

# Deep Learning Based Anomaly Detection for Securing ADS-B in NextGen Aviation

Jayesh Dinesh Kenaudekar  
Linköping University - LiU  
Linköping, Sweden  
jayesh@pryvx.com

Suleman Khan  
Linköping University - LiU  
Linköping, Sweden  
suleman.khan@liu.se,

Flavio Luiz dos Santos de Souza  
Federal Institute of Sao Paulo - IFSP  
Catanduva, Brazil  
flavio.souza@ifsp.edu.br

Andrei Gurtov  
Linköping University - LiU  
Linköping, Sweden  
andrei.gurtov@liu.se

**Abstract**—The Next Generation Air Transportation System relies extensively on the Automatic Dependent Surveillance–Broadcast (ADS-B) protocol to support ground-based air traffic surveillance and enhance air traffic operations. Although ADS-B provides precise, real-time aircraft position reports, its plain-text broadcast design leaves ground receivers vulnerable to message injection, modification, replay, jamming, and spoofing attacks. These vulnerabilities can degrade the integrity of the surveillance picture used by air traffic controllers, creating potential safety risks. To address these challenges without modifying the ADS-B protocol or requiring additional avionics changes, we propose a multilayer anomaly-detection framework designed explicitly for ground-based ADS-B surveillance systems. The first stage uses a Graph Convolutional Network to detect spatial inconsistencies in aircraft behavior across the monitored airspace. The second stage employs a WaveNet-based temporal model to identify anomalous deviations in kinematic features such as speed, altitude, and heading. The final stage integrates an RF fingerprinting classifier that analyzes raw IQ samples to distinguish legitimate aircraft from spoofed transmitters. Experimental results show that the framework achieves 99.99% accuracy on decoded ADS-B test data, 99.25% accuracy on raw IQ samples, and 87.2% accuracy in aircraft identification, with a False Alarm Rate as low as 0.25%. These findings demonstrate the effectiveness of the proposed approach for strengthening ground-based ADS-B surveillance and improving the reliability of air traffic management systems.

per year by the early 2030s [2]. In addition to this passenger growth, increased air cargo operations, military flights, and the use of unmanned aerial vehicles will further contribute to congestion in the skies. To ensure safe navigation and reduce the cost of air traffic control, the aviation community has transitioned from independent surveillance systems, such as primary and secondary surveillance radar (PSR/SSR), to cooperative and dependent surveillance (CDS) technologies, most prominently ADS-B [3].

ADS-B, endorsed by both the International Civil Aviation Organization (ICAO) and the FAA [4, 5], represents a modern evolution of SSR. Aircraft equipped with ADS-B transponders determine their location using satellite navigation and broadcast information such as identity, position, altitude, and velocity at an average rate of 4.2 messages per second. This continuous and precise reporting enables real-time tracking of aircraft, improving traffic management efficiency and flight safety in congested airspace. However, ADS-B was developed with an emphasis on cost-effectiveness and positional accuracy rather than security. As a result, the system remains vulnerable to multiple threats: messages are broadcast in plaintext without encryption or authentication, making them susceptible to eavesdropping, replay, modification, and injection; furthermore, the absence of entity authentication makes it difficult to distinguish legitimate transmitters from malicious ones. Previous studies [6–9] have shown that attackers equipped with inexpensive software-defined radio (SDR) devices can intercept, alter, and inject ADS-B signals, exposing billions of passengers to potential risks.

To mitigate these vulnerabilities, researchers have proposed several cryptography-based approaches. These include encryption and digital signatures to validate broadcast messages [10], challenge–response protocols for aircraft authentication [11], and message authentication codes for integrity assurance [7, 12]. While promising in theory, such methods require modifications to the ADS-B message structure or additional message types for key distribution, which is impractical given the FAA’s 2020 mandate for ADS-B equipage and the stringent certification requirements of avionics systems. Similarly, deploying supplementary sensors or infrastructure would involve high costs and complex regulatory approval processes.

