al [27]. However, the focus of their work and research was on maritime classification. In their paper, the researchers used ADS-B and AIS data to identify air and maritime traffic using kinematic data with an 87% and 86% success rate, respectively. For the aircraft classification results, they noted there was a significant class imbalance towards jet engines which likely contributed to the high accuracy rate of the ADS-B model.

The paper explained that multiple machine learning techniques have been used to tackle target classification problems. A few successful methods of maritime classification included the use of RNNs [67], RF with kinematic features [68] and RF with geographic and behavior features [29]. The algorithm proposed in the paper computed statistics (maximum, minimum, median, etc) to describe the kinematic properties of the target. Then, they applied a Gradient Boosting classifier on those features. Their model was able to output a prediction at each time step of the observation. The advantage of this method was that it could be used in real time.

2.3.2 Dual-Stage Deep Engine Classifier

One of the limitations with analysis on the ADS-B data set is that while the models have very few problems determining if the aircraft is a jet, they tend to have trouble predicting the difference between turboprop and piston engines. The reason for this is twofold. First, the data set is heavily imbalanced towards jet engines which causes the data to have fewer samples of turboprop and piston engines from which to learn. Second, the kinematic differences between piston and turboprop engines are minimal.

As a method to remedy this problem, Basrawi et al. developed a DSDEC [20].

Using the MALSTM-FCN algorithm as a basis for the model, the DSDEC algorithm employs a unique 2-stage approach. When using the model to make predictions, the first stage predicts if the aircraft is a jet or not a jet. Then, the “not jet” predictions are fed to the second stage. The second stage predicts if the aircraft has a piston or turboprop engine. The results from both the first and second stage are combined to provide an engine prediction for each observation. Figure 15 portrays the architecture that is used.

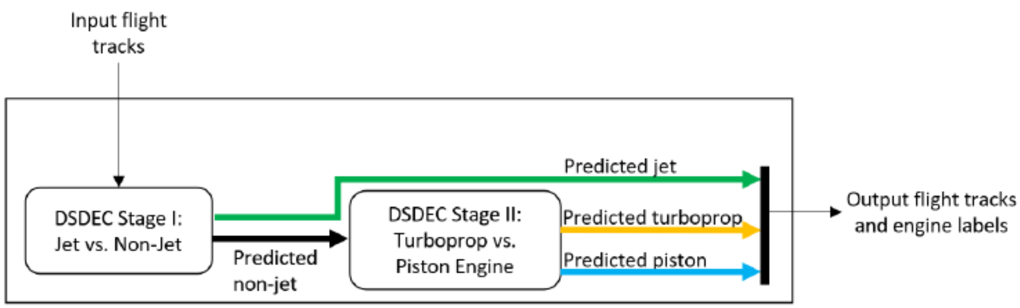


Figure 15: DSDEC architecture [16]

Basrawi et al. is the first to classify aircraft engine type using ADS-B data with a deep learning model. Using the DSDEC method, researchers were able to identify jets with 98.4% accuracy, turboprop aircraft with 79.2% accuracy, and piston engine aircraft with 89.9% accuracy. However, the DSDEC method would still confuse turboprop aircraft as piston engine aircraft 17.9% of the time [20]. Table 3 shows the results achieved by Basrawi et al. compared to a SVM and RF as baseline from previous research [28].

During the experiments, the researchers used static time steps (300 and 100 time steps) and each time step was separated by two seconds. Three hundred time steps would be equivalent to 600 seconds or 10 minutes of flight time. Similarly, 100 time steps would be 200 seconds or 3 minutes and 20 seconds of flight time. While Basrawi et al. indicated that more research would be needed to determine the influence of

Table 3: Results from Basrawi et al. [20]

Classifier	**Overall Accuracy**	**Jet Accuracy**	**Turboprop Accuracy**	**Piston Engine Accuracy**
DSDEC (300 time steps)	***89.2%***	***98.4%***	***79.2%***	***89.9%***
SVM (300 time steps)	77.4%	85.5%	72.2%	74.6%
SVM (100 time steps)	68.6%	81.1%	56.6%	68.1%
RF (300 time steps)	83.4%	94.1%	70.5%	85.8%
RF (100 time steps)	76.7%	90.0%	61.5%	78.6%

the time step size, their results point to the possibility that longer flight observations improve the model's predictive power. In fact, while the DSDEC algorithm was used against 21 different models to determine which hyperparameters led to the highest accuracy, none of the 100 time step models performed as well as the 300 time step examples.

2.3.3 Image Classification

Understanding how to classify aircraft images is critical for sensor fusion with ADS-B data. Classifying airplane models from images has been done successfully using a variety of methods. An image benchmark, gathered mostly from `https://www.airliners.net/`, was created by Maji et al. in 2013 to serve as a starting point for future researchers interested in aviation image categorization [17]. Nearly a thousand papers on aircraft classification have referred to the image database created by Maji et al known as the Fine-Grained Visual Classification of Aircraft (FGVC-Aircraft) benchmark. The benchmark includes 10,000 airplane images from 100 different aircraft models. Each image is organized in a hierarchical structure by model, variant, family and manufacturer. A graphical depiction of this hierarchy can be seen in Figure 16.

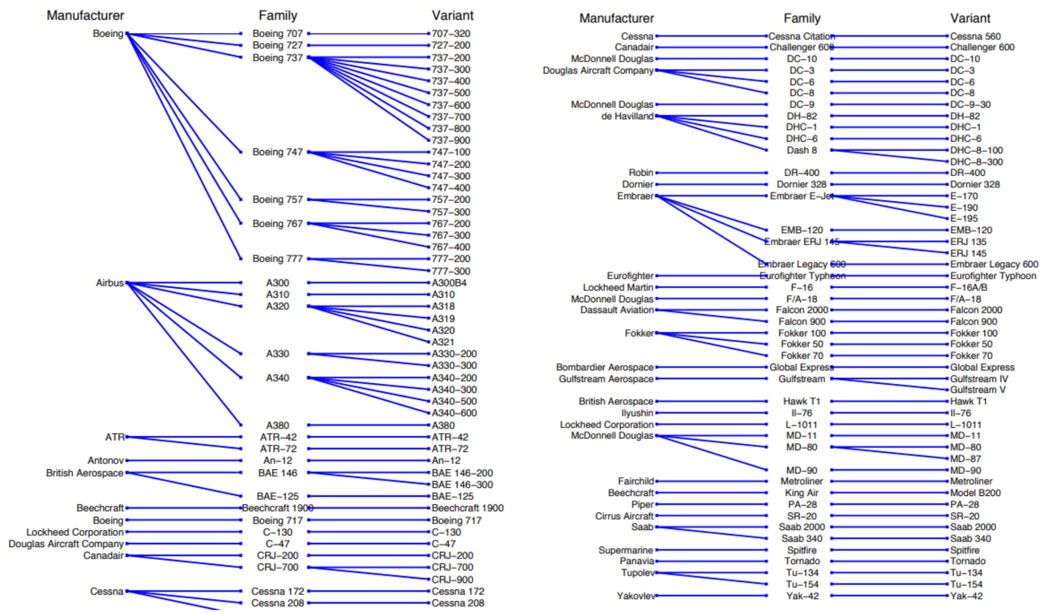


Figure 16: FGVC-Aircraft benchmark hierarchy shown as manufacturer, family and variant. The data set contains aircraft of 100 different variants grouped under 70 families and 30 manufacturers [17]

Maji et al. achieved a 48.69% variant classification accuracy with a non-linear SVM in their initial attempt [17]. Recent research has improved the accuracy to around 90%. Rong et al. achieved 93.3% accuracy using a Separated Smooth Sampling Network (SSSNet) [69], Yanfeng Wang et al. achieved 89.2% accuracy using a Bat Algorithm and Convolutional Neural Network (BA-CNN) [70], Lei Wang et al. achieved 91.4% accuracy using a Multilayer Feature Fusion (MFF) network with Parallel Convolutional Block (PCB) mechanism [71], and Liu et al. developed a Cross-Layer Mutual Attention Learning Network (CMAL-Net) with 94.7% accuracy [72]. The best performing of these networks, the CMAL-Net, uses a Residual Network (Resnet) as its backbone, which uses skip connections to solve the exploding and vanishing gradient problem. Multiple classifiers are used by the CMAL-Net and then trained to make predictions based on a particular layer from shallow to deep. The output of the entire model is the product of combining the predictions made by each successive layer. As has been shown to be effective in image classification, the common link between each of these methods is a CNN. Each researcher used either a hybrid or modified CNN to classify the images. This is important to note since when fusing ADS-B data with aircraft images, one of the aforementioned algorithms will likely be needed in conjunction with the methods that have already been proven to effectively classify ADS-B data.

2.3.4 Fuzzy Logic

Fuzzy logic is a type of mathematical logic that allows for multiple or subjective truth values. It allows for solving problems that do not have a precise solution. Fuzzy logic tries to provide the best answer given all possible input, but does not guarantee an exact answer. Instead of a typical "True" or "False" for each truth value, fuzzy logic truth values are assigned a degree of membership to each category. For example, when discussing someone's height with categories of "short," "average," and "tall," someone who is 4'2" would probably have values of "1, 0, 0" for each truth value, but a person who was 5'5" might have a truth value of "0.7, 0.3, 0".

In the case of ADS-B classification research, fuzzy values become important when trying to understand an aircraft's current phase of flight. In 2017, Junzi Sun, et al. came up with a method of predicting flight phase using fuzzy logic [18]. After separating observations from each aircraft, the flight phases can be determined from either a complete or partial flight path using Gaussian ($\mathcal{G}$), Z-shaped ($\mathcal{Z}$) or S-shaped ($\mathcal{S}$) membership functions. The membership function definitions are shown in Figure 17 and the relationships shown in Figure 18. In these figures, H, V, RoC and P represent altitude, velocity, rate-of-climb, and flight phase respectively.

$$
\begin{aligned}
H_{gnd}(\eta) &= \mathcal{Z}(\eta, 0, 200) \\
H_{lo}(\eta) &= \mathcal{G}(\eta, 10000, 10000) \\
H_{hi}(\eta) &= \mathcal{G}(\eta, 35000, 20000) \\
RoC_0(\tau) &= \mathcal{G}(\tau, 0, 100) \\
RoC_+(\tau) &= \mathcal{S}(\tau, 10, 1000) \\
RoC_-(\tau) &= \mathcal{Z}(\tau, -1000, -10) \\
V_{lo}(v) &= \mathcal{G}(v, 0, 50) \\
V_{mid}(v) &= \mathcal{G}(v, 300, 100) \\
V_{hi}(v) &= \mathcal{G}(v, 600, 100) \\
P_{gnd}(p) &= \mathcal{G}(p, 1, 0.2) \\
P_{clb}(p) &= \mathcal{G}(p, 2, 0.2) \\
P_{cru}(p) &= \mathcal{G}(p, 3, 0.2) \\
P_{des}(p) &= \mathcal{G}(p, 4, 0.2) \\
P_{lvl}(p) &= \mathcal{G}(p, 5, 0.2)
\end{aligned}
$$

Figure 17: Fuzzy Membership Functions for Flight Phase [18]

$$
\begin{array}{lcl}
if & H_{gnd} \wedge V_{lo} \wedge RoC_0 & then\ \ Ground \\
if & H_{lo} \wedge V_{mid} \wedge RoC_+ & then\ \ Climb \\
if & H_{hi} \wedge V_{hi} \wedge RoC_0 & then\ \ Cruise \\
if & H_{lo} \wedge V_{mid} \wedge RoC_- & then\ \ Descent \\
if & H_{lo} \wedge V_{mid} \wedge RoC_0 & then\ \ Level\ flight
\end{array}
$$

Figure 18: Fuzzy Relationships for Flight Phase [18]

2.4 Sensor Fusion

Information fusion is described as the blending of data acquired from multiple sources. Sensor fusion is a subset of information fusion and is the merging of data derived from only sensory sources [73]. In general, sensor/information fusion provides better reliability, improved coverage, increased confidence, reduced ambiguity, better resolution and is more robust [73].

Sensor fusion is a vast area of study with no common model or architecture. However, one way to split sensor fusion methods is by the two predominant areas when it can occur, pre-classification or post-classification [74]. Like the name implies, pre-classification occurs before the model makes a classification prediction on the model and post-classification occurs afterwards. Some research refers to pre-classification as low or intermediate-level fusion and post-classification as high or decision-level fusion [74, 73]. The most straightforward method is a pre-classification method called feature level fusion. In feature level fusion, the features of the data sets to be combined are merged prior to training the classifier. Post-classification can be broken into two overarching approaches. The first approach is generally used when the number of classes are large and mainly involves feature reduction schemes [74]. The second approach verifies the accuracy of the data being fused. In the latter approach, multiple classifiers are created with data fused via different methods or with different data. The classifiers are either selected based on their accuracy or consolidated to provide an improved decision [74]. In the case of ADS-B, this type of fusion could be invaluable since ADS-B suffers from potential data spoofing. Table 4 outlines a few of the common techniques.

Approach	Method	Description	Papers
Pre-classification	Data Level Fusion	Fuses raw related data	[74, 73]
	Feature Level Fusion	Extracts and combines features of related data using a data association technique such as clustering or grouping	[74, 73]
Post-classification	Abstract Level Fusion	Simplest type of fusion that uses majority voting to select classes	[74, 75]
	Rank Level Fusion	Assigns ranks to classes instead of selecting unique class choices. Common methods are the Borda count, the intersection, the highest rank and the union method	[74, 76]
	Measure Level Fusion	Sub-methods of this technique are classification and combination approaches. Each uses multiple classification methods to either score or combine the results produced by the classifier	[74, 77]
	Dynamic Classifier Selection	Another multi-classifier technique that uses the results of the classifier most likely to provide an accurate result. Essentially, a winner-takes-all or associative switch approach.	[74, 78, 79]

Table 4: Sensor Fusion Techniques

2.5 Summary

Since ADS-B data and its sister data type, AIS, are easy to obtain and in common use, many researchers have used them to gain insights into aviation, aircraft, ship or naval practices. To understand previous research efforts, this chapter presented a review of the relevant literature to ADS-B and AIS classification. The chapter reviewed current research with ADS-B data, pattern-of-life modeling, deep learning with time series data, fuzzy logic, data set size guidelines and data engineering techniques. The chapter also looked at methods of fusing ADS-B sensor data with other sources to improve the classifiers produced by the algorithm.

In many ways, the previous approaches used to classify ADS-B data have attempted to do so without a clear understanding of how the information within the data is related. Essentially, a trial-and-error approach was used. This research intends to take a step back to better understand the ADS-B data set by employing various data engineering techniques and attempting to predict more specific information about the aircraft in the data set. In chapter III, this research will look at methods to optimize features and data set size while predicting engine type. Chapter IV will attempt to predict more specific features of an aircraft such as the ICAO designator, WTC and description while also optimizing the input features and using data from the entire flight path. In chapter V, this research will also fuse other data sources such as images and weather with the ADS-B data set in order to build a model with enhanced predictive power. Exploring these areas will give researchers and analysts tools to use on other data sets when using ADS-B data as a surrogate and will allow the use of more than just the takeoff phase of flight. A summary of the papers that are pertinent to this research can be seen in Table 5.

Table 5: Comparative Study Table

Paper	**Aircraft & Maritime**	**Prediction & Classification**	**Improve Aircraft Operations**	**Transmissions Security**	**Data Set Size**	**Collection & Processing**	**Sensor Fusion**
Ginoulhac *et al* (2019) [27]	✓	✓					
Qian *et al* (2019) [28]	✓	✓					
Kumar *et al* (2021) [31]	✓	✓					
Basrawi (2022) [20]	✓	✓					
Sun *et al* (2016) [43]	✓	✓					
Sun *et al* (2017) [18]	✓	✓					
Zhan *et al* (2021) [80]	✓	✓					
Bækkegaard *et al* (2018) [67]	✓	✓					
Jochumsen *et al* (2016) [68]	✓	✓					
Kraus *et al* (2018) [29]	✓	✓					
Zhang *et al* (2021) [44]	✓	✓					
Zhao *et al* (2018) [45]	✓	✓					
Dästner *et al* (2018) [46]	✓	✓					
Dästner *et al* (2020) [81]	✓	✓					
Kanneganti *et al* (2018) [82]	✓	✓					
Maji *et al* (2013) [17]	✓	✓					
Wang *et al* (2022) [70]	✓	✓					
Wang *et al* (2022) [71]	✓	✓					
Rong *et al* (2021) [69]	✓	✓					
Liu *et al* (2022) [72]	✓	✓					
Ruseno *et al* (2022) [49]	✓		✓				
Filippone *et al* (2021) [50]	✓		✓				
McCallie (2011) [83]	✓		✓	✓			
Hasin *et al* (2021) [51]	✓			✓			
Pearce *et al* (2021) [52]	✓			✓			
Manesh *et al* (2019) [84]	✓			✓			
Sun (2021) [54]	✓					✓	
Karim *et al* (2017) [61]		✓					
Karim *et al* (2019) [15]		✓					
Goodfellow *et al* (2016) [10]		✓					
Hu *et al* (2018) [14]		✓					
Alwosheel *et al* (2018) [62]					✓		
Cho *et al* (2015) [63]					✓		
Jain *et al* (1982) [64]					✓		
Baum *et al* (1988) [65]					✓		
Haykin (2009) [66]					✓		
Raz *et al* (2021) [41]	✓		✓				✓
Elmenreich (2002) [73]							✓
Vakil *et al* (2021) [74]							✓
Przybyła-Kasperek *et al* (2017) [75]							✓
Marasco *et al* (2015) [76]							✓
Habtemariam *et al* (2015) [77]							✓
Xu *et al* (1991) [78]							✓
Gennarelli *et al* (2014) [79]							✓

Table 6: ADS-B Papers

Chapter	**ADS-B, Machine Learning and Data Engineering Contribution**
III	ADS-B Classification Using Multivariate Long Short-Term Memory Fully Convolutional Networks and Data Reduction Techniques
IV	Aircraft Classification Using Flight Phase Identification
V	Multi-Sensor Aircraft Classification

III. Study 1 - ADS-B Classification Using Multivariate Long Short-Term Memory Fully Convolutional Networks and Data Reduction Techniques

The following paper, "ADS-B Classification Using Multivariate Long Short-Term Memory Fully Convolutional Networks and Data Reduction Techniques," was published in Journal of Supercomputing on 12 August 2022. This paper is in Springer journal format. It can be found at the following Uniform Resource Locator (URL): https://link.springer.com/content/pdf/10.1007/s11227-022-04737-4.pdf

The paper discusses two experiments that were performed to determine the effects of reducing the data set size and the optimum feature combination for producing a deep learning model that predicts aircraft engine type during the first 10 minutes of flight. The experiments were performed on the same data set repository that was used in research by Basrawi et al. The repository contained Automatic Dependent Surveillance-Broadcast (ADS-B) training data from 1-8 December 2020 (350 GB) and ADS-B evaluation data from 16 November 2020 [20]. This research achieved slightly higher accuracy with fewer features and an eighth of the data. It was determined that the data set could be further reduced to a quarter of its original size of 44 GB before its accuracy was reduced to an unreasonable level. For the purposes of the experiment, "unreasonable" was defined as a reduction of no more than 10% accuracy. For the optimizing features experiment, it was determined that engine type could most accurately be predicted using airspeed, pressure, and vertical airspeed. It was also noted that the data set was not able to use the "time" feature in its native format of Unix or epoch time. The presence of the "time" feature created noise and, in most cases, severely reduced the accuracy of each model.

The following paper is broken up into five sections. It introduces the topic of ADS-B classification and pattern-of-life (POL) modeling, discusses the implications of limited computing resources on defense operations, and provides a solution through the optimization of ADS-B data using a deep learning model. A literature review is provided to provide relevant background for the two optimization experiments. Afterwards, the methods used to run the experiments are explained. The paper ends with a discussion of the experimental results and future work. The research questions associated with this paper can be seen in Table 7.

Table 7: Summary of Research Questions for Paper 1

ID	**Details**
RQ-1	What is the minimum amount of data needed to effectively classify engine type within 10% accuracy of previous efforts?
RQ-2	What is the smallest feature subset that predicts an aircraft's engine type with an accuracy within 5% accuracy of previous efforts?

The Journal of Supercomputing
https://doi.org/10.1007/s11227-022-04737-4

ADS-B classification using multivariate long short-term memory–fully convolutional networks and data reduction techniques

Sarah Bolton[1] · **Richard Dill**[1] · **Michael R. Grimaila**[2] · **Douglas Hodson**[1]

Accepted: 16 July 2022

Abstract
Researchers typically increase training data to improve neural net predictive capabilities, but this method is infeasible when data or compute resources are limited. This paper extends previous research that used long short-term memory–fully convolutional networks to identify aircraft engine types from publicly available automatic dependent surveillance-broadcast (ADS-B) data. This research designs two experiments that vary the amount of training data samples and input features to determine the impact on the predictive power of the ADS-B classification model. The first experiment varies the number of training data observations from a limited feature set and results in 83.9% accuracy (within 10% of previous efforts with only 25% of the data). The findings show that feature selection and data quality lead to higher classification accuracy than data quantity. The second experiment accepted all ADS-B feature combinations and determined that airspeed, barometric pressure, and vertical speed had the most impact on aircraft engine type prediction.

Keywords Multivariate long short-term memory–fully convolutional network · Automatic dependent surveillance-broadcast · Publicly available information · Open-source data · Classification · Machine learning

1 Introduction

Over the last three decades, storage on the internet increased by over 40,000% from 15.8 exabytes in 1993 to 6.8 zettabytes in 2020 [1]. While it is difficult to determine the exact number, as of February 2022, the size of the internet is

Richard Dill, Michael R. Grimaila and Douglas Hodson these authors have contributed equally to this work.

✉ Sarah Bolton
sarah.bolton@afit.edu

Extended author information available on the last page of the article

Published online: 12 August 2022

estimated to be about 21 zettabytes and is doubling every two years [2]. If we assume the average personal computer (PC) has a hard drive of one terabyte, 21 zettabytes is equivalent to 21 billion PCs, essentially three PCs for every person in the world. While a lot of this data is personal data, a large portion of it is considered publicly available information (PAI) and can be utilized by any internet user or organization.

This increase in available data has resulted in the study of identifying trends (i.e., data analytics), becoming more and more prevalent in multiple facets of society to include commerce and government. Researchers and major corporations have considered multiple ways to best utilize this massive resource, aptly referred to as 'Big Data.' Some areas that have shown promise include Internet of Things (IoT) analysis [3–5], traffic modeling [6], flight and maritime movement [7–11], image recognition [12], search engines [12] and natural language processing [12].

The increased focus on PAI and data analytics is recognized by military defense strategists who are responsible for making sound defense decisions. By incorporating PAI with the plethora of sensor data at their disposal, such as data from intelligence, surveillance, and reconnaissance platforms, it is possible to improve the predictive power of those resources. The need for data analytics is apparent in the United States Air Force and Space Force where multi-domain operations are integral to their defense strategies. In fact, the FY22 Posture Statement calls out Command and Control's need for the translation and sharing of data to provide 'real-time dissemination of actionable information' to provide 'joint warfighting across all domains at a pace faster than our competitors' [13]. Without recent advances in technology, artificial intelligence, and machine learning, this goal would be virtually impossible. Fortunately, new techniques can be used to filter the noise in big data much faster than human speed to quickly make inferences that are important to military decision makers.

To aid military leaders with analyzing the immense data at their disposal, we seek to improve military operations by providing enhanced capabilities for a major user of big data: intelligence analysts. One focus area important to intelligence analysts is pattern-of-life (POL) modeling. Some researchers seek to improve POL modeling via machine learning [14–18]. Recent research interests suggest analyzing ground-based and onboard aircraft sensors with deep learning to predict aircraft characteristics.

One stream of research for POL modeling is focused on exploiting automatic dependent surveillance-broadcast (ADS-B) data to make predictions about aircraft [6, 8, 11, 19]. Aircraft within certain airspace are required to broadcast ADS-B Out via an onboard transponder. The benefit of using ADS-B data for classification problems is that it is publicly available and aircraft flying in the USA and Europe are required to broadcast it in most classes of airspace [20, 21]. ADS-B data is collected from various sites worldwide where hobbyists and researchers maintain a receiver to collect it. ADS-B collectors submit their data to centralized repositories, such as the ADS-B Exchange [22], that aggregate the data for public use. In these repositories, both statistical and kinematic information about the broadcasting aircraft is available.

1.1 Problem description/objective

Pattern-of-life (POL) modeling is a research area with many techniques and best practices [14–18]. Military and defense personnel have an interest in POL modeling that includes more than modeling human day-to-day activities. For example, unattributed data from aircraft sensors, such as those collected from Air Traffic Control's (ATC) primary radar, can allow inferences to be made about the transmitting aircraft with some analysis. ATC's primary radar collects kinematic information such as location and airspeed, but is unable to obtain an aircraft's identification without the aircraft providing it via its transponder. With this basic kinematic aircraft data, models can predict information such as aircraft model or engine type without it being directly stated in the original dataset. The benefit of ADS-B data is that these features are present in the dataset and can be used as truth data to build models for datasets that do not have the truth data.

Since this type of processing can be resource intensive, it can be difficult or, in some cases, impossible to train a deep learning model when dealing with limited computing resources. The amount of computing resources required to train a model is heavily influenced by the size of the training data. For this reason, it is important to understand how to best utilize the available resources by minimizing the data used to train the model. There are two ways to minimize the data: limit the number of features or reduce the number of training samples. In this research, using aircraft kinematic data, we examine the impact of varying these factors when predicting engine type. Since reducing the training data will inevitably reduce the accuracy of the resulting model, for the purpose of this paper, we define an acceptable model as one that predicts within 10% of the previous baseline research results of 89.2% accuracy [23]. Therefore, models that can achieve at least a 79.2% accuracy will be considered 'acceptable.'

1.2 Research contribution

The research contribution of this paper can be summed up in the following points:

- Since there is no definitive guideline for minimum dataset size for deep learning classification problems, this research aims to determine a baseline for aircraft prediction models.
- This paper determines the baseline features to identify an aircraft with kinematic data: Speed, barometric pressure, and vertical speed
- This paper analyzes and reiterates the importance of selecting appropriate features. The 'noise' feature within this dataset severely limited the classification power of the network.

1.3 Organization

The rest of this paper is organized as follows: An exhaustive literature review and background information on ADS-B is provided in section two. In the third section, the methodology and process used to develop and evaluate each model is discussed. The fourth section presents the results. The conclusion is provided in the fifth section.

2 Background and literature review

This section describes automatic dependent surveillance-broadcast (ADS-B) data; previous attempts to classify engine type from it; relevant deep learning techniques; and the recommendations for dataset size when building a neural network model. Table 1 outlines the papers discussed in this section.

2.1 Automatic dependent surveillance-broadcast (ADS-B)

Automatic dependent surveillance-broadcast (ADS-B) is an aircraft sensor used throughout many regions of the world. Several researchers have utilized ADS-B data to identify flight patterns [8, 11, 24, 25], improve aircraft operations [26, 27], increase ADS-B transmission security [28, 29] and classify targets [7]. The transmissions are openly broadcast via an ADS-B Out transponder at 1090 MHz in many countries worldwide and at both 1090 MHz and 978 MHz within the USA. ADS-B data is received and collected via various methods to include other aircraft, ground stations, and satellites. Figure 1 depicts an ADS-B communications network.

Ground stations consist of sites where antennas collect ADS-B transmissions from passing aircraft. The ground stations are typically managed by commercial or government organizations, but hobbyists also collect the broadcasts. In the case of hobbyists, the aircraft transmissions are collected on a nearby device and transferred to one or multiple organizations that host the data for public use (e.g., the ADS-B Exchange [22]).

The ADS-B Exchange and similar services save the incoming data to their servers in two- to five-second intervals as a JavaScript Object Notation (JSON) file. Per the USA's DO-260B standard and Europe's ED-102A standard, each broadcast contains up to 70 features about the transmitting aircraft including the International Civil Aviation Organization (ICAO) identifier, altitude, airspeed, vertical speed, directional heading, position time, latitude, longitude, and ground status [22, 30].

Table 2 provides details on the countries that have an ADS-B mandate and the date when the mandate began or is projected to start. The chart shows that most countries, including the USA, Australia, the European Union Aviation Safety

Table 1 Comparative study table

Paper	ADS-B	Prediction & classification	Improve aircraft operations	Transmission security	Dataset size	Collection & processing
Ginoulhac et al. (2019) [7]	✓	✓				
Qian et al. (2019) [8]	✓	✓				
Kumar et al. (2021) [11]	✓	✓				
Basrawi (2021) [19]	✓	✓				
Basrawi et al. (2021) [23]	✓	✓				
Sun et al. (2016) [24]	✓	✓				
Sun et al. (2017) [25]	✓	✓				
Ruseno et al. (2022) [26]	✓		✓			
Filippone et al. (2021) [27]	✓		✓			
Hasin et al. (2021) [28]	✓			✓		
Pearce et al. (2021) [29]	✓			✓		
Sun (2021) [30]	✓					✓
Karim et al. (2017) [31]		✓				
Karim et al. (2019) [32]		✓				
Goodfellow et al. (2016) [33]		✓				
Hu et al. (2018) [34]		✓				
Alwosheel et al. (2018) [35]					✓	
Cho et al. (2015) [36]					✓	
Jain et al. (1982) [37]					✓	
Baum et al. (1988) [38]					✓	
Haykin (2009) [39]					✓	

Agency (EASA), and many East Asia and Pacific island countries, began their mandate on or before 2020 which forces aircraft that fly within those regions to install the appropriate ADS-B Out equipment.

Within the USA, the mandate is not required for low flying aircraft and aircraft operating out of rural airports. Table 3 explains the US requirements in more detail. Within the USA, ADS-B Out is required for aircraft flying above 10,000 feet MSL

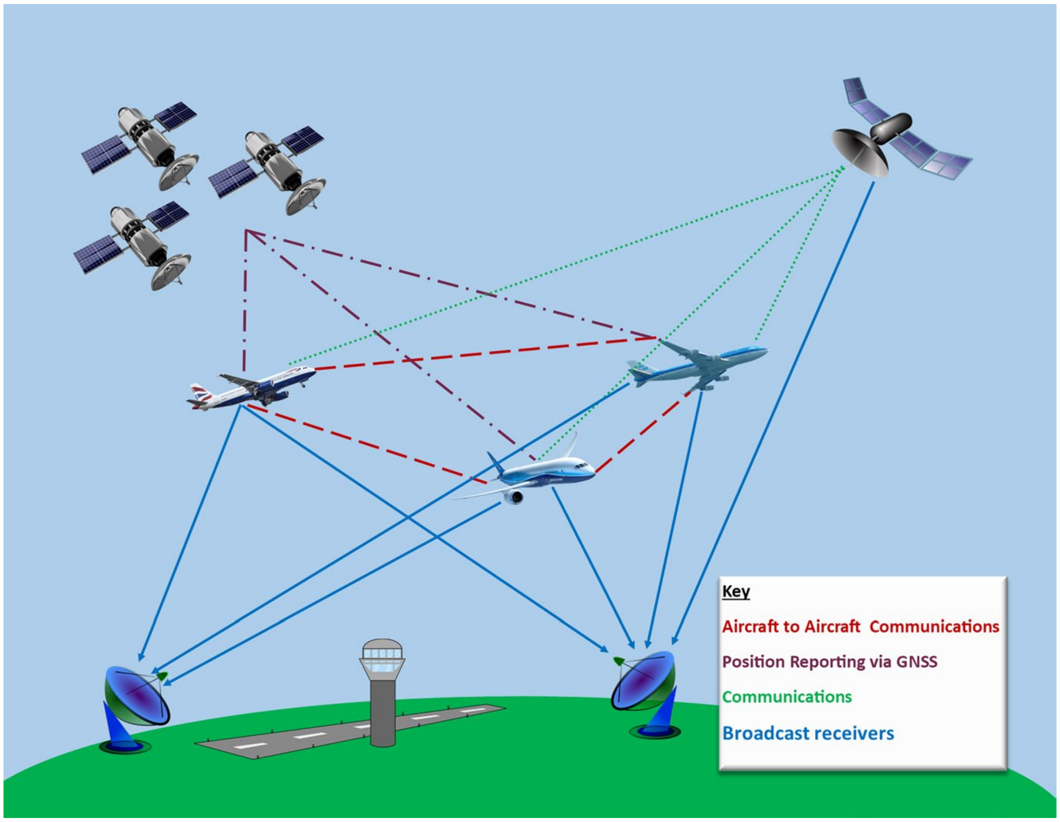


Fig. 1 ADS-B high-level operational concept graphic

and aircraft operating in locations near airports classified by Class B or Class C airspace.

2.2 Multivariate long short-term memory–fully convolutional neural networks

Artificial neural networks (ANNs) have been shown to be successful at classifying a variety of datatypes. Many ANN variations have been developed over the years to classify the infinite number of datatypes researchers encounter. Long short-term memory (LSTM) networks were developed to classify time series data and convolutional neural networks were developed to classify images. Combining these two networks, the multivariate long short-term memory–fully convolutional network (MLSTM–FCN) is shown to improve upon previous methods to classify time series data [31]. The algorithm, developed by Karim et al., combined fully convolutional networks (FCN), LSTM networks, and squeeze-and-excite blocks. FCNs were utilized to allow for the CNN benefit of class action maps without the requirement of extensive hyperparameter preprocessing that is normally required by a CNN. LSTMs were selected to detect the importance of sequences of observations in time series data. Squeeze-and-excite blocks were added to the algorithm to improve the classification power of multivariate datasets by ensuring feature maps have a similar impact to subsequent layers. In their study, Karim et al. tested four variations of the algorithm against 35 different datasets to include voice, human signal monitoring,

Table 2 ADS-B mandate timelines [40]

8. 1 Global ADS-B Out mandate summary

Mode S Transponders with Extended Squitter capability in accordance with RTCA DO-260 or DO-260B:	
Australia mandate for retrofit aircraft	12-Dec-2013
Singapore mandate for all aircraft	12-Dec-2013
Indonesia mandate for all aircraft	01-Jan-2018
Hong Kong mandate for all HKG-registered aircraft - Flying PBN routes L642 or M771 between FL290 and FL410 - Flying within HKG FIR between FL290 and FL410	 31-Dec-2013 31-Dec-2014
Australia mandate for SA Aware GNSS (new a/c)	08-Dec-2016
Mexico - SBAS GPS required as per FAA guidelines	January 1, 2020 (Pending)
Colombia - SA-ON acceptable	January 1, 2020 (Pending)
U.S. FAA mandate for 100% equipage - Class A, Class B and Class C airspace - 1090 ES (Extended Squitter) for FL180 and above - 1090 ES or UAT (Universal Access Transceiver) below FL180 - Must be compliant with TSO C166b (Mode S) or C154a (UAT)	01-Jan-2020
EASA mandate for *forward-fit and retrofit* aircraft	07-June-2020
China mandate for all aircraft	31-Dec-2022

and other time series data. Their algorithms outperformed the current state-of-the-art algorithm in 27 out of the 35 datasets [31]. Additionally, a variation of the algorithm, the multivariate attention long short-term memory–fully convolutional network (MALSTM–FCN), predicted aircraft engine type using ADS-B data [19]. This research will be further discussed in the next section.

While LSTMs and FCNs are well established, the use of squeeze-and-excite blocks is a much newer technique. To understand how they improve the MLSTM–FCN, it's important to understand how they work and what they provide. Squeeze-and-excite blocks were introduced by Hu et al. in 2018 to improve the representational capacity of a neural network [34]. They are used with convolutional networks to help model the dependencies between the channels. As the name implies, it consists of a squeeze mechanism followed by an excitation mechanism. In the squeeze step, the classifier uses global average pooling to aggregate the spatial information of each channel. Using the example of an image, where an image has dimensions of $H \times W \times C$ for height, width, and (normally 3) channels, the squeeze step would pass the image through a global average pooling operator where it would become a shape of $(1 \times 1 \times C)$. Then, the excite step uses a gating mechanism to capture channel-wise dependencies [32]. The excitation step uses a multi-layer neural network with one hidden layer. The input and output layer are the same shape $(1 \times 1 \times C)$, but the hidden layer reduces the space by a reduction factor, r, making the number of neurons in the hidden layer C/r. Karim et al. and Hu et al. use a reduction factor of 16 [32, 34]. This $(1 \times 1 \times C)$ value is multiplied element-wise with the original $(H \times W \times C)$ input. The graphical representation of the squeeze-and-excite block developed by Hu et al. can be seen in Fig. 2.

Using the squeeze-and-excite block in conjunction with an LSTM and FCN, the MLSTM–FCN was developed to create a model for time series classification. Figure 3 shows the architecture Karim et al. designed for their research. This algorithm is used to develop all of the models in this research.

2.3 Dual-stage deep engine classifier

Multiple researchers have focused on using the ADS-B dataset to predict engine type with ADS-B kinematic data [8, 19]. One of the limitations with analysis on this dataset is that while the models have very few problems determining if the aircraft is a jet, they tend to have trouble predicting the difference between turboprop and piston engines. The reason for this is twofold. First, the dataset is heavily imbalanced toward jet engines which causes the data to have fewer samples of turboprop and piston engines from which to learn. Second, the kinematic differences between piston and turboprop engines are minimal.

As a method to remedy this problem, Basrawi et al. developed a Dual-Stage Deep Engine Classifier (DSDEC) [19]. Using the MALSTM–FCN algorithm as a basis for the model, the DSDEC algorithm employs a unique 2-stage approach. When using the model to make predictions, the first stage predicts if the aircraft is a jet or not a jet. Then, the 'not jet' predictions are fed to the second stage. The second stage predicts if the aircraft has a piston or turboprop engine. The results from both the first and second stage are combined to provide an engine prediction for each observation. Figure 4 portrays the architecture that is used.

Basrawi et al. are the first to classify aircraft engine type using ADS-B data with a deep learning model. Using the DSDEC method, researchers were able to identify jets with 98.4% accuracy, turboprop aircraft with 79.2% accuracy, and piston engine aircraft with 89.9% accuracy. However, the DSDEC method would still confuse turboprop aircraft as piston engine aircraft 17.9% of the time [19]. Table 4 shows the results achieved by Basrawi et al. compared to a support vector machine (SVM) and random forest (RF) as baseline from previous research[8].

During the experiments, static time steps were used and each time step was separated by two seconds. 300 time steps would be equivalent to 600 seconds or 10 minutes of flight time. Similarly, 100 time steps would be 200 seconds or 3 minutes and 20 seconds of flight time. While Basrawi et al. indicated that more research would be needed to determine the influence of the time step size, their results point to the possibility that longer flight observations improve the model's predictive power. In fact, while the DSDEC algorithm was used against 21 different models to determine which hyperparameters led to the highest accuracy, none of the 100 time step models performed as well as the 300 time step examples.

2.4 Size of dataset

Determining the exact dataset size needed to build an artificial neural network (ANN) classifier is an open research problem that may never have a complete solution due to the infinite number of ANN combinations. However, experts suggest several guidelines to ensure enough data is available to train a model with sufficient accuracy value [35–39]. Guidelines can be broken into three important characteristics of the dataset: the number of prediction classes, the number of features, and the number of weights in the network. The following guidelines sum up the results from past research endeavors:

1. **Prediction Classes**: A sample size should have 50-1000 times the number of observations as prediction classes [36]. For this paper, there are three prediction classes: Jet, turboprop, and piston.
2. **Features**: There should be 10-100 times the number observations as features [37].
3. **Network Weights**: The sample size should be equivalent to 10 times the number of weights in the network [38, 39].
    (a) Another paper decided on a stricter limit stating that there should be 50 times the number of observations as network weights [35].

2.5 Summary

The information provided in this section shows that ADS-B sensor data is a useful resource when trying to understand POL modeling with aircraft. Previous research has used this data to predict aircraft characteristics [7, 8, 19]. When predicting engine type, researchers found that the MLSTM–FCN was able to train a model that achieved an overall accuracy of 89.2% [19]. This paper uses the information gleaned

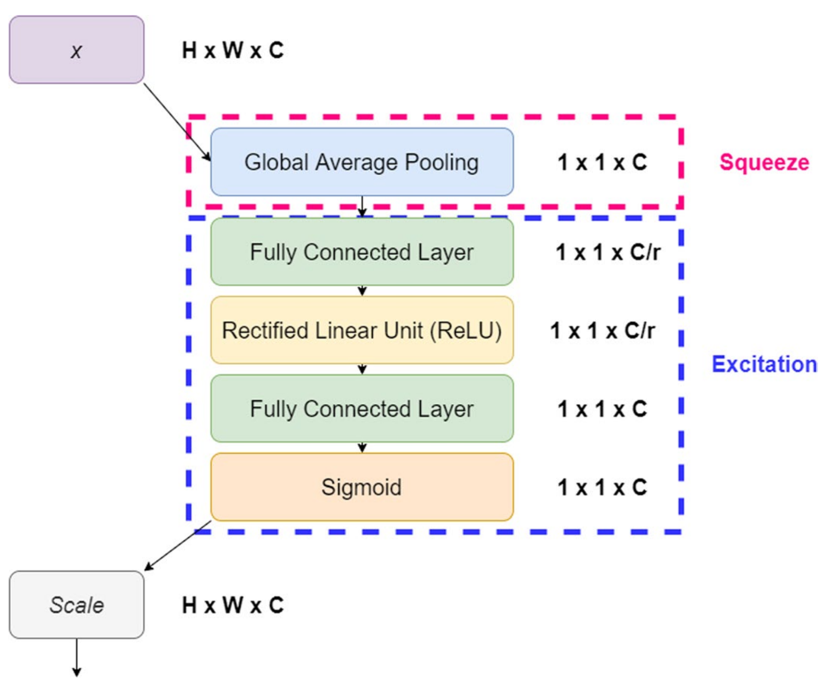


Fig. 2 The schema of the squeeze-and-excite model

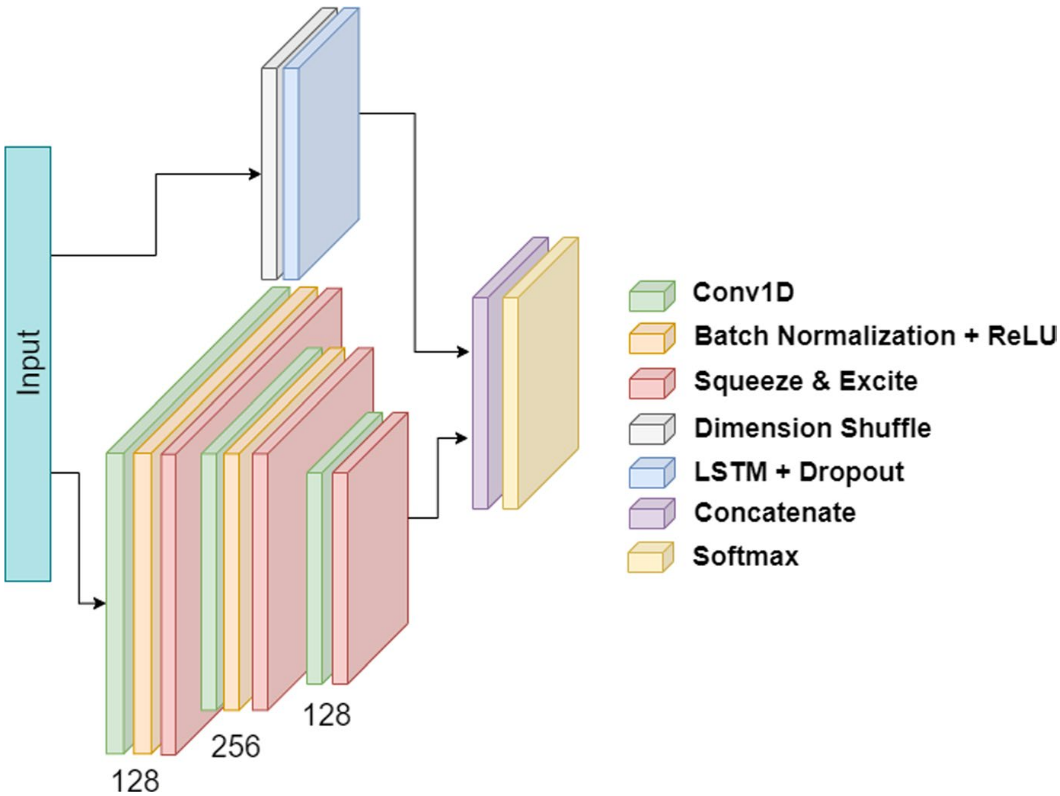


Fig. 3 MLSTM–FCN architecture [32]

Table 3 US ADS-B Airspace Requirements [20]

Airspace	Altitude
Class A	All
Class B	Generally, from surface to 10,000 feet mean sea level (MSL) including the airspace from portions of Class Bravo that extend beyond the Mode C Veil up to 10,000 feet MSL (e.g. LAX, LAS, PHX)
Class C	Generally, from surface up to 4,000 feet MSL including the airspace above the horizontal boundary up to 10,000 feet MSL
Class E	At and above 10,000 feet MSL over the 48 states and DC, excluding airspace at and below 2,500 feet AGL
	Over the Gulf of Mexico at and above 3,000 feet MSL within 12 nautical miles of the coastline of the United States
Mode C Veil	Airspace within a 30 NM radius of any airport listed in Appendix D, Section 1 of Part 91 (e.g. SEA, CLE, PHX) from the surface up to 10,000 feet MSL

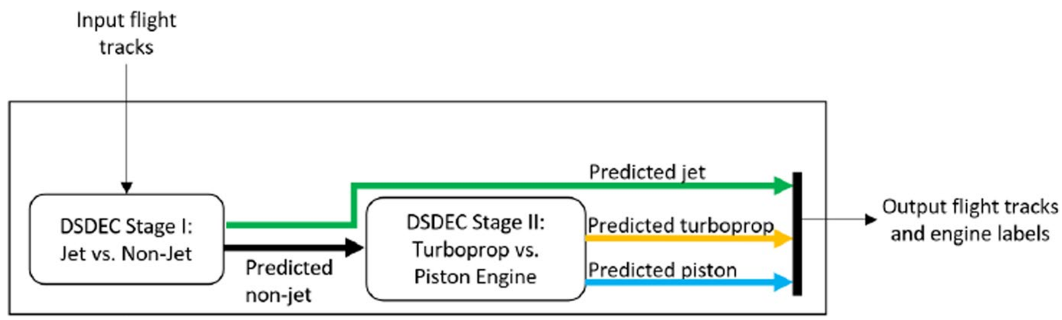


Fig. 4 DSDEC architecture [23]

Table 4 Results from Basrawi et al. [19]

Classifier	Overall accuracy	Jet Accuracy	Turboprop accuracy	Piston engine accuracy
DSDEC (300 time steps)	***89.2%***	***98.4%***	***79.2%***	***89.9%***
SVM (300 time steps)	77.4%	85.5%	72.2%	74.6%
SVM (100 time steps)	68.6%	81.1%	56.6%	68.1%
RF (300 time steps)	83.4%	94.1%	70.5%	85.8%
RF (100 time steps)	76.7%	90.0%	61.5%	78.6%

Bold italic values indicate the best performing examples

from research done with the MLSTM–FCN [32] and DSDEC [19] to learn how to minimize the input data size while maintaining a similar accuracy and loss.