In response to these challenges, recent work has shifted toward data-driven anomaly detection as a non-intrusive alternative. Darabseh [13] and Melo et al. [14] proposed verifying ADS-B messages using minimal additional sensors compared to multilateration methods. Habler and Shabtai [15] introduced a machine-learning–based solution using

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. ADS-B WORKING MECHANISM .....	2
3. RELATED WORK .....	3
4. PROPOSED METHODOLOGY .....	3
5. RESULTS AND DISCUSSION.....	6
6. CONCLUSION AND FUTURE WORK .....	12
ACKNOWLEDGEMENTS.....	12
REFERENCES .....	12
BIOGRAPHY .....	13

## 1. INTRODUCTION

The global demand for air travel has risen steadily over the past decades. According to the Federal Aviation Administration (FAA), by 2033 commercial aviation is expected to serve nearly 1.15 billion passengers annually [1]. In comparison, Eurocontrol forecasts 1.6 billion passengers

Long Short-Term Memory (LSTM) encoder–decoder models to capture flight-route patterns and detect deviations. Li et al. [16] applied Hierarchical Temporal Memory (HTM) to identify anomalies in flight paths. Although effective in specific scenarios, these approaches generally analyze data from individual aircraft in isolation, overlooking spatial and temporal dependencies between multiple aircraft operating within the same airspace. Moreover, most of these methods primarily focus on injection and modification attacks, often with relatively high false positive rates, which limits their operational deployment in real-world ATC systems.

In this work, we present a novel deep learning–based framework for anomaly detection in ADS-B communication that requires neither protocol modifications nor additional hardware. The framework is designed to detect injected, replayed, and spoofed messages originating from malicious actors and compromised aircraft. Our contributions are summarized as follows:

We design a graph-based detection module that models the airspace as a dynamic graph, enabling spatial correlation analysis among aircraft to identify abnormal message patterns.

We develop a sequence-level anomaly detector that leverages deep temporal models to forecast normal flight behavior and detect deviations in kinematic features such as speed, altitude, and heading.

We integrate an RF-fingerprinting classifier that extracts unique signal characteristics from raw IQ data to distinguish legitimate aircraft from impersonators, strengthening defenses against spoofing attacks.

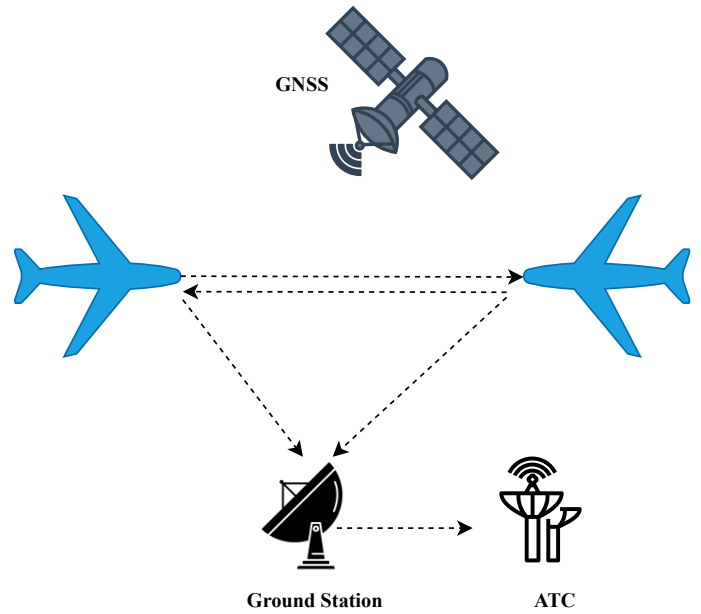
The paper is organized as follows. Section II introduces the working mechanism of ADS-B. Section III discusses related work in ADS-B security. Section IV presents the proposed methodology, including the multi-stage anomaly detection framework. Section V provides the experimental setup, results, and discussion. Finally, Section VI concludes the paper by summarizing the findings and outlining future research directions.

## 2. ADS-B WORKING MECHANISM

This section provides an overview of the relevant background information on the ADS-B protocol.

ADS-B is a modern surveillance system that relies on the Global Navigation Satellite System (GNSS) to automatically and continuously determine an aircraft’s position. This positional information is then broadcast to nearby aircraft and ground stations. The primary objective of ADS-B is to enhance aviation safety by reducing the risk of collisions and improving the accuracy and coverage of air traffic monitoring, even in challenging environments such as mountainous regions and oceanic airspace [4, 5]. In comparison with traditional PSR and SSR, ADS-B provides a more cost-effective and accurate alternative for ATM [3].

ADS-B consists of two complementary subsystems: ADS-B Out and ADS-B In. The ADS-B Out subsystem enables an aircraft to continuously broadcast unencrypted messages containing essential flight parameters such as identity, position, velocity, and altitude. The ADS-B In subsystem allows aircraft and ATC stations to receive these messages, thereby enhancing cooperative surveillance and situational awareness, as shown in Fig. 1 [1].



**Figure 1:** An illustration of how the ADS-B system works. The GNSS determines the aircraft’s position, which is processed onboard and broadcast by the ADS-B Out subsystem. The ADS-B In subsystem allows ground stations and nearby aircraft to receive these messages, which are then forwarded to ATC.

Globally, the most widely used ADS-B data link operates at 1090 MHz and uses Mode S surveillance technology. This employs an Extended Squitter (ES) message format with a total length of 112 bits and a transmission duration of 120 microseconds. ES messages are broadcast at least once per second, and more frequently depending on aircraft dynamics. Each ES message is divided into five fields: Downlink Format, Capability, Aircraft Address, ADS-B Data, and Parity Check [8]. These fields are summarized in Table 1.

In addition to the 1090 MHz Mode S link, the United States also uses the 978 MHz Universal Access Transceiver (UAT) link, primarily for general aviation aircraft flying below 18,000 feet. The UAT link supports larger message payloads and the broadcast of supplementary information such as weather and aeronautical data services. Still, it is not widely deployed outside the U.S. airspace [1].

**Table 1:** Format of a 1090 Extended Squitter Message.

Segment	Downlink Format	Capability	Aircraft Address	ADS-B Data	Parity Check
Bits	5	3	24	56	24

The ADS-B Data field typically contains the following aircraft attributes:

**ICAO Address:** A unique 24-bit identifier assigned to the aircraft transponder.

**Time:** Timestamp recorded in epoch milliseconds.

**Longitude:** East–west coordinate in degrees, within the range  $-180$  to  $+180$ .

**Latitude:** North–south coordinate in degrees, within the range  $-90$  to  $+90$ .

**Airborne Velocity:** Horizontal and vertical components

of velocity, measured in knots (nautical miles per hour), typically ranging from 0 to 500.

**Altitude:** Height above mean sea level, measured in feet, typically ranging from 0 to 40,000.

**Heading:** Direction of travel, measured in degrees from 0 to 360.

Through this design, ADS-B enables continuous, high-fidelity surveillance of aircraft movements, thereby complementing and replacing conventional radar-based systems in both domestic and international airspace.

### 3. RELATED WORK

The ADS-B system plays an important role in modern air traffic surveillance; however, its lack of built-in security makes it vulnerable to a wide range of cyberattacks. Researchers [7, 17–21] have extensively investigated these weaknesses and proposed countermeasures that generally fall into two categories: cryptography-based methods and machine-learning-based approaches.

Strohmeier et al. [17] examined ADS-B as part of next-generation ATM. While they emphasized its importance in handling the growing volume of flights and improving passenger safety, they also identified a fundamental weakness: ADS-B messages are broadcast in plain text, without encryption and authentication. This limitation has motivated various proposals based on cryptography. For instance, Finke et al. [18] introduced encryption schemes. At the same time, Costin et al. [7] and Feng et al. [18] suggested the adoption of Public Key Infrastructure (PKI) with digital signatures and elliptic-curve cryptography. Similarly, Kacem et al. [18–20] proposed symmetric-key and elliptic curve-based mechanisms. Although such approaches can enhance authentication and integrity, their integration requires modifications to the ADS-B protocol. Given ADS-B’s small payload size and strict avionics standards, these changes introduce significant communication overhead and make practical deployment challenging.