3 Methodology

This section outlines a method to efficiently classify engine type from ADS-B data. It improves on Basrawi et al [19] by reducing the model complexity and decreasing the requisite ADS-B dataset size to classify engine. Basrawi et al. used two feature sets, a limited and a full feature set, which had 6 and 9 features, respectively. Additionally, they used 9 days worth of ADS-B data to train a dual-stage classification algorithm. This method reduces the model complexity to a single stage and only requires a 24-hour sample size and three ADS-B features to achieve similar classification accuracy results. It uses data from 1 December 2020 to train the model and 16 November 2020 to test the model.

The rest of this section describes two experiments:

- **Experiment 1:** The first experiment creates 24 models with two different learning rates, three feature sets, and four data amounts. The models train over a period of 200 epochs and are evaluated on the testing data. The goal of the first experiment is to determine the effect of varying the number of training data observations. Effects to accuracy, loss, precision, and recall are recorded.

- **Experiment 2:** In the second experiment we develop models using all possible feature combinations (255). The goal of the second experiment is to determine which subset of features has the highest overall accuracy when predicting engine type. The models train for 50 epochs, which is where the first experiment shows training and validation data accuracy diverge.

The details of each experiment can be seen in Table 5.

3.1 Assumptions and limitations

The dataset used in this research is smaller than the one used by Basrawi et al. [19]. In this research, the training data from 1 December 2020 has 4,110 tracks/tensors, and the evaluation data from 16 November 2020 has 2,487 tracks/tensors. In the experiment completed by Basrawi et al., the training data ranged from 1 to 8 December 2020 consisting of 7,749 tracks/tensors. The evaluation data was also from 16 November 2020, but consisted of 4,158 tracks. This amounts to approximately half of the tracks for training and evaluation in this experiment. It is assumed that the effects of reducing the data size would be proportional regardless if one week or one day was used for training.

Although researchers have shown that the integrity of ADS-B sensor data is vulnerable to a variety of cyber attacks [41, 42], the dataset is assumed accurate for the purposes of this paper. While vulnerable, cyber attacks against ADS-B are not a common occurrence. Air traffic control-related attacks occur only a few times each year [43]. We consider this to be an acceptable assumption since the number of occurrences of message injects and other ADS-B attacks is low in comparison to the millions of flights that occur each year.

3.2 Process

The process for converting raw ADS-B data into a model that can predict engine type can be broken down into five steps that will be explained in the subsequent paragraphs Fig. 5.

3.2.1 Data preparation

The ADS-B data acquired by Basrawi et al. [19] contains aircraft observations from thousands of locations each day from November to December 2020. During the preparation phase, the data is converted from JavaScript Object Notation (JSON) files to Python Data Analysis Library (Pandas) data frames, sorted by unique International Civil Aviation Organization (ICAO) and time, and then saved as Comma-Separated Values (CSV) files. Then, the CSV files are reloaded as a Pandas data frames with invalid and irrelevant data (e.g., helicopter and glider data points)

Table 5 Experiment Details

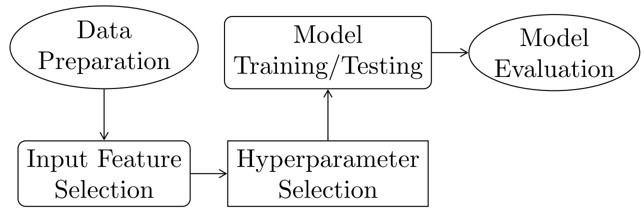
Exp #	Goals	# of Models	Parameters
1	Determine the effect of varying the amount of training data samples	24	Learning rate: 0.01 & 0.001 Drop out: 0.5 Time steps: 300 Feature set: limited, medium & full Optimizer: Adam Data size: Full, 1/2, 1/4, & 1/8 Epochs: 200
2	Determine the effect of each subset of features	255	Learning rate: 0.001 Drop out: 0.5 Time steps: 300 Feature set: all possible subsets Optimizer: Adam Data size: Full Epochs: 50

removed. Aircraft without at least 300 transmissions (10 minutes) are removed and any transmissions after 300 time steps are discarded to keep the data as similar as possible. For all aircraft, only the first 10 minutes of the takeoff segment of flight are used. The remaining data is balanced by engine type using an undersampling technique [44, 45] to reduce the impact of the heavily imbalanced dataset. The undersampling technique is selected over an oversampling or a hybrid method since the dataset is already very large and diverse. Thousands of 'Jet' observations remain in the dataset even with undersampling. Since they are randomly removed, overfitting is avoided. Then, the engine type feature is converted from a number to a one hot encoding of Boolean values with 'jet,' 'turboprop' and 'piston' as possible options. The data preparation step reduces 44 GB of raw data to 270 MB of processed data.

The processed data is formed into (n,t,v) tensors [46], where n is the number of aircraft tracks, t is the number of time steps for each track, and v is the number of features describing each track. In the previous version of this experiment t and v were varied across multiple trials to determine the best combination. It was determined that reducing the t variable lower than 300, lowers the accuracy. Therefore, $t = 300$ is maintained across all iterations of this experiment. n and v are altered by changing the amount of training data and changing the number of features, respectively.

Experts agree on best practices to follow when training classification models. The first recommendation is that the training dataset sample size be at least 50-1000 times the number of prediction classes [36]. Since there are only three prediction classes and over a million observations, this goal is met. The second suggestion is that the sample size is 10-100 times larger the number of features [37]. The feature set with the most number of features is 12. Since $12 \times 100 = 1200$ and there are 1,233,000 samples, this suggestion is also met. The final guideline states that the number of observations should be 10 times the number of weights in the model.

Fig. 5 Data Processing Steps



Since the model has 280,000-288,000 weights (depending on the number of features for that experiment), the number of samples needed is near 3 million [38, 39]. With less than half the requisite samples, this guideline is not met. However, the study that this experiment was built on only used 2,324,700 samples and was able to achieve a nearly 90% accuracy [19]; it is assumed that not meeting the 10x weight requirement, but meeting all other guidelines, is sufficient for this experiment.

In the final part of data preparation, we create four training datasets of varying sizes. The first training dataset includes observations from the entire 24-hour period on 1 December 2020. This dataset is the largest dataset and is referred to as the full dataset throughout this paper. In the creation of the next three datasets, we randomly remove tensors from the full dataset to create a new dataset that is half, a quarter, and an eighth the size of the full dataset. Since it is done randomly, the datasets are not equal to 12, 6, and 3 continuous hours of observations. Instead, the observations are 300 time step tensors taken at various points during the 24-hour time period. Proportions between engine types are maintained per the undersampling technique. The full dataset results in 4,110 tensors (1,233,000 samples), the half dataset has 2,055 tensors (616,500 samples), the quarter dataset has 1,026 tensors (307,800 samples), and the eighth dataset has 513 tensors (153,900 samples). Table 6.

3.2.2 Input feature selection

The raw ADS-B data collected during November and December 2020 contained a total of 57 features. Most of them were either not consistently transmitted by all aircraft or contained irrelevant or redundant information. There were also a few features that contained identifying information like ICAO, aircraft model, ID number, country of origin, inbound/outbound location, or other non-kinetic information that is not intended to be used for this research. This reduced the usable kinetic input features from 57 to 9. We also develop 3 other input features from these inputs to improve the location data. The definition of each feature is listed in Table 7.

The following features are selected for inclusion into the dataset due to their potential for predicting engine type.

1. Altitude and Ground Altitude—Since jet, turboprop, and piston engine aircraft tend to fly best at different altitudes, these features are important in distinguishing between them. [48]
2. Airspeed—Jets fly faster than piston or turboprop engines. Turboprop engines can reach greater speeds easier at higher altitudes than piston engines. [48]

3. Barometric Pressure—This feature is another way to measuring altitude. For every thousand feet of elevation, the pressure drops by 1 inHg. [49]
4. Vertical Speed—The importance of this feature is similar in nature to air/ground speed.
5. Time—This feature would allow the comparison of different chronological points of the flight
6. Track—With location and speed, track can be used to learn specific aircraft patterns.
7. Lat/Long—Aircraft may exhibit different behaviors depending on the geography. For example during the first ten minutes, an aircraft would takeoff differently from a mountainous region than an open field.
8. Location (X,Y,Z)—Lat/Long are normalized to better represent a 3-dimensional space. This feature was not in the raw data, but was instead generated to help better represent the data. The formulas used to normalize the lat/long are as follows:

    (a) $x = \cos(\text{lat}) * \cos(\text{lon})$
    (b) $y = \cos(\text{lat}) * \sin(\text{lon})$
    (c) $z = \sin(\text{lat})$

Feature combinations for the first experiment are tested based on their size and their possible effects on classification. The feature combinations experiment one uses are outlined in Table 8.

For the second experiment, all possible combinations of the full feature set are used to determine the best subset of features. However, latitude and longitude are not used for this experiment. They are replaced with the normalized location coordinates.

3.2.3 Hyperparameter selection

Since hyperparameters were already tested by Basrawi et al. to determine the best combination, this study uses those hyperparameters for the model creation with the exception of learning rate. Basrawi et al. found that the Adam optimizer and learning rates of 0.01 and 0.001 performed the best, while dropout rates were mostly inconsequential. Based on their findings, we train the model with a dropout rate of 0.5 and learning rates of 0.01 and 0.001 with the Adam optimizer.

Table 6 Dataset details from 1 december 2020

Size	Samples	Tensors
Full	1,233,00	4,110
Half	616,500	2,055
Quarter	307,800	1,026
Eighth	153,900	513

Table 7 ADS-B Data Features/Fields [22, 47]

Feature	Field	Data type	Description
Altitude	Alt	integer	The altitude in feet at standard pressure
Ground Altitude	Galt	integer	The altitude adjusted for local air pressure
Airspeed	Spd	knots (float)	The ground speed in knots
Barometric Pressure	InHg	float	The air pressure in inches of mercury that was used to calculate the AMSL altitude from the standard pressure altitude
Vertical Speed	Vsi	integer	Vertical speed in feet per minute
Time	PosTime	epoch (ms)	The time that the position was last reported by the aircraft.
Track	Trak	degrees (float)	Aircraft track angle across the ground clockwise from 0 north.
Latitude	Lat	float	The aircraft's latitude over the ground
Longitude	Long	float	The aircraft's longitude over the ground
Location	X,Y,Z	float	Cartesian coordinates of the aircraft. This data feature is derived from the lat/long coordinates and it is not originally part of the ADS-B broadcast

3.2.4 LSTM training, testing, and evaluation

Similar to Basrawi et al. [19], this study uses the algorithm developed by Karim et al. [31]. The two major differences are that this research omits the attention mechanism and does not use the two-phase approach suggested by Basrawi et al. [19]. In the first experiment, 24 separate models are created out of the data from 1 December 2020 to encompass the variations in learning rate (0.01 and 0.001), the feature set (limited, medium, and full), and the data amount used (24 hours, half of the data, a quarter of the data and an eighth of the data). Each model is trained for 200 epochs which in most cases is more than sufficient.

In the second experiment, 255 models are created using a 0.001 learning rate and trained on the full size dataset. The features vary between each model to evaluate all possible combinations of features. Each model is trained for 50 epochs which is the point when training and validation data accuracy deviate.

Accuracy, precision, loss, and recall are saved during the creation of the models for both experiments to show the models' histories during each epoch. Those metrics are also saved for the final models' evaluation. A k-fold method, where $k = 10$, tests each model with the data from 16 November 2020. Each model is compared to show how the change in both the input feature size and the amount of data points affected the aforementioned metrics.

4 Results and discussion

4.1 Experiment 1–24 models: data size comparison

Table 9 outlines the performance across all phases. Figures 6 and 7 represent the data from Table 9. The line colors represent the input feature size and the color's shade represents the learning rate. The 0.001 learning rate is darker shade and the 0.01 learning rate is the lighter shade. Blue represents the smallest input feature size, green the medium input feature size, and orange the full input feature set. As can be seen, the limited feature set has the best accuracy and loss with the medium and full feature set having very irregular performance. While not shown, recall and precision follow similar trends.

Model 1 performs the best. It has an 89.4% overall accuracy which is on par with the 89.2% accuracy achieved by Basrawi et al. [23], but with half the sample size. The confusion matrix for model 1 can be seen in Figure 8. The results for individual engine type are also similar to those collected by Basrawi et al. For comparison, they found that jet engines were predicted correctly 98.4% of the time, turboprop 79.2%, and piston 89.9% [23]. In this study, jets are accurately predicted 97.2% of the time, turboprops are 79.1% and pistons are 92.0%.

The results show that the limited feature dataset outperforms the medium and full feature sets. The larger feature sets never converge. This phenomenon is due to noise and overfitting to the training data. Experiment two is able to further analyze the lack of convergence. However, using the three feature sets created for this experiment, it is shown that altitude, airspeed, vertical speed, and location provide sufficient information about aircraft to differentiate engine type in most cases. The additional features do not provide any information that help the model learn more about aircraft engines. Instead, it makes the data more convoluted. This can be shown with the training history as seen in Figures 9, 10, and 11. These figures represent the training accuracy after each epoch where the feature set is varied between the figures, but learning rate and dataset size remain constant. The three images are from the models that had a 0.001 learning rate and used the full dataset (i.e., models 1, 3, and 5). While the training accuracy continues to increase, the validation accuracy does not. This indicates that the model is no longer improving and is instead overfitting to the training data. While only a subset of the models are shown in this manner, other models follow a similar pattern with the limited feature set performing the best in all cases.

The results from the limited feature set show that more data improves the classification power of the model. However, the reduction in accuracy is

Table 8 Experiment Trial Setup

Limited feature set	Medium feature set	Full feature set
Altitude Airspeed Vertical speed Location (X,Y,Z)	Altitude Airspeed Vertical speed Location (X,Y,Z) Barometric Pressure Time Track	Altitude Airspeed Vertical speed Location (X,Y,Z) Barometric Pressure Time Track Ground Altitude Lat/Long

gradual until the one-eighth size dataset. Based on our previous definition of an ‘acceptable’ model, the 1/4 size dataset meets the minimum accuracy with 83.9%. For this reason, if computational power was limited, someone could drop the number of inputs to 1/4 of the full dataset and still be guaranteed to have a reasonable prediction accuracy rate.

4.2 Experiment 2–255 models: feature set comparison

The goal for experiment two is to find the subset of features that best predict engine type from the full size dataset. All possible feature combinations from 8 available features result in the creation of 255 models. The top 20 results from the 255 models are presented in Table 10. Airspeed, pressure, and vertical airspeed create the best-performing dataset. This result is unexpected since it does not include altitude. Jets fly at a different altitude than piston or turboprop engines. For that reason, it would seem like altitude would be an important feature. A possible reason for this might be that the observations only include the first 10 minutes after takeoff. Jet engines would not have the time to reach cruising altitude until the end of the 10 minute time frame.

Additionally, the results from experiment show that a combination of all of the features except for ‘PosTime’ provide the third best feature indicator for accuracy. However, not only is ‘PosTime’ not included in the 3rd best combination, but it is also interesting to note that all of the worst-performing models contain ‘PosTime.’ Table 11 shows the 10 worst-performing models. It is not until the 43rd worst performer (the model with just the ‘Trak’ feature with a 48.9% accuracy rate) that the ‘PosTime’ feature is not present.

5 Conclusion

The goal of this research is to determine the effect of minimizing the size of the training data that was used to develop a MLSTM–FCN model. We do this with two separate experiments. In experiment one, we vary the number of training samples in the training dataset with three different feature sets. In experiment two, we vary the number of features present in the dataset. Those models are tested to determine how each change affects the accuracy and loss of the model against a separate test set. We deem a model to be ‘acceptable’ if the accuracy was within 10% of results from previous research [19].

There are a few main takeaways from this experiment. The first is that the quality of a dataset is more important than the quantity of data when building a neural network. Having redundant features or features that do not add pertinent information to the model can cause the model to perform poorly. This fact is convenient when it comes to minimizing a dataset, but can be frustrating when trying to determine

Table 9 Performance comparisons

Num	Learn rate	Feature set	Data amt	Mean accuracy	Mean loss
1	0.001	Limited	Full	**0.894 (+/−0.018)**	0.413 (+/−0.069)
2	0.01	Limited	Full	0.890 (+/−0.019)	**0.346 (+/−0.083)**
3	0.001	Medium	Full	0.513 (+/−0.031)	1.321 (+/−0.092)
4	0.01	Medium	Full	0.337 (+/−0.019)	3.604 (+/−0.117)
5	0.001	Full	Full	0.759 (+/−0.038)	0.675 (+/−0.100)
6	0.01	Full	Full	0.356 (+/−0.018)	2.960 (+/−0.210)
7	0.001	Limited	Half	0.848 (+/−0.019)	0.474 (+/−0.072)
8	0.01	Limited	Half	0.865 (+/−0.023)	0.367 (+/−0.051)
9	0.001	Medium	Half	0.492 (+/−0.029)	1.493 (+/−0.102)
10	0.01	Medium	Half	0.490 (+/−0.031)	1.622 (+/−0.102)
11	0.001	Full	Half	0.740 (+/−0.025)	0.595 (+/−0.053)
12	0.01	Full	Half	0.354 (+/−0.023)	3.071 (+/−0.149)
13	0.001	Limited	1/4	0.839 (+/−0.016)	0.473 (+/−0.098)
14	0.01	Limited	1/4	0.839 (+/−0.014)	0.478 (+/−0.054)
15	0.001	Medium	1/4	0.700 (+/−0.026)	0.789 (+/−0.078)
16	0.01	Medium	1/4	0.391 (+/−0.040)	2.211 (+/−0.152)
17	0.001	Full	1/4	0.658 (+/−0.028)	0.891 (+/−0.089)
18	0.01	Full	1/4	0.505 (+/−0.040)	1.645 (+/−0.121)
19	0.001	Limited	1/8	0.758 (+/−0.019)	0.897 (+/−0.124)
20	0.01	Limited	1/8	0.688 (+/−0.037)	1.525 (+/−0.224)
21	0.001	Medium	1/8	0.678 (+/−0.028)	1.008 (+/−0.088)
22	0.01	Medium	1/8	0.607 (+/−0.026)	1.397 (+/−0.175)
23	0.001	Full	1/8	0.595 (+/−0.038)	1.833 (+/−0.179)
24	0.01	Full	1/8	0.484 (+/−0.021)	2.120 (+/−0.116)

Bold value indicate the best performing examples

which features are actually important. Too many irrelevant features will add noise into the model and produce results that are less than ideal. Based on this experiment, speed, pressure and vertical speed are some of the more important features needed to identify an aircraft's engine type.