To address these limitations, researchers have increasingly turned to data-driven methods. Wang et al. [21] presented a spoofing detection technique using a LSTM network, capable of detecting ten different types of manipulated ADS-B messages without manual feature engineering and protocol modifications. However, their method was restricted to a narrow set of attack scenarios and suffered from a relatively high FAR. In a similar direction, Fried et al. [22] developed an LSTM encoder-decoder that reconstructs ADS-B messages and flags anomalies when reconstruction errors are high. While effective in identifying abnormal traffic, their approach often misclassifies new but legitimate traffic as suspicious, leading to operational concerns.

Beyond message-level analysis, other studies have explored the physical layer. Ying et al. [23], and Leonardi et al. [24] investigated Deep Neural Network (DNN) based classifiers applied to raw In-phase and Quadrature (IQ) samples, showing that phase and waveform characteristics can be used to distinguish genuine aircraft from spoofed transmitters. These approaches provide an additional layer of protection against impersonation attacks by exploiting PHY-layer radio signatures. However, such methods operate solely on individual signals and do not incorporate spatial relationships between aircraft or the temporal evolution of their kinematic behavior. In contrast, our framework integrates spatial

graph modeling, temporal anomaly detection, and RF fingerprinting, allowing the detection of both signal-level spoofing and message-level anomalies across the monitored airspace.

In summary, cryptography-based solutions offer strong theoretical protection but face deployment barriers due to protocol constraints and overhead. At the same time, machine learning-based methods avoid protocol modifications but often suffer from limited threat coverage and high FARs. This trade-off highlights a critical research gap: the need for solutions that are both practical and robust, ensuring strong security for ADS-B communications without requiring disruptive changes to the existing protocol.

### 4. PROPOSED METHODOLOGY

The architecture of the proposed framework, shown in Fig. 2, consists of three main stages: an attack classifier, an ADS-B feature analyzer, and an aircraft classifier. In the first stage, a GCN is employed to classify ADS-B messages as either normal or an attack. This enables early detection of malicious traffic patterns. If a message is classified as authentic, it proceeds to the second stage, where a WaveNet model is applied to perform time-series analysis of flight parameters such as speed, altitude, and heading. This stage detects anomalies by evaluating whether the temporal evolution of these features is consistent with expected aircraft behavior. Messages with no detected anomalies then move to the third stage, where a DNN is used to classify the aircraft based on ADS-B features. Any inconsistency at this stage, such as a mismatched aircraft identity or behavior, is flagged as anomalous. During runtime, any instance identified as an attack or anomaly at any stage immediately triggers a Stop/Alert action, while only traffic classified as authentic across all three stages is accepted. This multistage design ensures that message-level, feature-level, and aircraft-level aspects of ADS-B communication are jointly analyzed, providing robust detection of malicious or anomalous behavior.

#### *Data Collection*

To develop a robust ADS-B attack detection system, we leverage two forms of ADS-B data.

**1) IQ Samples.** The first dataset consists of quadrature signals, also referred to as I/Q data or IQ samples, extracted from the received aircraft signals during RF modulation. These signals preserve intrinsic transmitter-specific properties, allowing us to distinguish the signals of one aircraft from another. The dataset contains authentic ADS-B messages for a range of aircraft, recorded using an ADS-B antenna by authors [23]. In addition to authentic traffic, the dataset has been augmented with hardware-level attack scenarios, including rebroadcast, simulated, and relay attacks.

**2) Decoded ADS-B Messages.** The second dataset consists of decoded ADS-B messages, which provide structured kinematic features such as latitude, longitude, altitude, speed, and heading direction. These features are used to perform deep learning-based feature analysis for detecting deviations and anomalies. This dataset was open-sourced on GitHub by Jing Wang [21] and is publicly available for research use.

Figure 2: Proposed Framework: Multistage Intrusion Detection System for ADS-B

Table 2: Summary of ADS-B Datasets Utilized.

RF IQ Samples	Decoded ADS-B Messages
Quadrature (I/Q) data extracted from RF-modulated aircraft signals, capturing hardware-specific transmitter characteristics. 30,000 samples including authentic transmissions and attack variants (replay, rebroadcast, relay, ghost-aircraft).	Structured ADS-B messages containing kinematic features such as latitude, longitude, altitude, speed, and heading. 10,000 real samples per aircraft across 20 ICAO hardware-level addresses.

### Stage 1: GCN for ADS-B Attack Detection

The first stage of the framework is a graph-based message classifier, which performs binary classification to determine whether incoming ADS-B traffic is normal or malicious. This approach leverages the spatial relationships between aircraft by representing the airspace as a graph. Each unique aircraft, identified by its ICAO address, is described as a node, while edges are defined by the distance between aircraft within the radar coverage range, typically 370 km. For illustration, five aircraft are selected as nodes; however, in practice the number of nodes depends on the aircraft present in the monitored region.

The GCN architecture is employed to build the message classification model on graph-structured ADS-B data. To prepare the inputs, raw ADS-B messages from multiple aircraft are transformed into graphs that capture their relationships between aircraft within the airspace at a given time. Two datasets are used: one containing authentic ADS-B messages from real aircraft, and another containing adversarial messages. Rather than manually crafting attacks, a Generative Adversarial Network (GAN) is used to generate realistic fake ADS-B messages from decoded flight data.

### Figure 3: Airspace as Graph representation

Incorporating GAN-generated adversarial samples enhances the classifier's performance by exposing it to realistic yet manipulated traffic, thereby increasing its robustness and ability to detect previously unseen attack patterns. After preparing both authentic and adversarial datasets, each batch of ADS-B messages is converted into a graph. In this representation, nodes correspond to aircraft (identified by their ICAO addresses), and edges represent the distances between them. Graphs are labeled based on their composition: if all messages are authentic, the graph is classified as authentic; if one or more adversarial messages are present, the graph is classified as an attack. A new graph is created continuously as new ADS-B messages are received, ensuring real-time classification of the airspace.

The GCN model comprises three graph convolutional layers, followed by a global mean pooling operation. The convolutional layers enable nodes to aggregate and propagate information across their neighbors. With three layers, each