This dataset is reduced to a quarter of its original size before the model starts exhibiting severe negative effects. During both experiments, the best model achieves an accuracy of 89.4%. Reducing the data by half only reduces the accuracy by 4.6% to 84.8% accuracy which has a rate of change of 9.2% (Δ accuracy/Δ size). Reducing it to a quarter of its original size produces an 83.9% accuracy rate which is a 22.0% rate of change from the full size dataset. However, reducing the dataset to an eighth of its original size produces a 75.8% accuracy rate which is a 108.8% rate of change from the full size dataset. Since the raw data from 1 December 2020 was about 44 GB, this large file size could be reduced to 11 GB using the quarter size dataset and still effectively train a classification model. Most computers, laptops, or small handheld devices could perform data processing and train a model that was only 11 GB. This observation would work well for military operations in remote locations with low computational resources. If data and computation resources were

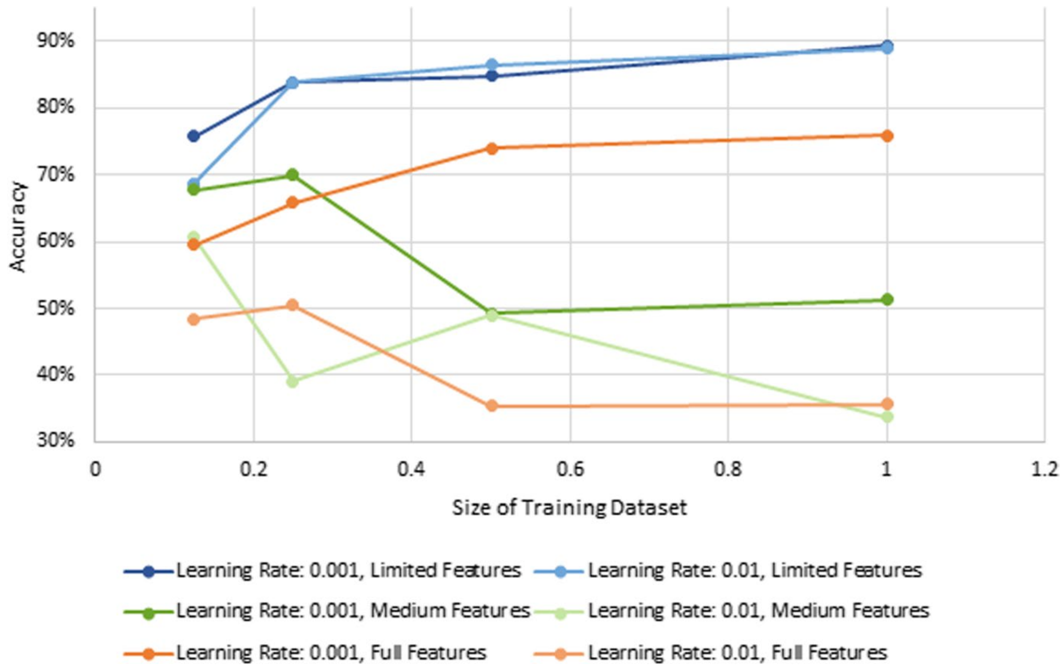


Fig. 6 Change in accuracy by change in sample size

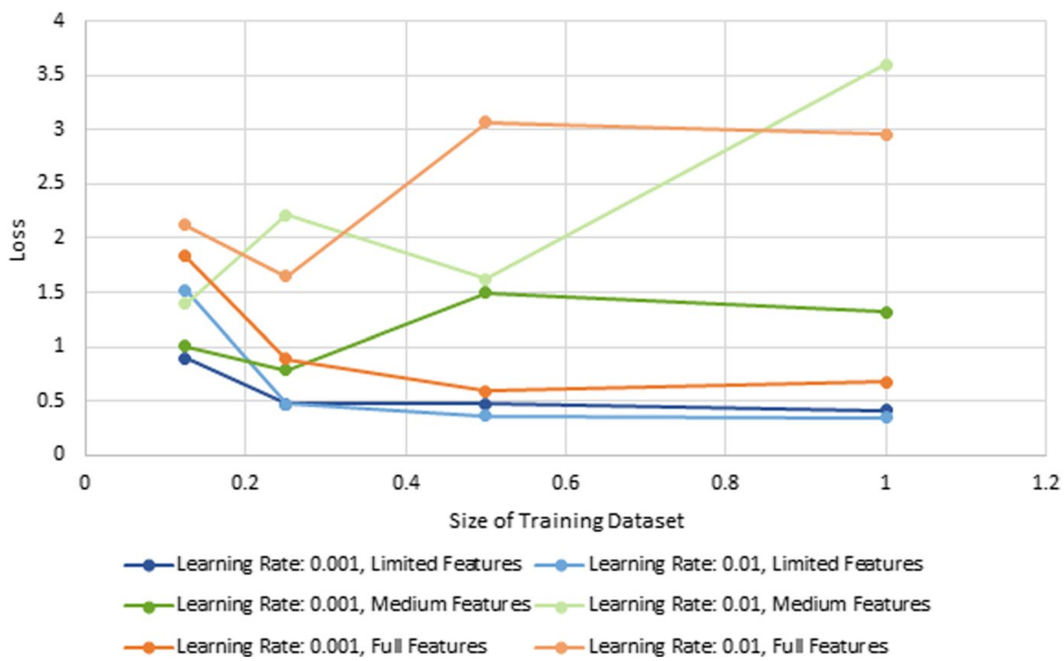


Fig. 7 Change in loss by change in sample size

not as plentiful, reducing the model to approximately 300,000 observations would create a model that could make observations with a small trade-off of less than 10% reduced accuracy.

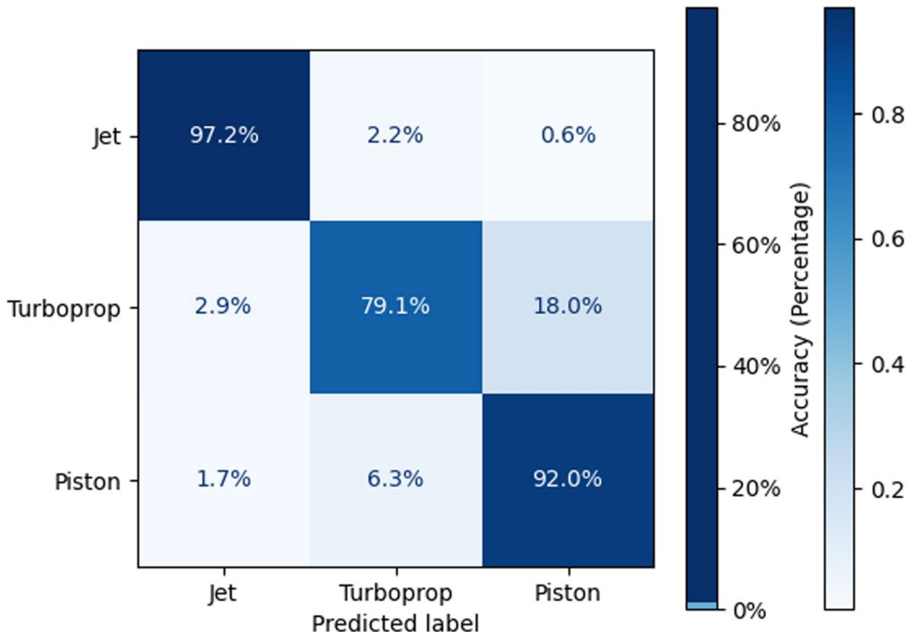


Fig. 8 Confusion Matrix for Model 1

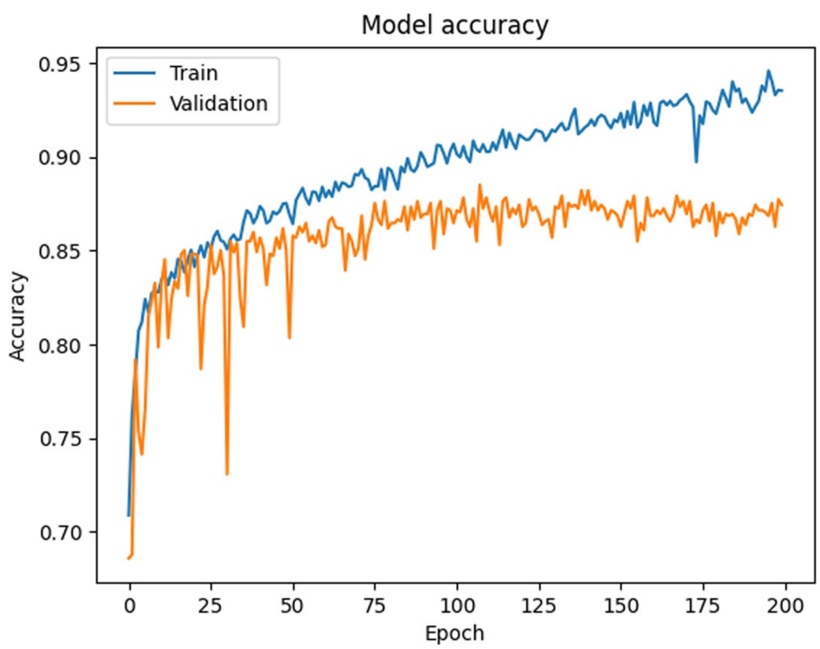


Fig. 9 Accuracy Training History on Limited Feature Set for 0.001 Learning Rate and Full Dataset

5.1 Future work

While the goal of this study is to determine how to minimize the input data to make best use of low computational resources, military operations tend to have access to a vast number of sensors. Combining other related sensors to the ADS-B data could improve results. Other potentially useful sensors include weather, radar, and image

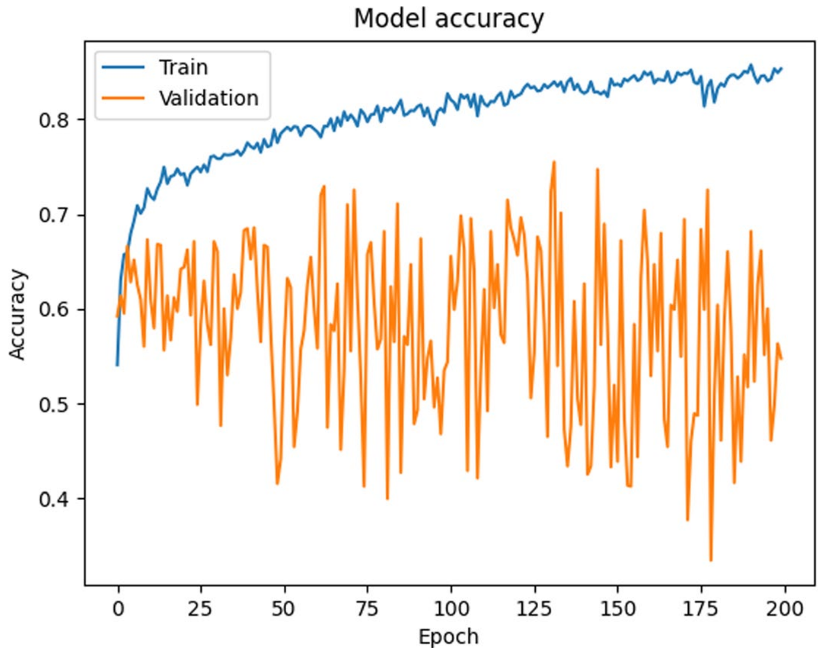


Fig. 10 Accuracy Training History on Medium Feature Set for 0.001 Learning Rate and Full Dataset

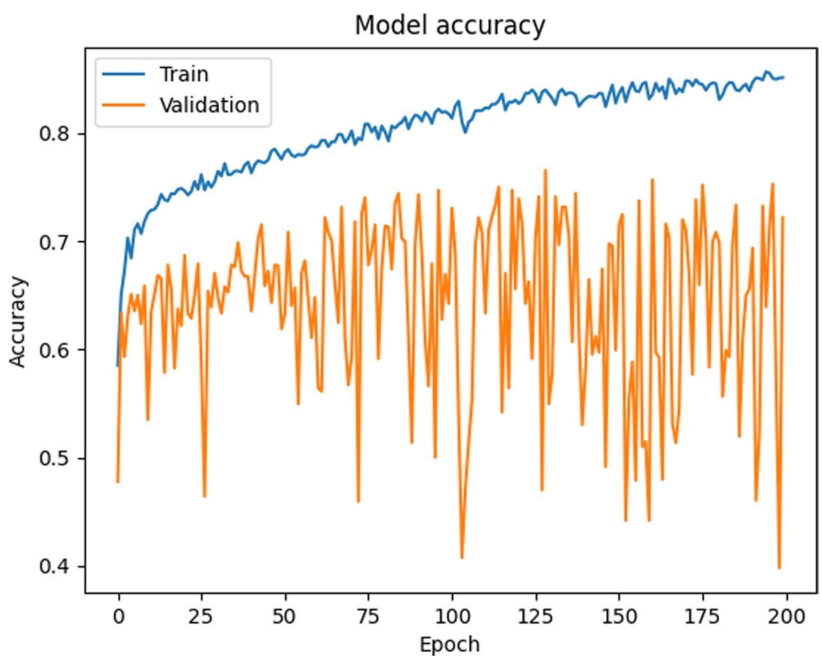


Fig. 11 Accuracy Training History on Full Feature Set for 0.001 Learning Rate and Full Dataset

data. Researchers should look at the effects of combining these sources to determine if they improve classification accuracy.

The raw ADS-B data represents the 'PosTime' feature as the number of milliseconds since Epoch (1970). During the processing of the data, 'PosTime' is modified to more closely compare to previous work with engine type prediction. Instead of milliseconds since Epoch, it represents the amount of milliseconds since takeoff.

Table 10 Top 20 Performance Comparisons

Ex	Feature set	Mean accuracy	Mean loss
56	Spd,InHg,Vsi	0.894 (+/−0.013)	0.304 (+/−0.038)
40	Spd,Vsi	0.887 (+/−0.025)	0.336 (+/−0.088)
251	Alt,Galt,Spd,InHg, Vsi,Trak,location	0.886 (+/−0.023)	0.319 (+/−0.056)
59	Spd,InHg,Vsi,Trak, location	0.886 (+/−0.015)	0.307 (+/−0.060)
186	Alt,Spd,InHg,Vsi,Trak	0.884 (+/−0.017)	0.330 (+/−0.059)
171	Alt,Spd,Vsi,Trak, location	0.884 (+/−0.016)	0.366 (+/−0.132)
43	Spd,Vsi,Trak,location	0.884 (+/−0.015)	0.326 (+/−0.066)
42	Spd,Vsi,Trak	0.877 (+/−0.017)	0.346 (+/−0.067)
185	Alt,Spd,InHg,Vsi, location	0.876 (+/−0.026)	0.318 (+/−0.056)
105	Galt,Spd,Vsi,location	0.876 (+/−0.019)	0.347 (+/−0.052)
105	Galt,Spd,Vsi,location	0.876 (+/−0.019)	0.347 (+/−0.052)
184	Alt,Spd,InHg,Vsi	0.875 (+/−0.020)	0.332 (+/−0.070)
235	Alt,Galt,Spd,Vsi,Trak, location	0.875 (+/−0.018)	0.362 (+/−0.060)
58	Spd,InHg,Vsi,Trak	0.875 (+/−0.015)	0.360 (+/−0.057)
248	Alt,Galt,Spd,InHg,Vsi	0.874 (+/−0.017)	0.342 (+/−0.056)
123	Galt,Spd,InHg,Vsi,Trak, location	0.874 (+/−0.015)	0.370 (+/−0.064)
169	Alt,Spd,Vsi,location	0.872 (+/−0.020)	0.364 (+/−0.077)
41	Spd,Vsi,location	0.871 (+/−0.023)	0.334 (+/−0.037)
104	Galt,Spd,Vsi	0.871 (+/−0.020)	0.369 (+/−0.060)
114	Galt,Spd,InHg,Trak	0.871 (+/−0.017)	0.368 (+/−0.062)
106	Galt,Spd,Vsi,Trak	0.871 (+/−0.013)	0.396 (+/−0.069)

When creating a model with an LSTM, this is not the best use of the time feature. Keeping track of time of day, day of the week, or even just the date would allow the model to better learn aircraft schedules. Future work should look at modifying this feature into something that would improve the classifier.

Table 11 Bottom 10 performance comparisons

Ex	Feature set	Mean accuracy	Mean loss
7	PosTime,Trak,location	0.337 (+/−0.038)	1.622 (+/−0.088)
29	InHg,Vsi,PosTime, location	0.337 (+/−0.037)	4.511 (+/−3.705)
21	InHg,PosTime,location	0.337 (+/−0.028)	38.038(+/−109.699)
22	InHg,PosTime,Trak	0.336 (+/−0.024)	1.615 (+/−0.117)
52	Spd,InHg,PosTime	0.334 (+/−0.024)	2.812 (+/−0.138)
127	Galt,Spd,InHg,Vsi, PosTime,Trak,location	0.334 (+/−0.019)	18.762 (+/−1.317)
37	Spd,PosTime,location	0.332(+/−0.016)	5.308(+/−0.238)
55	Spd,InHg,PosTime,Trak, location	0.114 (+/−0.022)	3.323 (+/−0.116)
38	Spd,PosTime,Trak	0.110 (+/−0.018)	2.742 (+/−0.072)
54	Spd,InHg,PosTime,Trak	0.074 (+/−0.015)	4.664 (+/−0.112)

This research uses only the first 10 minutes of flight to train each model. The first 10 minutes of flight consists of the takeoff phase and sometimes part of the cruising phase. Incorporating other phases, such as cruising and landing, could help to further separate the differences between engine types. The biggest benefit of using other phases is that jets, turboprops, and pistons have different characteristics when they get to the cruising phase. Jets fly at much higher altitudes during the cruising phase than other engine types and turboprop engines can reach greater speeds easier at higher altitudes than piston engines [48].

Funding This study was funded by the Air Force Research Laboratory.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Bartley K (2021) Big data statistics: how much data is there in the world? . https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/
2. Live-counter.com: how big is the internet? https://www.live-counter.com/how-big-is-the-internet/
3. Saleem TJ, Chishti MA (2019) Deep learning for internet of things data analytics. Proc Comput Sci 163:381–390. https://doi.org/10.1016/j.procs.2019.12.120
4. Dholey MK, Sinha D, Mukherjee S, Das AK, Sahana SK (2020) A novel broadcast network design for routing in mobile ad-hoc network. IEEE Access 8:188269–188283
5. Kumar V, Kalam S, Das AK, Sinha D (2021) Attack detection scheme using deep learning approach for iot. Adv Comput Sys Sec 14:17–30
6. Qian L, Gintautas V (2019) Fusing sensor data with publicly available information (PAI) for autonomy applications. In: SPIE Defense + Commercial Sensing, Baltimore, p. 22 . https://doi.org/10.1117/12.2518933
7. Ginoulhac R, Barbaresco F, Schneider JY, Pannier JM, Savary S (2019) Target classification based on kinematic data from AIS/ADS-B, using statistical features extraction and boosting. Proceedings International Radar Symposium 2019-June(681), 1–10 . https://doi.org/10.23919/IRS.2019.8768094
8. Qian L, Woods N, Rahman M (2019) Identifying ADS-B flight patterns. In: Paper presented at the 2019 MSS Joint (BAMS and NSSDF) Conference, San Diego, CA, 21–24 Oct 2019
9. Kraus P, Mohrdieck C, Schwenker F (2018) Ship classification based on trajectory data with machine-learning methods. In: Rohling, P.D.H. (ed.) Proceedings International Radar Symposium, vol. 2018-June, pp. 1–10. IEEE Computer Society, Bonn, Germany . https://doi.org/10.23919/IRS.2018.8448028
10. Andersson M, Johansson R (2010) Multiple sensor fusion for effective abnormal behaviour detection in counter-piracy operations. 2010 International Waterside Security Conference, WSS 2010 (December) . https://doi.org/10.1109/WSSC.2010.5730221
11. Kumar SG, Corrado SJ, Puranik TG, Mavris DN (2021) Classification and analysis of go-arounds in commercial aviation using ads-b data. Aerospace 8(10):291. https://doi.org/10.3390/aerospace8100291
12. Mohammadi M, Al-Fuqaha A, Sorour S, Guizani M (2017) Deep learning for IoT big data and streaming analytics: a survey. arXiv 20(4): 2923–2960.
13. Department of the air force posture statement fiscal year 2022. department of the air force. https://www.armed-services.senate.gov/imo/media/doc/FY22%20DAF%20Posture%20Statement%20-%20Final%20(v23.1)1.pdf
14. Silverman BG, Bharathy G, Weyer N (2019) What is a good pattern of life model? Guidance for simulations. Simulation 95(8):693–706. https://doi.org/10.1177/0037549718795040
15. Mikhailov S, Kashevnik A, Smirnov A (2020) Tourist behaviour analysis based on digital pattern of life. In: 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT), vol. 1, pp. 622–627 . https://doi.org/10.1109/CoDIT49905.2020.9263945
16. Hatcher WG, Yu W (2018) A survey of deep learning: Platforms, applications and emerging research trends. IEEE Access 6:24411–24432. https://doi.org/10.1109/ACCESS.2018.2830661
17. Forti N, Millefiori LM, Braca P (2019) Unsupervised extraction of maritime patterns of life from automatic identification system data. In: OCEANS 2019 - Marseille, pp. 1–5 . https://doi.org/10.1109/OCEANSE.2019.8867429
18. Faraj F (2021) Object detection and pattern of life analysis from remotely piloted aircraft system acquired full motion video. PhD thesis, Queen's University , (Canada)
19. Basrawi K (2021) Aircraft classification from ads-b kinematic data. PhD thesis, Air Force Institute of Technology
20. United States department of transportation: Airspace (2021). https://www.faa.gov/air_traffic/technology/equipadsb/research/airspace/
21. National archives and records administration: electronic code of federal regulations (2021). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-C/section-91.225
22. ADS-B Exchange: home - serving the flight tracking enthusiast - ADS-B exchange (2021). https://www.adsbexchange.com/
23. Basrawi K, Dill R, Peterson G, Borghetti B, Lopez J (2022) Aircraft identification from ADS-B Kinematic Data. J DoD Res Eng 5(2):31–43