**Algorithm 1 GCN for ADS-B Message Classification (Stage 1; M1, M2, M3)**

- 1: Input: Node features  $X \in \mathbb{R}^{N \times 6}$  (lat, lon, alt, speed, heading, climb)
- 2: Output:  $O \in \mathbb{R}^2$  fNormal; Attackg
- 3: Step 1: Preprocessing
- 4: Sort messages by ICAO and timestamp
- 5: Normalize features:  $\hat{X} = (X - X_{\min}) / (X_{\max} - X_{\min})$
- 6:  $H^{\text{in}} \in \mathbb{R}^{N \times 6}$
- 7: Step 2: Graph Construction
- 8: Build graph  $G = (V; E)$  with self-loops
- 9:  $V$  unique ICAO aircraft
- 10: Node features  $H^{\text{in}}$
- 11:  $E$  distance-based proximity
- 12: Graph label  $y \in \mathbb{R}^2$  fNormal; Attackg
- 13: Step 3: GCN Encoder (variant-dependent)
- 14: if M1 then
- 15:  $H^{(1)} = \text{ReLU}(\text{GCNConv}(6; 128; G; H^{\text{in}}))$
- 16:  $H^{(2)} = H^{(1)}$
- 17: else if M2 then
- 18:  $H^{(1)} = \text{ReLU}(\text{GCNConv}(6; 64; G; H^{\text{in}}))$
- 19:  $H^{(2)} = \text{ReLU}(\text{GCNConv}(64; 64; G; H^{(1)}))$
- 20:  $H^{(3)} = H^{(2)}$
- 21: else if M3 then
- 22:  $H^{(1)} = \text{ReLU}(\text{GCNConv}(6; 64; G; H^{\text{in}}))$
- 23:  $H^{(2)} = \text{ReLU}(\text{GCNConv}(64; 64; G; H^{(1)}))$
- 24:  $H^{(3)} = \text{ReLU}(\text{GCNConv}(64; 64; G; H^{(2)}))$
- 25:  $H^{(4)} = H^{(3)}$
- 26: end if
- 27:  $z = \text{GlobalMeanPool}(H^{(4)})$
- 28:  $\hat{y} = \text{Softmax}(\text{Linear}(z; H ! 2))$
- 29: Step 4: Training
- 30: for epoch = 1 to E do
- 31: for each graph G with label y do
- 32: Forward pass as above
- 33:  $L = \text{CrossEntropy}(\hat{y}; y)$
- 34: Update parameters using backpropagation on L
- 35: end for
- 36: end for
- 37: Step 5: Inference
- 38: Build  $G_{\text{new}}$  and  $H_{\text{new}}^{\text{in}}$
- 39: Compute  $H^{\text{in}}$ , then z, then  $\hat{y}$  as above
- 40: return  $O = \text{arg max}(\hat{y})$

is its ability to capture long-term temporal dependencies efficiently, without requiring excessively large models.

At the core of WaveNet are dilated causal convolutional layers. Unlike conventional 1-D convolutions that may allow information from the future to influence the past, causal convolutions ensure that the prediction at time step  $t$  depends only on inputs from time steps  $\leq t$ . Dilated convolutions extend this by introducing gaps between filter elements, which enlarges the receptive field exponentially with depth. A dilated causal convolution with kernel size  $K$  and dilation  $d$  is defined as:

$$y[t] = \sum_{k=0}^{K-1} w[k] x[t - d \cdot k]; \quad (1)$$

where  $w[k]$  are the learnable kernel weights, and  $d$  controls the spacing of past inputs considered. For example, using dilation factors of 1; 2; 4; 8, just four layers can cover 16 past time steps. This allows WaveNet to handle long ADS-B sequences without requiring a prohibitively deep network.

To model complex temporal patterns, WaveNet employs gated activation units. These are defined as:

$$Z = \tanh(W_{f;d} X) \odot (W_{g;d} X); \quad (2)$$

where  $\odot$  denotes a dilated causal convolution with dilation  $d$ ,  $\odot$  denotes element-wise multiplication,  $\tanh$  and  $\odot$  are activation functions, and  $W_f, W_g$  are learnable filter and gate weights, respectively. This combination introduces non-linearity, allowing the network to capture temporal correlations in ADS-B sequences more effectively.

WaveNet also incorporates residual and skip connections. Each residual block produces a skip output that is summed across all blocks and a residual output that is added back to the input of the block. These additions mitigate the vanishing gradient problem, accelerate convergence, and make it possible to train deeper models effectively.

For ADS-B forecasting, the output is continuous rather than categorical. Therefore, the final softmax layer used in the original WaveNet is omitted, and the network directly predicts future values of kinematic attributes (e.g., speed, altitude, heading).

**Aircraft Classification via RF Fingerprinting**

node can integrate information from up to three hops away, capturing both local and broader airspace interactions. Each hidden layer uses a 128-dimensional feature representation. The pooled representation is then passed to a classifier that produces the final binary output, identifying whether the observed airspace state is normal or under attack.