24. Sun J, Ellerbroek J, Hoekstra J (2016) Large-scale flight phase identification from ads-b data using machine learning methods. In: 7th International Conference on Research in Air Transportation, pp. 1–8
25. Sun J, Ellerbroek J, Hoekstra J (2017) Flight extraction and phase identification for large automatic dependent surveillance-broadcast datasets. J Aerospace Inform Sys 14(10):566–572
26. Ruseno N, Lin C-Y, Chang S-C (2022) Uas traffic management communications: the legacy of ads-b, new establishment of remote id, or leverage of ads-b-like systems? Drones 6(3):57. https://doi.org/10.3390/drones6030057
27. Filippone A, Parkes B, Bojdo N, Kelly T (2021) Prediction of aircraft engine emissions using ads-b flight data. Aeronautical J 125(1288):988–1012. https://doi.org/10.1017/aer.2021.2
28. Hasin F, Munia TH, Zumu NN, Taher KA (2021) Ads-b based air traffic management system using ethereum blockchain technology. In: 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), pp. 346–350 . https://doi.org/10.1109/ICICT4SD50815.2021.9396828
29. Pearce N, Duncan KJ, Jonas B (2021) Signal discrimination and exploitation of ads-b transmission. In: SoutheastCon 2021, pp. 1–4 . https://doi.org/10.1109/SoutheastCon45413.2021.9401909
30. Sun J (2021) The 1090 megahertz riddle: a guide to decoding mode s and ads-b signals. TU Delft OPEN Publishing
31. Karim F, Majumdar S, Darabi H, Chen S (2018) LSTM fully convolutional networks for time series classification. IEEE Access 6:1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939
32. Karim F, Majumdar S, Darabi H, Harford S (2019) Multivariate lstm-fcns for time series classification. Neural Net 116:237–245. https://doi.org/10.1016/j.neunet.2019.04.014
33. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, Cambridge
34. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
35. Alwosheel A, van Cranenburgh S, Chorus CG (2018) Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. J Choice Model 28:167–182. https://doi.org/10.1016/j.jocm.2018.07.002
36. Cho J, Lee K, Shin E, Choy G, Do S (2015) How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv preprint arXiv:1511.06348
37. Jain AK, Chandrasekaran B (1982) 39 dimensionality and sample size considerations in pattern recognition practice. In: Elsevier (ed.) Classification pattern recognition and reduction of dimensionality. handbook of statistics, vol. 2, pp. 835–855. Elsevier B.V., North Holland, Amsterdam . https://doi.org/10.1016/S0169-7161(82)02042-2
38. Baum EB, Haussler D (1989) What size net gives valid generalization? Neural Comput 1:151–160
39. Haykin SS (2009) Neural networks and learning machines. Prentice Hall, New York
40. Honeywell aerospace: automatic dependent surveillance-broadcast will transform air traffic control. technical report, honeywell international Inc, Phoenix, Arizona (2018). https://aerospace.honeywell.com/content/dam/aerobt/en/documents/landing-pages/white-papers/DO-260-mandates-white-paper.pdf
41. Kamaruzzaman MRB, Fall D, Hossain MD, Taenaka Y, Kadobayashi Y (2022) Modulated synchronous taxiing: Mitigating uncertainties amid ads-b spoofing. In: 2022 Integrated Communication, Navigation and Surveillance Conference (ICNS), pp. 1–13 . https://doi.org/10.1109/ICNS54818.2022.9771477
42. Riahi Manesh M, Kaabouch N (2017) Analysis of vulnerabilities, attacks, countermeasures and overall risk of the automatic dependent surveillance-broadcast (ads-b) system. Int J Crit Infrastruct Prot 19:16–31. https://doi.org/10.1016/j.ijcip.2017.10.002
43. Wu Z, Shang T, Guo A (2020) Security issues in automatic dependent surveillance - broadcast (ads-b): a survey. IEEE Access 8:122147–122167. https://doi.org/10.1109/ACCESS.2020.3007182
44. Kaur H, Pannu HS, Malhi AK (2019) A systematic review on imbalanced data challenges in machine learning: applications and solutions. ACM Comput Surv (CSUR) 52(4):1–36
45. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G (2017) Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl 73:220–239
46. Weinman JJ, Lidaka A, Aggarwal S (2011) TensorFlow: Large-scale machine learning. GPU Computing Gems Emerald Edition (November), 277–291 https://arxiv.org/abs/1603.04467v21603.04467v2
47. Brooks TN (2018) Using autoencoders to learn interesting features for detecting surveillance aircraft . https://doi.org/10.48550/ARXIV.1809.10333

48. Haygood J (2021) Turboprop Vs piston airplanes . https://www.skytough.com/post/turboprop-vs-piston-airplanes
49. Service NW (2022) Pressure definitions . https://www.weather.gov/bou/pressure_definitions/

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sarah Bolton[1] · **Richard Dill**[1] · **Michael R. Grimaila**[2] · **Douglas Hodson**[1]

Richard Dill
richard.dill@afit.edu

Michael R. Grimaila
michael.grimaila@afit.edu

Douglas Hodson
douglas.hodson@afit.edu

[1] Department of Engineering and Computer Science, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB 45433, OH, USA

[2] Department of Systems Engineering and Management, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB 45433, OH, USA

IV. Study 2 - Aircraft Prediction

The following paper, "Aircraft Classification Using Flight Phase Identification," was accepted by the 2023 National Aerospace and Electronics Conference (NAECON). This paper is in IEEE format.

Unlike previous research which focused on the first 10 minutes of flight, this paper presents a method of aircraft classification using all flight phases with Automatic Dependent Surveillance-Broadcast (ADS-B) kinematic data. First, the research utilizes a previously proven fuzzy logic method of flight phase prediction to incorporate the flight phase into the ADS-B observations. Then, the study uses an Multivariate Long Short-Term Memory - Fully Convolutional Network (MLSTM-FCN) to predict an aircraft's Wake Turbulence Category (WTC), description and designator, as described in the ICAO Document 8643 [48]. To analyze the model, the research develops three experiments. The first experiment analyzes the value of including flight phase by building two models for each of the class types (WTC, description and designator); one with flight phase and one without flight phase. The second experiment focuses on the class imbalance within each of the class types and adds weights to the classes based on their presence within the data. The last experiment completes a flight phase analysis to determine which flight phase provides the most accurate predictions. It was shown that, WTC depended highly on the addition of flight phase and that the climb and descent phases are the most influential phases for accurately predicting aircraft characteristics.

The following paper is broken up into five sections. First, it introduces ADS-B and its vulnerabilities. Then, it provides a background on previous research in ADS-B classification and flight phase prediction. Next, the paper presents the experimental setup to include data preparation steps and associated algorithms. The paper ends with the results and conclusion. The research questions associated with this paper can be seen in Table 8.

Table 8: Summary of Research Questions for Paper 2

ID	Details
RQ-3	How does the identification of flight phase assist in aircraft prediction of description, designator and WTC?
RQ-4	Which phases work best to predict aircraft by their wake turbulence category (WTC), aircraft description and type designator (as listed in ICAO Doc 8643)?

Aircraft Classification Using Flight Phase Identification

Sarah Bolton
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
sarah.bolton@afit.edu

Richard Dill
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
richard.dill@afit.edu

Michael R. Grimaila
Department of Systems Engineering and Management
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
michael.grimaila@afit.edu

Douglas D. Hodson
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
douglas.hodson@afit.edu

***Abstract*—Flight data collected from an Automatic Dependent Surveillance–Broadcast (ADS-B) can be used to make inferences about aircraft characteristics. This work uses all flight phases to identify an aircraft's Wake Turbulence Category (WTC), aircraft description and type designator as described in the International Civil Aviation Organization (ICAO) Document 8643.**

***Index Terms*—Multivariate Long Short-Term Memory – Fully Convolutional Network, Automatic Dependent Surveillance-Broadcast, open-source data, classification, machine learning**

I. Introduction

Automatic Dependent Surveillance–Broadcast (ADS-B) provides aircraft tracking and identification data that can be received by satellites and ground stations. It is open source and can be captured by any receiver within line-of-sight of an aircraft that is transmitting ADS-B data. While it is not used globally, much of North America, Europe, Asia and all of Australia have some form of ADS-B usage mandate. In January 2020, the United States mandated the use of ADS-B within controlled airspace with Federal Regulation 14 CFR 91.225 and 14 CFR 91.227.

Since it is widely available and setting up an ADS-B receiver is relatively simple, thousands of receiver owners share their data to create a full air picture in any place that ADS-B is used. Two common repositories of aircraft data are the ADS-B Exchange and FlightAware [1], [2]. Each of these sites accept data from the ADS-B receiver owners, compile the data, and provide the compiled data to other feeders or to paying customers. Per the ADS-B Exchange, as of March 2023, there are over 9,500 active feeds that track over 15,000 aircraft [1]. ADS-B has become a powerful resource for aircraft researchers and operators.

One important feature of ADS-B is the transmission of identifying information within the data. One of the more useful pieces of information is the aircraft's International Civil Aviation Organization (ICAO) 24-bit aircraft address or Mode S hex code. The ICAO 24-bit identifier is like a phone number or Internet Protocol (IP) address for an aircraft. As the name implies, it is a 24-bit address that is textually represented by 6 hexadecimal digits. ICAO Annex 10, Volume 3 outlines the regulations regarding Mode S code allocation [3]. Having an aircraft's Mode S code allows the lookup and tracking of the specific aircraft with that address. This provides the ability to determine aircraft characteristics such as designator, model, manufacturer, weight and any other performance parameters that have been previously assigned to that hex code.

Two downsides to ADS-B are that the broadcast is vulnerable to cyber attacks and the broadcast is not passive. It is not passive since the aircraft has to turn on its ADS-B Out device for the broadcast to occur. If an aircraft chooses not to turn on their ADS-B, no identifying information about the aircraft is transmitted and it must be tracked by a more passive form of aircraft tracking such as radar. When ADS-B is broadcast, invalid data can be easily transmitted either intentionally (i.e. cyber attacks) or unintentionally (i.e. typos or miscalibrated sensors). While passive aircraft tracking does not require input from the pilot and can be done without the pilot's knowledge, passive tracking can not provide the specific details that are found in an ADS-B transmission like the ICAO hex code. Since ADS-B provides truth data, usage of ADS-B opens an area of research where the kinematic data found in ADS-B can be used as a surrogate for more passive forms of aircraft tracking. Aircraft classification models can be built to be used by passive forms of aircraft tracking, which can then validate the authenticity of ADS-B.

This paper presents methods to use the kinematic data found in ADS-B broadcasts to predict aircraft characteristics using deep learning. It is structured as follows: Section II discusses previous work with ADS-B classification. Section III outlines the process and discusses the models that were used to convert raw ADS-B files into predictive models. Section IV reviews

Funding provided by the Air Force Research Laboratory.

the findings from 3 different experiments that were created to determine the effectiveness of the process described in this paper. Finally, section V concludes the paper.

II. Background & Previous Work

ADS-B has been used in research for a variety of purposes. Notably, ADS-B data has been used to predict flight patterns [4]–[8], improve ADS-B transmission security [9]–[11], enhance aircraft operations [12]–[15] and to classify aircraft characteristics [16]–[19]. Due to the complex nature of the data found in ADS-B broadcasts, previous research in aircraft classification has fallen victim to two challenges. First, ADS-B data is extremely imbalanced towards jet aircraft. In the research completed in [16], Ginoulhac et al. created 6 categories of aircraft based on the type of aircraft (airplane/helicopter), wake turbulence category (WTC) and engine type. They found that 94% of their dataset contained medium WTC airplanes with jet engines. They were able to achieve a mean precision, recall and f1-score of 87%, but did not account for the imbalanced nature of the data. While their research was a great start in the field of aircraft classification, not accounting for such a severe imbalance made the results less reliable. While their accuracy for the dominant class was 98%, the average accuracy for the other 5 classes was 56% (ranged from 86% to 37%).

The second challenge in ADS-B classification research is the extreme differences in aircraft operations during each flight phase. For example, an aircraft will fly at a different velocity, altitude and rate of climb during the take-off phase than it will during its cruise phase. This can complicate classification model development if researchers do not account for these differences. To circumvent this problem, previous research has attempted to classify aircraft characteristics based on just the take-off phase and has omitted data from other phases [17]–[19]. However, it is important to note that if an aircraft needed to be identified without access to its ADS-B data (i.e. the aircraft cannot or will not transmit it), the take-off and ground phases would be the least useful phases to classify since the aircraft could easily been seen and identified by the Air Traffic Control (ATC) tower without the need for ADS-B. For that reason, it is important to include all flight phases in research.

To include all flight phases, knowing which phase an aircraft is in during a given time period is important to annotate within the data. Various researchers have effectively developed methods to predict flight phase using fuzzy logic [5], [7], [20]. The input values commonly used for the fuzzy sets are *altitude, velocity* and *vertical rate/rate-of-climb*. The outputs are the flight phases. Research completed by [7] categorized the flight phases into *ground/taxi, climb, cruise, descent, level flight* and *unknown*. The phases were defined by the rules outlined in table I. Research completed by [20] used similar flight phases, but removed *level flight* (i.e. leveling out temporarily during climb or descent) and broke *descent* into two parts. This paper will use the design created by [7] in table I.

Accounting for the imbalanced data with an undersampling technique but using only the take-off phase, researchers were able to effectively predict aircraft engine type using a Multivariate Long Short-term Memory–Fully Convolutional Network (MLSTM-FCN) [17]–[19]. This type of neural network was developed in 2019 to classify time series data [21]. The MLSTM-FCN implements a *squeeze-and-excitation* block to improve accuracy beyond the capability of just the LSTM and FCN. Work from [17] used a dual stage deep engine classifier (DSDEC) where a binary classification model was created that first predicted *jet* or *not jet*. If the observation was determined to be *not jet*, it was fed into another model that was built to predict the difference between *turboprop* and *piston* engines. The model achieved an 89.2% accuracy which surpassed previous statistical machine learning (SML) classification techniques [4], [17]. Research completed by [19] focused on reducing the input data and the feature set. The feature set that provided the highest accuracy was *speed*, *barometric pressure* and *vertical speed*. Using half the number of tensors (4,110 compared to 7,749), a third of the features and a multi-class classifier instead of the DSDEC technique, [19] obtained similar results to [17] with an accuracy of 89.4%.

TABLE I
Flight Phase Fuzzy Logic Rules [7]

Flight Phase	***Altitude***	***Velocity***	***Vertical Speed***
Ground	Ground	Low	Zero
Climb	Low	Medium	Positive
Cruise	High	High	Zero
Descent	Low	Medium	Negative
Level Flight	Low	Medium	Zero
Unknown[a]	—-	—-	—-

[a] *Unknown* is defined by any aircraft states that are not defined by the rows above it

III. Methodology

The goal behind this research is to predict an aircraft's WTC, description, and type designator as listed in the ICAO doc 8643 using a deep learning model during all flight phases. Table II defines each of these aircraft characteristics. This paper focuses on the use of the MLSTM-FCN for classification due to its performance. During the course of the research, other deep learning architectures, such as a few variations of a bidirectional LSTM, Temporal Convolutional Network (TCN), and a binary classifier modeled after the DSDEC method [17] were attempted, but if after 15 epochs they did not show improvement over the MLSTM-FCN, training was discontinued. Ultimately, this paper solely discusses the use and training of an aircraft classifier with the MLSTM-FCN architecture as shown in figure 1.

A. Data Preparation

Provided by the ADS-B exchange [1], the data in this research spans the months of October and November 2021. October 2021 is used as the training set and November 1-2, 2021 is used for the test set. Raw data files are in a JSON format with 5 seconds between each observation. First, these files are converted to compressed CSV files for usage in our

TABLE II
ICAO DOC 8643 [22], [23]

Name	***Definition***	***Options***
Wake Turbulance Category (WTC)	Maximum certified take-off mass	ICAO 8643 defines WTC by heavy, medium or light. For the purposes of this research, medium is separated into two categories (small and large). • Light (< 15500 lbs) • Small (15500 to 75000 lbs) • Large (75000 to 300000 lbs) • Heavy (> 300000 lbs) • High performance (> 5g acceleration and 400 kts)
Description	3-characters used to describe aircraft type, number of engines and engine type	First character (aircraft type): • L - Landplane • S - Seaplane • A - Amphibian • H - Helicopter • G - Gyrocopter • T - Tilt-wing Aircraft Second character (# of engines): 1, 2, 3, 4, 6, 8 Third character (engine type): • P - Piston • T - Turboprop/turboshaft • J - Jet • E - Electric Example: L2T - a landplane with two turboprop engines
Type Designator	Up to a 4-character code that identifies an aircraft's model	There are 653 different designators in our dataset. The five most common ones are B738, A320, A321, B737 and A319.

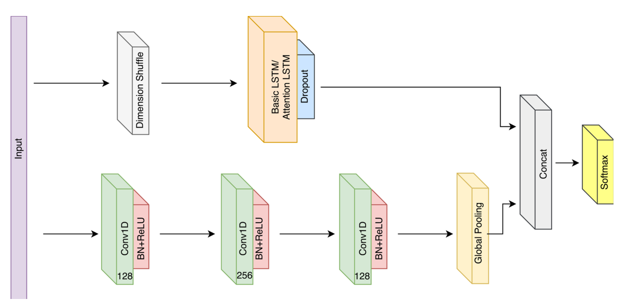


Fig. 1. MLSTM-FCN [21]

model. Next, all non-kinematic and non-time-based features are removed from the data, but the ICAO hex code and aircraft designator are kept. Finally, air tracks are analyzed to remove tracks that are either impossible or inconsistent.

The next phase of data preparation focuses on feature generation. Various features that are not available in the original data set are either inferred and generated using the data that is available. Latitude and longitude is a feature within the ADS-B dataset that has been determined to not be useful for classification purposes without manipulation [17], [19]. To improve its utility, latitude and longitude is altered to create a location that takes into account the Earth's 3-dimensional nature. Assuming latitude is ϕ and longitude is λ (1), (2) and

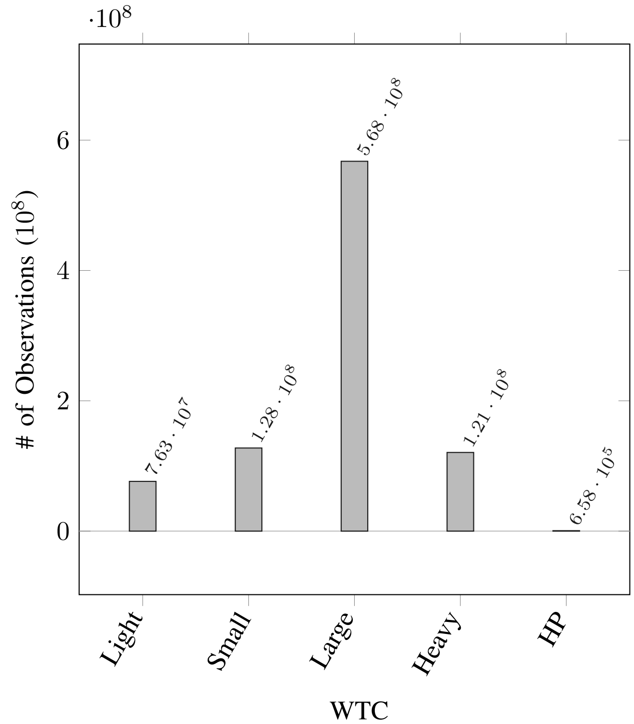


Fig. 2. Number of observations per WTC class

(3) become the new location coordinates.

$$x = \cos\phi * \cos\lambda \quad (1)$$

$$y = \cos\phi * \sin\lambda \quad (2)$$

$$z = \sin\phi \quad (3)$$

Since flight phase is not a feature that is inherent to ADS-B, the fuzzy logic flight phase method developed by [7] is used to determine the phase of each observation in the dataset. The rules for each phase are shown in table I. These features are added to the already processed data in preparation for the final steps.

In the last two steps, WTC and description are obtained from the Opensky Network Aircraft database [24]. These results are double checked with the ICAO Doc 8643 and added to the database. Finally, files are broken up into 300 time-step tensors and saved for model development.

For training, there are a total of 2,976,027 tensors available made up of 892,808,100 observations. However, since some aircraft do not have enough samples to train the model, only the 130 most common aircraft and the 9 high performance aircraft are used for training. This leaves 870,378,000 observations which form 2,901,260 tensors. Even with removing aircraft, there is still a huge imbalance between datasets. Figures 2 and 3 show the number of observations per class.

To handle the imbalanced dataset, previous research used the undersampling technique [17], [19]. However, as can be

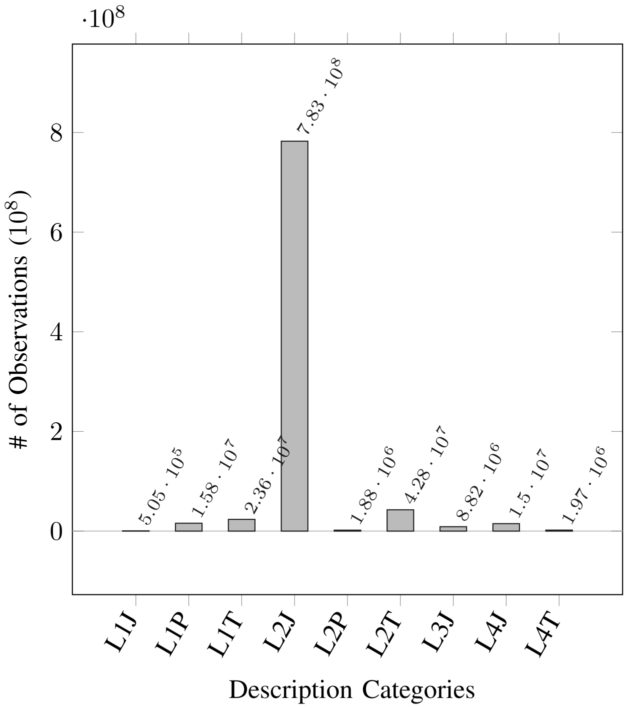


Fig. 3. Number of observations per description class

seen in figures 2 and 3, the undersampling technique would result in a large number of lost observations. For this research, class weights are used during the model's training to lessen the impact of the majority class instead of removing those observations entirely. The use of class weights is a function in Tensorflow to aid in the classification of imbalanced datasets [25].

B. Model Development

Previous work has shown the MLSTM-FCN to perform well with the ADS-B dataset [17], [19]. With the exception of the feature set selection, 50 epochs are used to train each model with data from the entire month of October 2021. November 1-2, 2021 is used to test the model. The models are built using dropout and learning rates that were found to work well in previous work [17]–[19]. The dropout rate remains at 0.5 and the learning rate is 0.001. Timesteps per tensor are kept at 300. While multiple feature sets are used to determine the one that provided the most accurate model, table III shows the features that were eventually settled on for analysis purposes. These features align closely to the most useful features found in [19].

C. Experiments

In order to predict WTC, description and designator, 3 experiments are created.

1) Compare two models to determine the effectiveness of identifying flight phase
2) Develop and analyze models for WTC, description and type desginator while accounting for class imbalances
3) Analyze and report the accuracy per flight phase

TABLE III
FEATURES [1]

Feature	*Definition*
alt_baro	Barometric altitude in feet as a number OR "ground" as a string
alt_geom	Geometric (GNSS / INS) altitude in feet referenced to the WGS84 ellipsoid
gs	Ground speed in knots
ias	Indicated air speed in knots
tas	True air speed in knots
track	True track over ground in degrees (0-359)
track_rate	Rate of change of track, degrees/second
baro_rate	Rate of change of barometric altitude, feet/minute
geom_rate	Rate of change of geometric (GNSS / INS) altitude, feet/minute
x, y, z	normalized location based on latitude and longitude
phase	flight phases listed in table I

1) Experiment 1 - Flight Phase Features: During the exploratory phase of this research, we build two models for each classifier without adding class weights. The two models use the features listed in table III. One of the models includes the flight phase and the other does not. The goal of this experiment is to determine the effectiveness of including flight phase as a one-hot-encoded feature within the dataset.