**Building WaveNet Model for ADS-B Data Forecasting**

WaveNet is a deep autoregressive generative model introduced by DeepMind for the initial generation of raw audio [25]. Although initially designed for speech, its fully convolutional architecture makes it well-suited for time-series forecasting tasks, such as predicting future ADS-B message attributes. The key advantage of WaveNet

To ensure robustness against spoofed identifiers, the aircraft identification module is designed from the outset to avoid any dependence on message-level fields such as the ICAO address. Instead of exploiting raw IQ magnitudes and other features tied to the broadcast content, it leverages RF-layer signatures that remain invariant to the transmitted data. In particular, phase-based features extracted from the I/Q stream capture hardware-induced imperfections such as oscillator frequency offsets and Doppler shifts that act as unique fingerprints and are extremely difficult to forge.

For the  $k^{\text{th}}$  I/Q sample pair, the instantaneous phase is defined as [23]

$$\phi[k] = \text{atan2}(Q[k]; I[k]); \quad (3)$$

Algorithm 2 WaveNet Model for ADS-B Time-Series Forecasting (K=2, Dilations f1; 2; 4; 8g)

```

Require: Input  $X \in \mathbb{R}^{T \times F}$  in (causal order), all convs use
causal padding
Ensure: Output  $\hat{Y} \in \mathbb{R}^{T \times F}$  out (reshape if multi-step)
Hyperparameters: residual=32, skip=64, head=128
1: procedure WAVENET( $X$ )
2:    $H \leftarrow \text{Conv1D}_{\text{causal}}(X; \text{Iters} = 32; \text{kernel} = 1)$ 
3:    $S_{\text{sum}} \leftarrow 0$ 
4:   for  $d \in \{2, 4, 8\}$  do
5:      $F \leftarrow \text{Conv1D}_{\text{causal}}(H; \text{Iters} = 32; \text{kernel} = 2; \text{dilation} = d)$ 
6:      $G \leftarrow \text{Conv1D}_{\text{causal}}(H; \text{Iters} = 32; \text{kernel} = 2; \text{dilation} = d)$ 
7:      $Z \leftarrow \tanh(F) \cdot G$ 
8:      $S_{\text{blk}} \leftarrow \text{Conv1D}(Z; \text{Iters} = 64; \text{kernel} = 1)$ 
9:      $R_{\text{blk}} \leftarrow \text{Conv1D}(Z; \text{Iters} = 32; \text{kernel} = 1)$ 
10:     $S_{\text{sum}} \leftarrow S_{\text{sum}} + S_{\text{blk}}$ 
11:     $H \leftarrow H + R_{\text{blk}}$ 
12:  end for
13:   $Y \leftarrow \text{ReLU}(S_{\text{sum}})$ 
14:   $Y \leftarrow \text{Conv1D}(Y; \text{Iters} = 128; \text{kernel} = 1)$ 
15:   $Y \leftarrow \text{ReLU}(Y)$ 
16:   $\hat{Y} \leftarrow \text{Conv1D}(Y; \text{Iters} = F_{\text{out}}; \text{kernel} = 1)$ 
17:  return  $\hat{Y}$ 
18: end procedure

```

where  $I[k]$  and  $Q[k]$  denote the in-phase and quadrature components, respectively. The use of  $\text{atan2}$  ensures robust angle estimation across all four quadrants.

The complex baseband signal can be expressed as

$$x(t) = I(t) + jQ(t); \quad (4)$$

For ADS-B, the passband signal at carrier frequency  $f_c$  is modeled as

$$x_p(t) = \left\langle \frac{n_p}{2} x(t) e^{j2\pi f_c t} \right\rangle; \quad (5)$$

In practice, the transmitter (TX) and receiver (RX) local oscillators are not perfectly aligned. Let  $f_{TX}$  and  $f_{RX}$  denote the actual TX and RX LO frequencies near  $f_c$ , and define the LO mismatch

$$f_{LO} = f_{TX} - f_{RX}; \quad (6)$$

Let  $f_D$  denote the Doppler shift due to relative motion. The effective carrier-frequency offset observed at baseband is then

$$f = f_{LO} + f_D; \quad (7)$$

and  $\phi$  captures the residual constant phase (initial phase and static LO phase error).