2) Experiment 2 - Model Development: In experiment 2, we develop models for each classifer with class weights. The class weights were calculated with (4). In this equation, w is weight, l is the size of the largest class, and c is the size of the class being calculated. This ensures that the largest class has a weight of one, and all smaller classes have a weight that is larger than one. The smaller the presence of the class, the greater the weight of that class.

$$w = \frac{l}{c} \tag{4}$$

3) Experiment 3 - Flight Phase Analysis: The accuracy per flight phase is calculated and reported using the test set to determine which flight phase provides the most accurate predictions of WTC, description and designator.

IV. Results and Discussion

A. Experiment 1 - Flight Phase Features

In experiment 1, we build two models for each prediction class. One that includes the flight phase in the feature set, and one that does not include flight phase. The phases are one-hot-encoded into one of the six options listed in table I. The results from this experiment are shown in table IV. As can be seen, WTC heavily depended on flight phase to provide improved results. With description and designator, the results were negligible.

B. Experiment 2 - Model Development

Experiment 2 accounts for the imbalanced dataset by weighing the classes to balance the impact each class has on the dataset. The results can be seen in table V. WTC improves with the weighted model, but description and designator have

TABLE IV
FLIGHT PHASE RESULTS AND STANDARD DEVIATIONS

Class	*Phase*	*Loss*	*Precision*	*Recall*	*Accuracy*
WTC	yes	.592 (.007)	.829 (.003)	.760 (.003)	.798 (.003)
	no	1.439 (.018)	.683 (.004)	.640 (.004)	.667 (.004)
Descr	yes	.252 (.006)	.935 (.006)	.925 (.007)	.929 (.007)
	no	.274 (.003)	.937 (.006)	.916 (.007)	.926 (.006)
Desig	yes	3.804 (.034)	.420 (.008)	.114 (.003)	.273 (.003)
	no	2.627 (.012)	.687 (.012)	.086 (.001)	.282 (.003)

much less favorable outcomes. Since weighing the classes fixes the imbalance problem, this is likely due to the fact that the imbalance between classes in WTC is less significant than balances between description and designator. However, weighing the classes is necessary to prevent the model from having inflated results for the majority class.

TABLE V
WEIGHTED MODEL RESULTS AND STANDARD DEVIATIONS

Class	*Loss*	*Precision*	*Recall*	*Accuracy*
WTC	.515 (.005)	.845 (.003)	.773 (.003)	.817 (.003)
Description	.743 (.008)	.781 (.005)	.620 (.006)	.722 (.004)
Designator	2.582 (.014)	.473 (.023)	.060 (.004)	.187 (.005)

C. Experiment 3 - Flight Phase Analysis

The final experiment focuses on the flight phases to determine how much the model's accuracy depends on that particular phase. The results of this analysis are shown in table VI. As expected from previous research [17], [19], the climb and descent phases produce the most accurate models in both WTC and description. The ground phase produces the least accurate model for WTC and description. The ground phase producing the least effective model is not surprising. The difference between aircraft as they are parked or taxiing is not significant. Most of the classification power of aircraft on the ground is likely related to other observations within the same tensor as the ground observations. Since the tensors are 300 timesteps long, each tensor lasts about 25 minutes. Over this 25 minute period, many of the tensors with ground observations also contain observations from the climb or descent phase.

V. CONCLUSION AND FUTURE WORK

Including all phases of flight creates difficulties in making predictions about aircraft characteristics due to the dynamic nature of air flight. However, the ability to predict aircraft characteristics anytime during flight is important since only obtaining data from takeoff or landing is not always feasible or desirable. This research shows that depending on which characteristic is being classified, sometimes generating features to predict flight phase is useful, but sometimes it is not. In this case, WTC benefits from flight phase prediction, but classification is not improved when predicting description and designator. When predicting characteristics from the entire flight path, it is important to determine if the characteristic being predicted is heavily influenced by flight phase or not.

TABLE VI
RESULTS PER PHASE

Class	*Phase*	*Precision*	*Recall*	*Accuracy*
WTC	Climb	0.82	0.89	0.92
	Cruise	0.75	0.75	0.82
	Descent	0.72	0.82	0.87
	Ground	0.60	0.74	0.73
	Level	0.71	0.78	0.79
	Unknown	0.81	0.84	0.87
Description	Climb	0.44	0.66	0.88
	Cruise	0.29	0.48	0.67
	Descent	0.34	0.56	0.80
	Ground	0.43	0.65	0.65
	Level	0.42	0.55	0.71
	Unknown	0.36	0.57	0.45
Designator	Climb	0.27	0.28	0.21
	Cruise	0.24	0.22	0.15
	Descent	0.25	0.24	0.14
	Ground	0.25	0.24	0.25
	Level	0.24	0.24	0.38
	Unknown	0.23	0.23	0.21

Having a dataset feature imbalance can falsely inflate the classification power of a model. In the case of aircraft description, the majority class accounts for 87.7% of the dataset. When weighing the classes so that they provide equal input, the accuracy of the model is reduced by 20% from 92.9% to 72.2%. The majority class in WTC is only 63.6% of the total observations causing the weighted version of the model to be less affected by the modification. However, the weighing of the classes ensures that the model learns to predict the actual features needed to learn about the aircraft instead of learning to favor the majority class.

With 139 different class options, creating an effective model to classify an aircraft designator is a difficult problem. During the model building phase, various architectures are attempted such as a Temporal Convolutional Network (TCN), a bidirectional LSTM, and a binary classifier modeled after the DSDEC method [17]. None of architectures showed promise after a handful of epochs and were abandoned to focus on the MLSTM-FCN [21]. Not accounting for the imbalance, we were able to achieve an accuracy of 28%, but with a recall of only 9%. Future research will look at including other sensors and data types such as weather or images to improve the classification ability when predicting aircraft designator.

ACKNOWLEDGMENT

The views expressed are those of the authors and do not reflect the official guidance or position of the United States Government, the Department of Defense or of the United States Air Force. The material was assigned a clearance of CLEARED on 21 March 2023. Originator Reference Number: AFIT-23-ENG03071. Number: 88ABW-2023-0237

REFERENCES

[1] ADS-B Exchange, "Home - serving the flight tracking enthusiast - ads-b exchange," Mar 2023. [Online]. Available: https://www.adsbexchange.com/

[2] FlightAware, "Fightaware - flight tracker/flight status," March 2023. [Online]. Available: https://flightaware.com/

[3] "Icao annex 10 volume 3," International Civil Aviation Organization, Quebec, Canada, Standard, 2007.

[4] L. Qian, N. Woods, and M. Rahman, "(U) identifying ads-b flight patterns," 2019, paper presented at the 2019 MSS Joint (BAMS and NSSDF) Conference, San Diego, CA, 21–24 October 2019.

[5] J. Sun, J. Ellerbroek, and J. Hoekstra, "Large-scale flight phase identification from ads-b data using machine learning methods," in *7th International Conference on Research in Air Transportation*, 2016, pp. 1–8.

[6] V. Kumar, S. Kalam, A. K. Das, and D. Sinha, "Attack detection scheme using deep learning approach for iot," *Advanced Computing and Systems for Security*, vol. 14, pp. 17–30, 2021.

[7] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight extraction and phase identification for large automatic dependent surveillance–broadcast datasets," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 566–572, 2017.

[8] S. T. Kanneganti, P. B. Chilson, and R. Huck, "Visualization and prediction of aircraft trajectory using ads-b," in *NAECON 2018-IEEE National Aerospace and Electronics Conference*. IEEE, 2018, pp. 529–532.

[9] F. Hasin, T. H. Munia, N. N. Zumu, and K. A. Taher, "Ads-b based air traffic management system using ethereum blockchain technology," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 346–350.

[10] N. Pearce, K. J. Duncan, and B. Jonas, "Signal discrimination and exploitation of ads-b transmission," in *SoutheastCon 2021*, 2021, pp. 1–4.

[11] T. Wu, S. Zhang, J. Yang, and P. Lei, "An ads-b signal poisoning method based on generative adversarial network," *Electronics Letters*, vol. 59, no. 2, p. e12699, 2023. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ell2.12699

[12] N. Ruseno, C.-Y. Lin, and S.-C. Chang, "Uas traffic management communications: The legacy of ads-b, new establishment of remote id, or leverage of ads-b-like systems?" *Drones*, vol. 6, no. 3, 2022. [Online]. Available: https://www.mdpi.com/2504-446X/6/3/57

[13] A. Filippone, B. Parkes, N. Bojdo, and T. Kelly, "Prediction of aircraft engine emissions using ads-b flight data," *The Aeronautical Journal*, vol. 125, no. 1288, p. 988–1012, 2021.

[14] F. Peyrin, P. Fréville, N. Montoux, and J.-L. Baray, "Original and low-cost ads-b system to fulfill air traffic safety obligations during high power lidar operation," *Sensors*, vol. 23, no. 6, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/6/2899

[15] M. Yue, H. Zheng, H. Cui, and Z. Wu, "Gan-lstm-based ads-b attack detection in the context of air traffic control," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[16] R. Ginoulhac, F. Barbaresco, J. Y. Schneider, J. M. Pannier, and S. Savary, "Target classification based on kinematic data from AIS/ADS-B, using statistical features extraction and boosting," *Proceedings International Radar Symposium*, vol. 2019-June, no. 681, pp. 1–10, 2019.

[17] K. Basrawi, R. Dill, G. Peterson, B. Borghetti, and J. Lopez, "Aircraft identification from ads-b kinematic data," *Journal of DoD Research and Engineering*, vol. 5, no. 2, pp. 31–43, 2022.

[18] K. Basrawi, "Aircraft classification from ads-b kinematic data," Ph.D. dissertation, Air Force Institute of Technology, 2021.

[19] S. Bolton, R. Dill, M. R. Grimaila, and D. Hodson, "Ads-b classification using multivariate long short-term memory–fully convolutional networks and data reduction techniques," *The Journal of Supercomputing*, vol. 79, no. 2, pp. 2281–2307, 2023.

[20] Q. Zhang, J. H. Mott, M. E. Johnson, and J. A. Springer, "Development of a reliable method for general aviation flight phase identification," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[21] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate lstm-fcns for time series classification," *Neural Networks*, vol. 116, pp. 237–245, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019301200

[22] "Icao document 8643," International Civil Aviation Organization, Quebec, Canada, Standard, 2016.

[23] L. Ren, M. Castillo-Effen, H. Yu, E. Johnson, Y. Yoon, T. Nakamura, and C. Ippolito, "Small unmanned aircraft system (suas) categorization framework v2.0," 02 2019.

[24] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, and M. Wilhelm, "Bringing up opensky: A large-scale ads-b sensor network for research," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, 2014, pp. 83–94.

[25] "Classification on imbalanced data: Tensorflow core," Nov 2022. [Online]. Available: https://www.tensorflow.org/tutorials/structured$_{data}/imbalanced_{data}$

V. Study 3 - Sensor Fusion

The following paper, “Multi-sensor Aircraft Classification,” was submitted and accepted by the 2023 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE’23). This paper is in IEEE format.

This research uses Automatic Dependent Surveillance-Broadcast (ADS-B) kinematic data, weather and aircraft images to predict Wake Turbulence Category (WTC), description and designator as depicted in the ICAO Document 8643 [48] to validate the integrity of ADS-B observations. The paper notes that ADS-B is vulnerable to inaccuracies (both intentional and unintentional), but more passive forms of aircraft tracking such as 3D primary radar, are not. Fusing other data with kinematic ADS-B data allows ADS-B to be used as a surrogate for more accurate data. If other data types are not available, the fusion of other senors can also validate the accuracy of ADS-B tracks. Similar to previous work, the Multivariate Long Short-Term Memory - Fully Convolutional Network (MLSTM-FCN) is used to build the models with ADS-B and weather data. The Cross-Layer Mutual Attention Learning Network (CMAL-Net) is used for image classification. It was found that the addition of weather data improved designator and description, but did not improve WTC. The addition of images improved the results in all class types over just ADS-B or ADS-B and weather. In the case of description and WTC, the addition of ADS-B improved the classification accuracy when compared to only images.

The paper below is broken up into five sections. First, it introduces ADS-B regulations and vulnerabilities. Then, it provides a background on previous research in aircraft prediction using ADS-B, an overview of the algorithms involved in ADS-B classification, research in aircraft image classification and research with sensor fusion. Next, the paper presents the experimental setup for ADS-B classification models, aircraft image classification models, and the fusion of those models. The paper ends with the results and conclusion. The research questions associated with this paper can be seen in Table 9.

Table 9: Summary of Research Questions for Paper 3

ID	**Details**
RQ-5	Does fusing weather and image data improve the predictive power of a model that uses ADS-B data to predict description, designator and WTC?
RQ-6	Using pre- and/or post-classification fusion methods, how can a model be developed to predict description, designator and WTC using ADS-B, aircraft images and weather?

Multi-Sensor Aircraft Classification

Sarah Bolton
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
sarah.bolton@afit.edu

Richard Dill
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
richard.dill@afit.edu

Michael R. Grimaila
Department of Systems Engineering and Management
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
michael.grimaila@afit.edu

Douglas D. Hodson
Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA
douglas.hodson@afit.edu

***Abstract*—Automatic Dependent Surveillance–Broadcast (ADS-B) is a useful tool to for air traffic controllers, military and other sources that are invested in understanding a national or global air picture. While it is highly available, it can sometimes lack integrity due to hacking, spoofing or, even, unintentional inaccuracies in the broadcast. Unlike primary radar, ADS-B's lack of trustworthiness makes it not feasible to rely on it alone. Fusing other data sources with ADS-B can help confirm the accuracy of the broadcasts or allow ADS-B to act as a surrogate for primary radar and bolster the information that primary radar can provide. This paper presents an effective method of using ADS-B data as a surrogate for primary 3D radar by combining the kinematic information that ADS-B data provides with weather and aircraft images to make predictions about aircraft characteristics.**

***Index Terms*—Multivariate Long Short-Term Memory – Fully Convolutional Network, Automatic Dependent Surveillance-Broadcast, open-source data, classification, machine learning, sensor fusion**

I. Introduction

Automatic Dependent Surveillance–Broadcast (ADS-B) has become an important research tool over the last decade. Since its inception in 1995, many countries have begun to mandate the use of ADS-B within controlled airspace. Notably, the United States mandated the use of ADS-B within controlled airspace on 1 January 2020 with Federal Regulation 14 CFR 91.225 and 14 CFR 91.227 which was followed by Europe's ADS-B mandate in June 2020 with the Commission Implementing Regulation (EU) Number 1028/2014. These mandates apply to a large number of aircraft even if they are not based in Europe or the United States since international flights that fly through either of these areas are also required to comply with the mandate. While the exact number of aircraft that must comply with these regulations is difficult to determine, ADS-B repositories, such as the ADS-B Exchange, collect flight information on over 15,000 aircraft daily [1]. In the US alone, 163,216 aircraft were ADS-B compliant as of April 2023 [2].

Funding provided by the Air Force Research Laboratory.

One of the main features of the ADS-B system is open sharing and wide availability. For ease of access, there is no encryption or authentication. This makes ADS-B extremely vulnerable to cyber attacks and other inaccuracies. Jamming, eavesdropping, spoofing and injections of fake tracks would be easy for an attacker since ground stations cannot distinguish the difference between a real broadcast or a fake one. Since much of the broadcasted data is in the form of free text, even typos are an issue. Costin et al proved the vulnerabilities to ADS-B in 2012 [3]. For this reason, it's important to have a backup source of information when using ADS-B to ensure an accurate air picture. While ADS-B provides plenty of useful information for aircraft tracking, it cannot entirely replace primary radar due to its lack of integrity.

ADS-B's vulnerability concerns make fusing the ADS-B broadcast with other related sensors an important research area. While ADS-B on its own is not entirely trustworthy, using a backup sensor, like primary radar or images of the aircraft, allows ADS-B receivers to confirm the information being provided by the broadcast. For this research, we use only the kinematic data (speed, altitude, track and location) within ADS-B so that our model could act as a surrogate for a non-ADS-B data source. The kinematic data is able to simulate primary 3D radar and potentially eliminate the need to rely on ADS-B altogether. Using the kinematic data within ADS-B, this paper presents a method to predict an aircraft's wake turbulence category (WTC), description and designator as described in the ICAO doc 8643 [4]. To improve the accuracy of predictions using only kinematic data, we present a method of combining weather with the ADS-B data. Then, to ensure the integrity of the data, we use images to verify the authenticity of the ADS-B transmission. Section II discusses relevant literature. Section III reviews how the research and experiments were setup. Section IV presents the results from the research. Finally, we conclude in section V.

II. Background & Previous Work

A. Aircraft Prediction with ADS-B

ADS-B provides aircraft tracking and identification information via an onboard ADS-B Out transponder. It is broadcast in the clear at 1090 MHz in many countries worldwide and at both 1090 MHz and 978 MHz within the United States. In recent years, ADS-B has become an important tool for researchers. Due to its prevalence, ease of use, wide availability, and the dense information contained within the broadcast, ADS-B has been shown to be useful for a wide variety of aircraft research topics. The areas of research most related to this paper include enhancing aircraft operations [5]–[8], predicting flight patterns [9]–[13], identifying vulnerabilities [3], improving ADS-B transmission security [14]–[17] and predicting aircraft characteristics [18]–[22].

Created by Karim et al. in 2017, the Multivariate Long Short-Term Memory - Fully Convolutional Network (MLSTM-FCN) has been shown be effective in predicting characteristics about aircraft using the kinematic data found within ADS-B [19]–[23]. [20] and [19] used a Dual-Stage Deep Engine Classifier (DSDEC) that implemented the MLSTM-FCN to make predictions about aircraft engine types. They were able to achieve an overall accuracy of 89.2% with jet engine accuracy at 98.4%, turboprop accuracy at 79.2% and piston engine accuracy at 89.9% when looking at only the take-off phase of flight. The research completed in [22] predicted more definitive aircraft characteristics, specifically, WTC, description and designator using all flight phases instead of only the take off phase. While the accuracy was reduced overall due to the use of the entire flight path and the increase in prediction classes (3 engine types vs 5 WTC classes and 9 description classes), the take-off phase was able to produce slightly improved accuracy to [19] when predicting WTC and description.

B. Image Classification

While unrelated to ADS-B aircraft classification, aircraft image classification is important to understand for the purposes of sensor fusion. Many techniques have been shown to be effective in classifying an aircraft model based on its image. In 2013, Maji et al. developed an image benchmark that was gathered mostly from https://www.airliners.net/ to facilitate a baseline for other researchers with a focus on aircraft image classification [24]. The image repository, called the Fine-Grained Visual Classification of Aircraft (FGVC-Aircraft) benchmark, has been cited nearly 1,000 times in other aircraft classification research. The benchmark includes 10,000 airplane images spanning 100 different models. Each image is organized hierarchically by model, variant, family and manufacturer.

In their initial attempt at classifying the images in the benchmark, Maji et al. achieved a variant classification accuracy of 48.69% with a non-linear support vector machine [24]. More recent research improved the classification accuracy to about a 90% variant classification accuracy. In 2021, Rong et al. achieved a 93.3% accuracy using a Separated Smooth Sampling Network (SSSNet) [25]. In 2022, Yanfeng Wang et al. achieved a 89.2% accuracy using a Bat Algorithm and Convolutional Neural Network (BA-CNN) [26]. Also in 2022, Lei Wang et al. achieved a 91.4% accuracy using a Multilayer Feature Fusion (MFF) network with Parallel Convolutional Block (PCB) mechanism [27]. A more recent effort was completed by Liu et al. in March 2023 that developed a Cross-Layer Mutual Attention Learning Network (CMAL-Net) which achieved 94.7% accuracy [28]. Their CMAL-Net uses a Residual Network (Resnet) as its backbone. The architecture develops multiple classifiers that are trained to make predictions based on a specific layer from shallow to deep. The prediction from each layer is combined to create the final output. The backbone, the Resnet, is a Convolutional Neural Network (CNN) that uses a technique called skip connections to solve the exploding and vanishing gradient problems.

C. Sensor Fusion

Information fusion is described as the blending of data acquired from multiple sources. Sensor fusion is a subset of information fusion and is the merging of data derived from only sensory sources [29]. In general, sensor/information fusion provides better reliability, improved coverage, increased confidence, reduced ambiguity, better resolution and is more robust than using a single data source [29].

Sensor fusion is a vast area of study with no common model or architecture. However, one way to split sensor fusion methods is by the two predominant areas when it can occur, pre-classification or post-classification [30]. Like the name implies, pre-classification occurs before the model is built and post-classification occurs afterwards. Some research refers to pre-classification as low or intermediate-level fusion and post-classification as high or decision-level fusion [29], [30].

The most straightforward method to fuse data is a pre-classification method called data level fusion. In data level fusion, the raw data is combined before training the model. The model is built using data from all fused sources. Another pre-classification method, feature level fusion, uses data association and grouping techniques to improve performance over just the use of the raw data. Post-classification fuses the output or prediction from models that were built using a single data source. A few common methods include voting, ranking, fuzzy logic and other statistical methods. In the case of ADS-B, fusion from other sensor sources could be invaluable since the integrity of ADS-B data cannot be guaranteed.

III. Methodology

The intent behind this research is to fuse multiple aircraft data sources to predict aircraft characteristics. We will make predictions by training a model to determine an aircraft's WTC, description and designator as listed in the ICAO doc 8643. ADS-B, weather and images are selected as the data sources to be combined. Weather and ADS-B are fused during pre-processing, while aircraft images are fused during post-processing. Due to the success in previous research, the MLSTM-FCN model trains the ADS-B and weather models while the CMAL-Net trains the image models [21], [28], [31]. The steps required to fuse the three data sources can be seen in Figure 1.

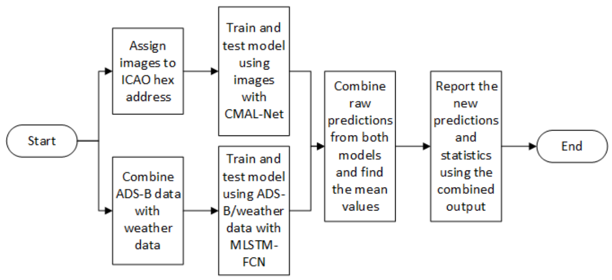


Fig. 1. Flowchart to combine ADS-B, weather and image data

A. Data Preparation

The ADS-B data in this research is obtained from the ADS-B Exchange [1]. The original dataset spans 1 Jan 2021 to 17 Jan 2022 and is 3.7 TB of ADS-B data. However, this research focuses on 1 Oct 2021 to 2 November 2021 with October being used for the training set and November being used for the test set. After obtaining the data, the files are converted from JSON to CSVs, cleaned and broken into 300-time step tensors.

The weather data, obtained from Meteostat [32], is shown in Table I. Calls are made to Meteostat for each ADS-B tensor in order to obtain weather for the timeframe and location of the aircraft during the tensor. The weather data is combined with the ADS-B data and saved as a CSV file.

TABLE I
WEATHER FEATURES USED FROM METEOSTAT

Feature	***Description***	***Type***
temp	The air temperature in °C	Float64
dwpt	The dew point in °C	Float64
rhum	The relative humidity in percent (%)	Float64
prcp	The one hour precipitation total in mm	Float64
snow	The snow depth in mm	Float64
wdir	The average wind direction in degrees (°)	Float64
wspd	The average wind speed in km/h	Float64
pres	The average sea-level air pressure in hPa	Float64

The image data is obtained from the FGVC-Aircraft benchmark [24]. Since there are aircraft included in the FGVC-Aircraft benchmark where the ADS-B data does not have enough samples to train a model, and there are aircraft present in ADS-B data that the FGVC-Aircraft benchmark does not utilize (i.e. the two sets are not equal), the intersection of aircraft classes between the ADS-B data and the FGVC-Aircraft images is found. The similarities among the two sets results in 55 remaining aircraft for the classification model. All other aircraft samples are removed from both datasets. In the training set this produces a total of 325,822,800 observations which translates to 1,086,076 tensors or approximately 452,5322 flight hours. For testing, the intersection yields 67,435 ADS-B/weather tensors which is equivalent to 20,230,500 timesteps or roughly 28,098 hours of flight data. In the FGVC-Benchmark, the remaining images include 3,736 training images and 1,864 testing images.

Since ICAO hex address is not annotated in the FGVC-Aircraft benchmark, each ICAO hex address within the ADS-B data is randomly assigned an image that matches its designator. Due to the number of ICAO hex addresses outnumbering the number of images, each image could be assigned to multiple ICAO hex addresses. This is assumed to be a negligible concern due to the random assignment.

B. Model Development

A total of nine models are created. Three models are trained on ADS-B only, three models are trained on both ADS-B and weather, and three models are trained on the image benchmark. The six models created with the ADS-B and weather data use the MLSTM-FCN algorithm to classify each of the aircraft characteristics of interest (WTC, description and designator). Each ADS-B/weather model is trained for 150 epochs with a dropout of 0.5 and a learning rate of 0.001. These parameters are selected due to success with previous research in this area [21], [22]. Using the test set, precision, recall, accuracy, loss and F1 scores are found. Raw predictions are recorded for each model. Similarly, three models are trained on the FGVC-Aircraft images using the CMAL-Net architecture for 200 epochs. Loss, precision, recall, accuracy, F1 scores and raw predictions are recorded. The list of models can be seen in table II.

TABLE II
MODELS

Model #	***Input Data***	***Algorithm***	***Prediction***
1	ADS-B	MLSTM-FCN	WTC
2	ADS-B	MLSTM-FCN	Description
3	ADS-B	MLSTM-FCN	Designator
4	ADS-B & Weather	MLSTM-FCN	WTC
5	ADS-B & Weather	MLSTM-FCN	Description
6	ADS-B & Weather	MLSTM-FCN	Designator
7	Images	CMAL-Net	WTC
8	Images	CMAL-Net	Description
9	Images	CMAL-Net	Designator

C. Post-processing

Since the results for the fusion between the weather and ADS-B models is found by running the MLSTM-FCN model on the test set, the bulk of post-processing focuses on combining the outputs from the models trained by the MLSTM-FCN and the image models trained by the CMAL-Net. To combine the outputs from the nine models, the raw predictions for each class are normalized per tensor. The predictions from the six ADS-B models and the three image models are combined and the element-wise mean is found for each class per tensor. The class with the largest mean value per tensor is selected as the predicted class for that tensor. The predicted classes are compared to the actual classes for each tensor to determine the precision, recall, accuracy and F1 score.

IV. Results and Discussion

The results from the combination of weather and ADS-B data using the MLSTM-FCN can be seen in Table III. Both predictions on ADS-B alone and predictions with ADS-B and weather are represented in the table. As can be seen, adding the weather data provided improvements with the models that predicted description and designator, but did not improve the WTC model.

TABLE III
ADS-B and Weather Results

Class	***Wx***	***Loss***	***Precision***	***Recall***	***Accuracy***	***F1***
WTC	No	0.370	0.873	0.869	0.870	0.871
	Yes	0.389	0.859	0.853	0.857	0.856
Description	No	0.223	0.953	0.941	0.947	0.947
	Yes	0.233	0.958	0.929	0.949	0.943
Designator	No	1.849	0.619	0.201	0.410	0.303
	Yes	1.748	0.638	0.253	0.448	0.362

The results for the FGVC-Aircraft benchmark image classification can be seen in Table IV. The models performed slightly better than previous research due to the reduction in the number of aircraft samples [24]–[28].

TABLE IV
FGVC-Aircraft Image Results

Class	***Loss***	***Precision***	***Recall***	***Accuracy***	***F1 Score***
WTC	0.090	0.983	0.984	0.983	0.984
Description	0.083	0.989	0.969	0.987	0.978
Designator	0.283	0.954	0.953	0.953	0.953

The fusion between all three data sources can be seen in Table V. In all cases, the image data improved the results compared to the ADS-B and weather data alone. Additionally, adding the ADS-B and weather data to the image data further improved the results for WTC and description when compared to just images alone. With designator, the accuracy was reduced by 3% compared to the image classification by itself. This is likely due to the low accuracy of predicting designator using only ADS-B and weather.

TABLE V
ADS-B with FGVC Aircraft Results

Class	***Wx***	***Precision***	***Recall***	***Accuracy***	***F1 Score***
WTC	No	0.9880	0.9750	0.9898	0.9814
	Yes	0.9847	0.9770	0.9890	0.9808
Description	No	0.9693	0.9113	0.9956	0.9334
	Yes	0.9736	0.9323	0.9960	0.9493
Designator	No	0.8972	0.9435	0.9245	0.9152
	Yes	0.8893	0.9316	0.9201	0.9049

V. Conclusion

In this work we propose an architecture for combining the kinematic data found within ADS-B with weather and aircraft images. Experiments using this method demonstrate when looking at sensor fusion that occurs during pre-processing (i.e combining weather with ADS-B kinematic data) there are negligible improvements when predicting WTC and description, but there is a slight (4%) improvement when predicting designator. When looking at the entire architecture, there is near perfect accuracy when predicting description and 99% accuracy when predicting WTC. This is a 1-2% improvement over using only images. Designator's accuracy is slightly reduced from the accuracy with only images (-5%), but is nearly doubled when making predictions from only ADS-B and weather.

The architecture has two main advantages. First, with only using the kinematic data found within ADS-B, it allows the model to utilize ADS-B as a surrogate for primary 3D radar and potentially eliminate the need for ADS-B altogether. Second, it allows the use of ADS-B in combination with other sources of aircraft data to verify the accuracy of ADS-B. This method could easily be modified to provide more weight to the image results which would further improve the accuracy. It could also allow for predictions to be made without images, but would provide a smaller confidence value on its accuracy.

Acknowledgment

The views expressed are those of the authors and do not reflect the official guidance or position of the United States Government, the Department of Defense or of the United States Air Force. The material was assigned a clearance of CLEARED on 08 JUN 2023. Originator Reference Number: AFIT-23-ENV03100. Number: 88ABW-2023-0601

References

[1] ADS-B Exchange, "Home - serving the flight tracking enthusiast - ads-b exchange," Mar 2023. [Online]. Available: https://www.adsbexchange.com/

[2] United States Department of Transportation, "Current equipage levels," April 2023. [Online]. Available: https://www.faa.gov/air_traffic/technology/equipadsb/installation/current_equipage_levels

[3] A. Costin and A. Francillon, "Ghost in the air(traffic): On insecurity of ads-b protocol and practical attacks on ads-b devices," in *BLACKHAT 2012, July 21-26, 2012, Las Vegas, NV, USA*, EURECOM, Ed., Las Vegas, 2012, © EURECOM. Personal use of this material is permitted. The definitive version of this paper was published in BLACKHAT 2012, July 21-26, 2012, Las Vegas, NV, USA and is available at :.

[4] "Icao document 8643," International Civil Aviation Organization, Quebec, Canada, Standard, 2016.

[5] N. Ruseno, C.-Y. Lin, and S.-C. Chang, "Uas traffic management communications: The legacy of ads-b, new establishment of remote id, or leverage of ads-b-like systems?" *Drones*, vol. 6, no. 3, 2022. [Online]. Available: https://www.mdpi.com/2504-446X/6/3/57

[6] A. Filippone, B. Parkes, N. Bojdo, and T. Kelly, "Prediction of aircraft engine emissions using ads-b flight data," *The Aeronautical Journal*, vol. 125, no. 1288, p. 988–1012, 2021.

[7] F. Peyrin, P. Fréville, N. Montoux, and J.-L. Baray, "Original and low-cost ads-b system to fulfill air traffic safety obligations during high power lidar operation," *Sensors*, vol. 23, no. 6, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/6/2899

[8] M. Yue, H. Zheng, H. Cui, and Z. Wu, "Gan-lstm-based ads-b attack detection in the context of air traffic control," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[9] L. Qian, N. Woods, and M. Rahman, "(U) identifying ads-b flight patterns," 2019, paper presented at the 2019 MSS Joint (BAMS and NSSDF) Conference, San Diego, CA, 21–24 October 2019.

[10] J. Sun, J. Ellerbroek, and J. Hoekstra, "Large-scale flight phase identification from ads-b data using machine learning methods," in *7th International Conference on Research in Air Transportation*, 2016, pp. 1–8.

[11] V. Kumar, S. Kalam, A. K. Das, and D. Sinha, "Attack detection scheme using deep learning approach for iot," *Advanced Computing and Systems for Security*, vol. 14, pp. 17–30, 2021.

[12] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight extraction and phase identification for large automatic dependent surveillance–broadcast datasets," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 566–572, 2017.

[13] S. T. Kanneganti, P. B. Chilson, and R. Huck, "Visualization and prediction of aircraft trajectory using ads-b," in *NAECON 2018-IEEE National Aerospace and Electronics Conference*. IEEE, 2018, pp. 529–532.

[14] F. Hasin, T. H. Munia, N. N. Zumu, and K. A. Taher, "Ads-b based air traffic management system using ethereum blockchain technology," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 346–350.

[15] N. Pearce, K. J. Duncan, and B. Jonas, "Signal discrimination and exploitation of ads-b transmission," in *SoutheastCon 2021*, 2021, pp. 1–4.

[16] T. Wu, S. Zhang, J. Yang, and P. Lei, "An ads-b signal poisoning method based on generative adversarial network," *Electronics Letters*, vol. 59, no. 2, p. e12699, 2023. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ell2.12699

[17] J. Wang, Y. Zou, and J. Ding, "ADS-B spoofing attack detection method based on LSTM," *EURASIP Journal on Wireless Communications and Networking*, August 2020.

[18] R. Ginoulhac, F. Barbaresco, J. Y. Schneider, J. M. Pannier, and S. Savary, "Target classification based on kinematic data from AIS/ADS-B, using statistical features extraction and boosting," *Proceedings International Radar Symposium*, vol. 2019-June, no. 681, pp. 1–10, 2019.

[19] K. Basrawi, R. Dill, G. Peterson, B. Borghetti, and J. Lopez, "Aircraft identification from ads-b kinematic data," *Journal of DoD Research and Engineering*, vol. 5, no. 2, pp. 31–43, 2022.

[20] K. Basrawi, "Aircraft classification from ads-b kinematic data," Ph.D. dissertation, Air Force Institute of Technology, 2021.

[21] S. Bolton, R. Dill, M. R. Grimaila, and D. Hodson, "Ads-b classification using multivariate long short-term memory–fully convolutional networks and data reduction techniques," *The Journal of Supercomputing*, vol. 79, no. 2, pp. 2281–2307, 2023.

[22] S. Bolton, R. Dill, M. Grimaila, and D. Hodson, "Aircraft classification using flight phase identification," 2023, unpublished.

[23] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[24] S. Maji, E. Rahtu, J. Kannala, M. B. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," *CoRR*, vol. abs/1306.5151, 2013. [Online]. Available: http://arxiv.org/abs/1306.5151

[25] S. Rong, Z. Wang, and J. Wang, "Separated smooth sampling for fine-grained image classification," *Neurocomputing*, vol. 461, pp. 350–359, 2021.

[26] Y. Wang, Y. Chen, and R. Liu, "Aircraft image recognition network based on hybrid attention mechanism," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[27] L. Wang, K. He, X. Feng, and X. Ma, "Multilayer feature fusion with parallel convolutional block for fine-grained image classification," *Applied Intelligence*, vol. 52, no. 3, pp. 2872–2883, 2022.

[28] D. Liu, L. Zhao, Y. Wang, and J. Kato, "Learn from each other to classify better: Cross-layer mutual attention learning for fine-grained visual classification," *Pattern Recognition*, vol. 140, p. 109550, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320323002509

[29] W. Elmenreich, "An introduction to sensor fusion," *Vienna University of Technology, Austria*, vol. 502, pp. 1–28, 2002.

[30] A. Vakil, J. Liu, P. Zulch, E. Blasch, R. Ewing, and J. Li, "A survey of multimodal sensor fusion for passive rf and eo information integration," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 7, pp. 44–61, 2021.

[31] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate lstm-fcns for time series classification," *Neural Networks*, vol. 116, pp. 237–245, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019301200

[32] C. S. Lamprecht, "Meteostat Python."

VI. Conclusion

This dissertation has discussed methods of aircraft pattern-of-life (POL) modeling using Automatic Dependent Surveillance-Broadcast (ADS-B) data and deep learning. Chapter I provided an overview of the importance of POL modeling to the Department of the Air Force (DAF) and other defense leaders. It discussed how ADS-B data could be used as a surrogate data set for other kinematic data that is not attributable. Finally, it reviewed the research questions that will provide defense personnel with tools to improve Intelligence, Surveillance and Reconnaissance (ISR) operations. Chapter II surveyed relevant literature in aircraft standards, deep learning, previous aircraft classification research and sensor fusion. Chapter III presented the journal article published in the Journal of Supercomputing that discussed data engineering techniques to optimize ADS-B features and data set size to predict engine type during takeoff. Chapter IV presented the paper submitted to the National Aerospace and Electronics Conference (NAECON), which used ADS-B data with flight phase prediction to classify Wake Turbulence Category (WTC), description and designator using the entire flight path. Chapter V presented the paper published for the World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'23) which summarized the research completed with sensor fusion of ADS-B data, aircraft images and weather. This section summarizes the findings of this dissertation, discusses the contributions and suggests future research.

6.1 Summary of Findings

The findings of this dissertation provide a method to use ADS-B data as a surrogate for classification tasks on unattributed aircraft data. The next few subsections are separated by the study that was completed to summarize the findings from each.

6.1.1 Study 1 - Feature Reduction and Data Minimization

The goal of this paper is to determine the effect of minimizing the size of the training data that is used to develop a Multivariate Long Short-Term Memory - Fully Convolutional Network (MLSTM-FCN) model. We did this with two separate experiments. In experiment one, we vary the number of training samples in the training data set with three different feature sets. In experiment two, we vary the number of features present in the data set. Those models are tested to determine how each change affected the accuracy and loss of the model against a separate test set. We deem a model to be "acceptable" if the accuracy was within 10% of results from previous research. Since previous research was able to achieve an 89.2% accuracy, any accuracy above 79.2% would be considered acceptable [20].

There are a few main findings from this experiment. The first is that this experiment is able to achieve an 89.4% accuracy with its best model. This is a slight improvement over the model developed by Basrawi et al which achieved 89.2% accuracy [20]. The worst performing models perform poorly due to the quality of certain features. The experiments show that quality of a data set is just as important as the quantity of data when building a neural network. Having redundant features, badly optimized features or features that do not add pertinent information to the model can cause the model to perform poorly. Too many irrelevant features will add noise into the model and produce results that are less than ideal. Based on this experiment, speed, pressure and vertical speed are shown to be the most important features needed to identify an aircraft's engine type while achieving a 89.4% accuracy. The noisy feature in the data is time. The time feature was not properly optimized and detracted from the training of the model. All of the 42 worst performing models (out of 255 total models) include the time feature with the worst performing accuracy at 7.4%.

This data set reduces to a quarter of its original size before the model starts exhibiting severe negative effects. To compare each model, we use the rate of change formula as can be seen in equation 5.

$$\frac{\Delta accuracy}{\Delta size} \tag{5}$$

During both experiments, the best model achieves an accuracy of 89.4%. Reducing the data by half only reduces the accuracy by 4.6% to 84.8% accuracy which is a rate of change of 9.2%. Reducing it to a quarter of its original size produces an 83.9% accuracy rate which is a 22.0% rate of change from the full size data set. However, reducing the data set to an eighth of its original size produces a 75.8% accuracy rate which is a 108.8% rate of change from the full size data set. Since the raw data from 1 December 2020 was about 44 GB, this large file size could be reduced to 11 GB using the quarter size data set and still effectively train a classification model. Most computers, laptops, or small handheld devices could perform data processing and train a model that was only 11 GB. This observation would work well for military operations in remote locations with low computational resources. If data and computation resources were not as plentiful, reducing the model to approximately 300,000 observations would create a model that could make observations with a small trade off of less than 10% reduced accuracy.

6.1.2 Study 2 - Aircraft Prediction

The aircraft prediction study uses a much larger data set than study 1 with the goal of classifying aircraft with more specificity. This experiment uses ADS-B data from 1-31 October 2021 to train various models to classify an aircraft's WTC, description and type designator. The models are tested with data from 1-2 November 2021. The study is broken into three experiments. In the first experiment, flight phase features are added to the data set using fuzzy logic. Then, models are built with and without flight phase to determine the impact of having flight phase part of the data set. It was found that WTC depended heavily on flight phase, whereas it was less significant for description and designator.

The second experiment focuses on the class imbalance. Previous research either disregarded the imbalance or used the undersampling technique to remedy the imbalance. Ignoring the imbalance inflates the accuracy of the model and undersampling unnecessarily removes a large amount of available data. This research uses a weight system instead where classes that are less present in the data are given a higher weight than ones that saturate the data. Accuracy for WTC improved with the weighted model, but description and designator did not. However, loss improved with both WTC and designator.

The final experiment focuses on determining which flight phase provides the highest accuracy predictions. It does this by providing the accuracy, precision and recall during each flight phase. As expected from previous research, the climb and descent phases produced the most accurate models. The ground phase produced the least accurate models.

As hypothesized in the prospectus, WTC and description are easy to predict with high accuracy. WTC, which the ADS-B data set expresses as emitter category, has only 5 values. These values represent the size of the aircraft. Since aircraft of different sizes, generally fly at different altitudes and speeds, the models are able to determine the difference between categories. Similarly, the description has at most 120 unique values. However, this research does not consider helicopters or gyrocopters, which makes only 12 of those values applicable to this research. Those 12 values describe the type of aircraft (i.e. jet, amphibious, seaplane, etc), the number of engines and the type of engine. Due to the relatively low number of options, models that predict description are able to achieve a high accuracy. To the contrary, designator has a large number of options. Per Uniting Aviation, there are nearly 10,000 type designators in use worldwide [85]. 750 of those type designators show up in the ADS-B data set in October 2021. Since many of the aircraft in the data set fly in a similar manner, it is difficult to achieve a high accuracy level.

This study provides a means to utilize all flight phases to predict WTC and description while accounting for the class imbalance. The accuracy when trying to predict designator was low and will require more research before aircraft designator can be predicted from only kinematic data. Likely, the fusion of additional data set types with ADS-B will be required to help distinguish between between aircraft that fly in a similar manner. However, predicting WTC and description is a huge step in classifying an aircraft. While the specific air frame cannot be detected, its general shape and size can be classified with a high degree of accuracy.

6.1.3 Study 3 - Sensor Fusion

Study 3 uses the same ADS-B data set as study 2 to predict an aircraft's WTC, description and type designator. The difference between this study and study 2 is that the goal of this study is to fuse the ADS-B data set with image and weather data. Study 3 focuses on obtaining, downloading, formatting and building deep learning models from image and weather data to work with the ADS-B data set. Ten-thousand attributable aircraft images are identified and downloaded from the Fine-Grained Visual Classification of Aircraft (FGVC-Aircraft) Benchmark that has been used for previous aircraft classification research efforts and deep learning image classification challenges [17]. Recent classification work with the FGVC-Aircraft Benchmark using a Cross-Layer Mutual Attention Learning Network (CMAL-Net) has achieved an 94.7% accuracy [72]. Weather data is obtained from Meteostat [86].

Using pre-processing techniques, nine models are built that predict WTC, description and designator on only ADS-B data, ADS-B data and weather, and only images. During post-processing, ADS-B data, weather and images are combined for the final results. When combining ADS-B data with weather, description and designator showed improvement with the addition of weather, but WTC did not. In the combination of all three data types, image predictions improved the ADS-B/weather predictions in all cases. Additionally, ADS-B/weather predictions improved upon image predictions in the case of WTC and description. These results are important since with the addition of images to the ADS-B data set, predicting designator becomes possible where it wasn't with only ADS-B data. With this research, intelligence analysts would be able to use ADS-B data as a surrogate for combining kinetic aircraft data with images and weather to predict aircraft characteristics.

6.2 Research Contributions

This section reviews the contributions discussed in Chapters III, IV and V. Table 10 summarizes those contributions and references the corresponding chapter and research question (RQ) where appropriate.

The primary contribution of this dissertation is the development of models and techniques that allow for the identification and classification of aircraft using kinetic and unattributed data. The models developed in this dissertation can be used as surrogate data for sources of kinematic aircraft data, such as 3D primary radar, and will allow for improved capabilities within the intelligence community. Since, unlike computers, the human brain struggles to process large amounts of textual data, intelligence analysts will have another tool at their disposal to assist in making educated analyses. Combining this work with already completed flight trajectory prediction research [82] will give defense personnel the ability to know more about enemy aircraft. Not only will they know where the aircraft has been, but, now, they will have a reasonable idea of what type of aircraft it is and where it is probably going.

The second contribution tries to assist in an area that will likely always be an open problem. Due to the infinite number of possibilities in the creation of a deep learning model and the type of data that is being processed, determining minimum data set size guidelines that apply to all variations of deep learning models is not feasible. The research discussed in chapter III provides a baseline for this type of deep learning problem. While a full 24-hours of data provided the best accuracy, if collection time and data processing equipment are limited, sufficient accuracy can be achieved with just a quarter of the full amount, or approximately 6 hours of data.

In chapter III, we found a baseline list of features that can be used to classify an aircraft's engine type during takeoff. The best performing model achieved an 89.4% accuracy which was a higher accuracy than previous research and used only speed, pressure and vertical speed. The second best performing model was able to perform almost as well with an 88.7% accuracy using only speed and vertical speed. These features were likely the best performing since during the takeoff phase, features such as altitude, location and track vary less between aircraft. The finding allows for the simple prediction of engines with simply two or three features.

Research from chapter III achieved an 89.4% accuracy in classifying engine type during takeoff with its best performing model. This is a slight improvement over the model developed by Basrawi et al which achieved 89.2% accuracy [20]. The slight improvement by itself is not laudable, but the accuracy was achieved with a more streamlined model. The more streamlined model did not use the attention mechanism nor the dual-stage approach. The benefit is that since the dual-stage approach involves the development of two distinct models and costs nearly twice as much time to build, this extra training time and complexity is saved.

The research described in chapter III only used data from the first 10 minutes of flight for each aircraft in order to maintain similarity across observations. The problem with this method is that the first 10 minutes is only the takeoff and climb segment of flight. Since the collection of ADS-B broadcasts requires that the aircraft be within line-of-sight of the receiver, broadcasts can only be collected at airports or from satellites to be used for these models. For the classification of domestic or friendly aircraft, this is not an issue. For the classification of enemy aircraft, we are unlikely to have the broadcasts from enemy airfields available for classification purposes. Adding a flight phase feature using a fuzzy logic algorithm [18], as was done in chapters IV and V, allows for the classification of all flight segments.

In chapter IV, models were created to predict more precise aircraft characteristics. With a high level of accuracy, we classified aircraft description and WTC, as described in section 2.1.2. Designator could not be classified with a reasonable level of accuracy. However, identifying an aircraft's description and WTC helps to understand general characteristics about the aircraft.

Research from V fused other data types with ADS-B kinematic data to improve classification results. Pre-classification fusion methods are required for weather data since weather cannot predict aircraft on its own. Post-classification fusion methods were required for images since images and ADS-B data require different deep learning algorithms. The addition of images increased classification accuracy of designator from 36.2% to 91.5%. Additionally, adding ADS-B data to images improved the image prediction accuracy by 1-2%.

Table 10: Summary of Research Contributions

ID	**Details**	**Ch.**	**RQ Ref**
C-1	Development of various models that provide aircraft classification with kinematic and unattributed data that can be used as a surrogate for other data types	III, IV, V	RQ-1
C-2	Finding a minimum data set size for all types of deep learning classification problems is an open problem. This research provides a baseline for aircraft prediction models.	III	RQ-1
C-3	This research discovered the baseline features to identify an aircraft engine type during takeoff with kinematic data: Speed, barometric pressure and vertical speed	III	RQ-2
C-4	This research reimplemented the DSDEC model, and found that the attention mechanism and the dual-stage approach were unnecessary. This research achieved a slight improvement in accuracy by excluding those methods	III	
C-5	Adding a flight phase feature allowed for the use of the entire flight path and improved classification accuracy	IV	RQ-3
C-6	Using only ADS-B data to train and classify, we were able to more precisely classify aircraft using aircraft description and WTC, as described in section 2.1.2	IV	RQ-4
C-7	Weather improved classification accuracy of WTC. The addition of image data improved classification accuracy for all categories. ADS-B improved the classification accuracy of the images for WTC and description	V	RQ-5
C-8	Weather must be fused with ADS-B during pre-processing and images must be fused during post-processing	V	RQ-6

6.3 Future Work

Suggestions for future post-dissertation implementations are as follows:

1. **Genetic Algorithms:** This research builds off of previous research that used neural networks to classify engine type from the ADS-B data. However, another area that shows promise for this type of data is genetic algorithms. Future research should look at developing and testing a genetic algorithm with the modified ADS-B data set that will be created from this research.

2. **Discontinuous data transmissions:** Using discontinuous data is a difficult problem that will not be touched upon in this research. Instances of aircraft operations will need at least 10 minutes of continuous transmissions in order to be used in the training of the mode. Since ADS-B can lose an aircraft's signal in areas with no ADS-B receivers, future research should look into the ability to modify the algorithm to use discontinuous data.

3. **Social Media fusion:** Social media is a large, but complicated data source. Future research should look at how to process and integrate social media data with ADS-B model.

4. **Data-in-motion:** The method in outlined in this paper will classify ADS-B using data-at-rest. To make this design effective for military decision makers, modifying the model to classify ADS-B with data-in-motion will be necessary to make predictions as events occur.

Bibliography

1. *117th Congress (2021-2022): National Defense Authorization Act for Fiscal Year 2022*. 117th Congress, December 2022.

2. *Department of the Air Force Posture Statement Fiscal Year 2020*.

3. *Department of the Air Force Posture Statement Fiscal Year 2021*.

4. *Department of the Air Force Posture Statement Fiscal Year 2022*.

5. *Department of the Air Force Posture Statement Fiscal Year 2023*.

6. Airlive.net. Alert over 400 flights canceled in the u.s. due to ads-b navigation system issue, Jun 2019.

7. Jan Stibor, Thomas Koerber, Eric Vallauri, and Katia Orteca. Home - ads-b, Jan 2022.

8. United States Department of Transportation. Airspace, Dec 2021.

9. Fjodor Van Veen. Neural network zoo, 2019.

10. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.

11. fdeloche. Long short-term memory.svg, 2017. [Online; accessed 20-May-2022].

12. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

13. Sumit Saha. A comprehensive guide to convolutional neural networks — the eli5 way, Dec 2018.

14. Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

15. Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. Multivariate lstm-fcns for time series classification. *Neural Networks*, 116:237–245, 2019.

16. Khaled Basrawi, Richard Dill, Gilbert Peterson, Brett Borghetti, and Jennifer Lopez. Aircraft identification from ads-b kinematic data. *Journal of DoD Research and Engineering*, 5(2):31–43, 2022.

17. Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew B. Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *CoRR*, abs/1306.5151, 2013.

18. Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Flight extraction and phase identification for large automatic dependent surveillance–broadcast datasets. *Journal of Aerospace Information Systems*, 14(10):566–572, 2017.

19. Honeywell Aerospace. Automatic dependent surveillance-broadcast will transform air traffic control. Technical report, Honeywell International Inc, Phoenix, Arizona, 2018.

20. Khaled Basrawi. *Aircraft Classification from ADS-B Kinematic Data*. PhD thesis, Air Force Institute of Technology, 2021.

21. Kevin Bartley. Big data statistics: How much data is there in the world?, Nov 2021.

22. Live-Counter.com. How big is the internet?

23. Tausifa Jan Saleem and Mohammad Ahsan Chishti. Deep Learning for Internet of Things Data Analytics. *Procedia Computer Science*, 163:381–390, 2019.

24. Milan Kumar Dholey, Ditipriya Sinha, Sankar Mukherjee, Ayan Kumar Das, and Sudip Kumar Sahana. A novel broadcast network design for routing in mobile ad-hoc network. *IEEE Access*, 8:188269–188283, 2020.

25. Vikash Kumar, Sidra Kalam, Ayan Kumar Das, and Ditipriya Sinha. Attack detection scheme using deep learning approach for iot. *Advanced Computing and Systems for Security*, 14:17–30, 2021.

26. Lei Qian and Vadas Gintautas. Fusing sensor data with publicly available information (PAI) for autonomy applications. In *SPIE Defense + Commercial Sensing*, number May 2019, page 22, Baltimore, 2019.

27. Raphael Ginoulhac, Frederic Barbaresco, Jean Yves Schneider, Jean Marie Pannier, and Sebastien Savary. Target classification based on kinematic data from AIS/ADS-B, using statistical features extraction and boosting. *Proceedings International Radar Symposium*, 2019-June(681):1–10, 2019.

28. Lei Qian, Noah Woods, and Mazaya Rahman. (U) identifying ads-b flight patterns, 2019. Paper presented at the 2019 MSS Joint (BAMS and NSSDF) Conference, San Diego, CA, 21–24 October 2019.

29. Paul Kraus, Camilla Mohrdieck, and Friedhelm Schwenker. Ship classification based on trajectory data with machine-learning methods. In Prof. Dr. Hermann Rohling, editor, *Proceedings International Radar Symposium*, volume 2018-June, pages 1–10, Bonn, Germany, Aug 2018. IEEE Computer Society.

30. Maria Andersson and Ronnie Johansson. Multiple sensor fusion for effective abnormal behaviour detection in counter-piracy operations. *2010 International Waterside Security Conference, WSS 2010*, (December), 2010.

31. Satvik G. Kumar, Samantha J. Corrado, Tejas G. Puranik, and Dimitri N. Mavris. Classification and analysis of go-arounds in commercial aviation using ads-b data. *Aerospace*, 8(10), 2021.

32. Mehdi Mohammadi, Ala Al-Fuqaha, Sameh Sorour, and Mohsen Guizani. Deep learning for IoT big data and streaming analytics: A survey. *arXiv*, 20(4):2923–2960, 2017.

33. Barry G Silverman, Gnana Bharathy, and Nathan Weyer. What is a good pattern of life model? guidance for simulations. *SIMULATION*, 95(8):693–706, 2019.

34. Sergei Mikhailov, Alexey Kashevnik, and Alexander Smirnov. Tourist behaviour analysis based on digital pattern of life. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 622–627, 2020.

35. William Grant Hatcher and Wei Yu. A survey of deep learning: Platforms, applications and emerging research trends. *IEEE Access*, 6:24411–24432, 2018.

36. Nicola Forti, Leonardo M. Millefiori, and Paolo Braca. Unsupervised extraction of maritime patterns of life from automatic identification system data. In *OCEANS 2019 - Marseille*, pages 1–5, 2019.

37. Fouad Faraj. *Object Detection and Pattern of Life Analysis from Remotely Piloted Aircraft System Acquired Full Motion Video*. PhD thesis, Queen's University (Canada), 2021.

38. National Archives and Records Administration. Electronic code of federal regulations, Sept 2021.

39. ADS-B Exchange. Home - serving the flight tracking enthusiast - ads-b exchange, Dec 2021.

40. Guided by national defense strategy, defense department increases force lethality, Aug 2020.

41. Ali K Raz and Roberto Sabatini. Information fusion as an autonomy enabler for uav traffic management. In *AIAA Scitech 2021 Forum*, page 0658, 2021.

42. David Cenciotti. Using ads-b data and satellite imagery to "observe" interesting military aircraft movements (and intercepts), Jul 2021.

43. Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Large-scale flight phase identification from ads-b data using machine learning methods. In *7th International Conference on Research in Air Transportation*, pages 1–8, 2016.

44. Qilei Zhang, John H Mott, Mary E Johnson, and John A Springer. Development of a reliable method for general aviation flight phase identification. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

45. Ying Zhao, Richard Wu, Andrew Polk, Matthew Xi, and Tony Kendall. Big data and deep learning models for automatic dependent surveillance broadcast (ads-b).

46. Kaeye Dästner, Susie Brunessaux, Elke Schmid, Bastian von Haßler zu Roseneckh-Köhler, and Felix Opitz. Classification of military aircraft in real-time radar systems based on supervised machine learning with labelled ads-b data. In *2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6. IEEE, 2018.

47. Ardhendu Behera, Zachary Wharton, Pradeep Hewage, and Asish Bera. Context-aware attentional pooling (cap) for fine-grained visual classification. *arXiv preprint arXiv:2101.06635*, 2021.

48. Icao document 8643. Standard, International Civil Aviation Organization, Quebec, Canada, 2016.

49. Neno Ruseno, Chung-Yan Lin, and Shih-Cheng Chang. Uas traffic management communications: The legacy of ads-b, new establishment of remote id, or leverage of ads-b-like systems? *Drones*, 6(3), 2022.

50. A. Filippone, B. Parkes, N. Bojdo, and T. Kelly. Prediction of aircraft engine emissions using ads-b flight data. *The Aeronautical Journal*, 125(1288):988–1012, 2021.

51. Farah Hasin, Tahmina Haque Munia, Nazmun Nahar Zumu, and Kazi Abu Taher. Ads-b based air traffic management system using ethereum blockchain technology. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 346–350, 2021.

52. Nolan Pearce, Kate J. Duncan, and Bryan Jonas. Signal discrimination and exploitation of ads-b transmission. In *SoutheastCon 2021*, pages 1–4, 2021.

53. Sarah Bolton, Richard Dill, Michael R Grimaila, and Douglas Hodson. Ads-b classification using multivariate long short-term memory–fully convolutional networks and data reduction techniques. *The Journal of Supercomputing*, 79(2):2281–2307, 2023.

54. Junzi Sun. The 1090 megahertz riddle: a guide to decoding mode s and ads-b signals. *TU Delft OPEN Publishing*, 2021.

55. International Civil Aviation Organization. Vision and mission, 2022.

56. Icao wake turbulence category.

57. Niklas Donges. A guide to rnn: Understanding recurrent neural networks and lstm networks, Jul 2021.

58. Tiago Miguel. How the lstm improves the rnn, Jan 2021.

59. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

60. Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.

61. Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. LSTM fully convolutional networks for time series classification. *IEEE Access*, 6:1662–1669, 2018.

62. Ahmad Alwosheel, Sander van Cranenburgh, and Caspar G. Chorus. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 28:167–182, 2018.

63. Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv preprint arXiv:1511.06348*, 2015.

64. A.K. Jain and B. Chandrasekaran. 39 dimensionality and sample size considerations in pattern recognition practice. In Elsevier, editor, *Classification Pattern*

Recognition and Reduction of Dimensionality, volume 2 of *Handbook of Statistics*, pages 835–855. Elsevier B.V., North Holland, Amsterdam, 1982.

65. Eric B. Baum and David Haussler. What size net gives valid generalization? In *Advances in Neural Information Processing Systems 1*, page 81–90, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

66. Simon S Haykin. *Neural networks and learning machines.* Prentice Hall,, New York, 2009.

67. Simon Bækkegaard, Jeppe Blixenkrone-Møller, Jakob Juul Larsen, and Lars Jochumsen. Target classification using kinematic data and a recurrent neural network. In *2018 19th International Radar Symposium (IRS)*, pages 1–10. IEEE, 2018.

68. Lars W Jochumsen, Jan Østergaard, Søren H Jensen, Carmine Clemente, and Morten Ø Pedersen. A recursive kinematic random forest and alpha beta filter classifier for 2d radar tracks. *EURASIP Journal on Advances in Signal Processing*, 2016(1):1–12, 2016.

69. Shenghai Rong, Zilei Wang, and Jie Wang. Separated smooth sampling for fine-grained image classification. *Neurocomputing*, 461:350–359, 2021.

70. Yanfeng Wang, Yinan Chen, and Runmin Liu. Aircraft image recognition network based on hybrid attention mechanism. *Computational Intelligence and Neuroscience*, 2022, 2022.

71. Lei Wang, Kai He, Xu Feng, and Xitao Ma. Multilayer feature fusion with parallel convolutional block for fine-grained image classification. *Applied Intelligence*, 52(3):2872–2883, 2022.

72. Dichao Liu, Longjiao Zhao, Yu Wang, and Jien Kato. Learn from each other to classify better: Cross-layer mutual attention learning for fine-grained visual classification. *Pattern Recognition*, 140:109550, 2023.

73. Wilfried Elmenreich. An introduction to sensor fusion. *Vienna University of Technology, Austria*, 502:1–28, 2002.

74. Asad Vakil, Jenny Liu, Peter Zulch, Erik Blasch, Robert Ewing, and Jia Li. A survey of multimodal sensor fusion for passive rf and eo information integration. *IEEE Aerospace and Electronic Systems Magazine*, 36(7):44–61, 2021.

75. M Przybyła-Kasperek and Alicja Wakulicz-Deja. Comparison of fusion methods from the abstract level and the rank level in a dispersed decision-making system. *International Journal of General Systems*, 46(4):386–413, 2017.

76. Emanuela Marasco, Ayman Abaza, and Bojan Cukic. Why rank-level fusion? and what is the impact of image quality?

77. B Habtemariam, Ratnasingham Tharmarasa, Michael McDonald, and Thia Kirubarajan. Measurement level ais/radar fusion. *Signal Processing*, 106:348–357, 2015.

78. Lei Xu, Adam Krzyzak, and Ching Y Suen. Associative switch for combining multiple classifiers. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume 1, pages 43–48. IEEE, 1991.

79. Gianluca Gennarelli, Moeness G Amin, Francesco Soldovieri, and Raffaele Solimene. Passive multiarray image fusion for rf tomography by opportunistic sources. *IEEE Geoscience and Remote Sensing Letters*, 12(3):641–645, 2014.

80. Weijie Zhan, Jianxin Yi, Xianrong Wan, and Yunhua Rao. Track-feature-based target classification in passive radar for low-altitude airspace surveillance. *IEEE Sensors Journal*, 21(8):10017–10028, 2021.

81. Kaeye Dästner, Elke Schmid, Bastian von Haßler zu Roseneckh-Köhler, and Felix Opitz. Analysis of time series of statistical air traffic data. In *2020 21st International Radar Symposium (IRS)*, pages 157–162. IEEE, 2020.

82. Sai Teja Kanneganti, Phillip B Chilson, and Robert Huck. Visualization and prediction of aircraft trajectory using ads-b. In *NAECON 2018-IEEE National Aerospace and Electronics Conference*, pages 529–532. IEEE, 2018.

83. Donald L McCallie. Exploring potential ads-b vulnerabilites in the faa's nextgen air transportation system. Technical report, AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH DEPT OF ELECTRICAL AND ..., 2011.

84. Mohsen Riahi Manesh, Mahdi Saeedi Velashani, Elias Ghribi, and Naima Kaabouch. Performance comparison of machine learning algorithms in detecting jamming attacks on ads-b devices. In *2019 IEEE International Conference on Electro Information Technology (EIT)*, pages 200–206. IEEE, 2019.

85. Steven Laskie. Neural network zoo, 2019.

86. Christian Sebastian Lamprecht. Meteostat Python.

Acronyms

ABMS Advanced Battle Management System. 4

ADS-B Automatic Dependent Surveillance-Broadcast. iv, v, xi, 2, 3, 5, 7, 8, 10, 11, 12, 13, 15, 28, 29, 30, 31, 32, 34, 36, 37, 38, 40, 41, 42, 70, 78, 79, 85, 87, 88, 89, 90, 92, 93

AFIT Air Force Institute of Technology. x, 11, 12, 13

AFRL Air Force Research Laboratory. x, 11, 12, 13

AIS automatic identification system. 10, 30, 37

ANN Artificial Neural Network. 17, 18, 19, 28

ATC Air Traffic Control. 5

BA-CNN Bat Algorithm and Convolutional Neural Network. 33

CMAL-Net Cross-Layer Mutual Attention Learning Network. 33, 78, 89

CNN Convolutional Neural Network. 10, 17, 23, 24, 25, 27, 34

DAF Department of the Air Force. iv, 4, 85

DCGS Distributed Common Ground System. 2

DSDEC Dual-Stage Deep Engine Classifier. 29, 30, 31, 93

EASA European Union Aviation Safety Agency. 13

FCN Fully Convolutional Network. 25, 26, 27

FGVC-Aircraft Fine-Grained Visual Classification of Aircraft. x, 33, 89

FNN Feed-Forward Neural Network. 17, 18, 20, 25

GPU graphics processing unit. 22

ICAO International Civil Aviation Organization. 2, 10, 11, 12, 14, 15, 38

IoT Internet of Things. 1

ISR Intelligence, Surveillance and Reconnaissance. iv, 4, 85

JADC2 Joint All-Domain Command and Control. 4

JSON JavaScript Object Notation. 12

LSTM Long Short-Term Memory. 10, 17, 21, 22, 26, 27

LSTM-FCN Long Short-Term Memory - Fully Convolutional Network. 27

MALSTM-FCN Multivariate Attention Long Short-Term Memory - Fully Convolutional Network. 28, 29, 31

MFF Multilayer Feature Fusion. 33

MLSTM-FCN Multivariate Long Short-Term Memory - Fully Convolutional Network. iv, x, 28, 70, 78, 86

MSL mean sea level. 13

NDAA National Defense Authorization Act. iv, 3

PAI publicly available information. 1

PC personal computer. 1

PCB Parallel Convolutional Block. 33

POL pattern-of-life. iv, 2, 3, 4, 17, 41, 85

Resnet Residual Network. 33

RF Random Forest. 29, 30, 31

RNN Recurrent Neural Network. 10, 17, 20, 22, 23, 30

RQ research question. 90

SSSNet Separated Smooth Sampling Network. 33

SVM Support Vector Machine. 29, 31, 33

URL Uniform Resource Locator. 41

WTC Wake Turbulence Category. v, 5, 6, 15, 17, 38, 70, 78, 79, 85, 87, 88, 89, 90, 92, 93

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
14-09-2023	Doctoral Dissertation	Sept 2020 — Sept 2023

4. TITLE AND SUBTITLE

Pattern-of-Life Modeling with Automatic Dependent Surveillance-Broadcast (ADS-B)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Sarah J. Bolton, Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-DS-23-S-063

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/Sensors Directorate
2241 Avionics Circle, Bldg 600
WPAFB OH 45433-7318
COMM: 937-713-8554
Email: edmund.zelnio@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RYA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This dissertation and research were sponsored by the Air Force Research Laboratory Layered Sensing Exploitation Branch (AFRL/RYA) to investigate the utility of using the data found within aircraft secondary radar to make predictions about aircraft characteristics and intent. The research focuses on making predictions on aircraft characteristics using only the kinetic data within one type of secondary radar, Automatic Dependent Surveillance-Broadcast (ADS-B), as a surrogate for primary radar. The results from this research provide a means to reduce the reliance on a type of aircraft tracking that is vulnerable to cyber attack and other integrity concerns.

15. SUBJECT TERMS

ADS-B, Deep Learning, publicly available information, classification

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	120	Major Richard Dill, AFIT/ENG (937) 255-3636, ext 3652; richard.dill@afit.edu