With these impairments, the received passband signal becomes

$$x_p(t) = \left\langle \frac{n_p}{2} x(t) e^{j2\pi(f_c + f)t + j\phi} \right\rangle; \quad (8)$$

Equivalently, the phase evolution can be written as

$$\phi(t) = 2\pi f t + \phi; \quad (9)$$

so that the instantaneous frequency is

$$f(t) = 2f; \quad (10)$$

The derivative  $\dot{\phi}(t)$  therefore reveals the effective carrier-frequency offset, which reflects both TX/RX oscillator mismatches (via  $f_{LO}$ ) and Doppler shifts ( $f_D$ ), while also being influenced by the propagation channel. These phase dynamics, together with I/Q imbalance patterns, form a discriminative feature set that a classifier can use to infer the transmitter's true identity [23].

## 5. RESULTS AND DISCUSSION

**Hardware setup:** All experiments were carried out on Google Colaboratory using its GPU runtime. Each session provided access to an NVIDIA Tesla T4 GPU with about 15 GB of memory and a 6-hour time limit, which was enough to run several training and testing cycles each day. To maximize the effectiveness of these resources, we designed the model architecture to be lightweight and employed efficient training methods, thereby reducing both computation time and energy consumption.

**Evaluation Metrics.** Since the attack classifier functions as a detection system, we evaluate its performance using the following two standard metrics:

1. Detection Probability ( $P_d$ ): the proportion of malicious messages correctly identified as malicious.
2. False Alarm Probability ( $P_{fa}$ ): the proportion of authentic messages incorrectly classified as malicious.

**Graph data transformation** using decoded real ADS-B messages

For each time slice  $t$ , we construct a graph  $G(V_t; E_t)$ , where nodes  $V$  represent the aircraft (identified by ICAO) and edges  $E$  encode pairwise proximity. Each node carries a six-dimensional feature vector: latitude; longitude; speed; altitude; climb rate; heading while each edge stores the inter-aircraft distance (optionally normalized). We first build graphs from authentic decoded ADS-B messages, then augment the dataset with adversarial graphs generated per aircraft using a GAN trained on decoded traffic. To emulate realistic conditions, we also form mixed graphs by interleaving authentic and simulated messages. The final corpus is split randomly into 80% training and 20% testing.

**GCN model variants:** We evaluate three configurations, where  $L$  denotes the number of graph-convolution layers (each followed by ReLU), and a global mean pooling + linear classifier completes the model:

1. M1:  $L=1$  (128 hidden units).
2. M2:  $L=2$  (64 hidden units per layer).
3. M3:  $L=3$  (64 hidden units per layer).

With  $L$  layers, each node aggregates information up to its  $L$ -hop neighborhood, enabling the model to capture both local and broader airspace interactions.

As depicted in Table 3, the one-layer GCN model (M1) delivers robust baseline performance, achieving an accuracy of 99.12%. Increasing the depth to two layers (M2) does not result in a significant improvement, as accuracy remains

(a) GCN 1 Training and Validation Loss (b) GCN 2 Training and Validation Loss (c) GCN 3 Training and Validation Loss

Figure 4: M1, M2 and M3 Model Training and Validation Loss on ADS-B Decoded Messages

(a) GCN 1 Training and Validation Loss (b) GCN 2 Training and Validation Loss (c) GCN 3 Training and Validation Loss

Figure 5: M1, M2 and M3 Model Training and Validation Loss on ADS-B IQ Data

Table 3: GCN (Multi-hop layers) Training Loss and Detection Accuracy on Decoded ADS-B Messages

	1-layer	2-layer	3-layer
Avg. Loss	0.094	0.071	0.055
Accuracy (%)	99.12	98.01	99.45

nearly unchanged. The best overall result is obtained with the three-layer model (M3), which reports an accuracy of 99.45%. However, adding further multi-hop layers beyond this point does not enhance performance. This observation is consistent with the phenomenon of over-smoothing in graph neural networks. As the number of layers increases, node embeddings become progressively similar due to repeated message passing from each node to its neighbors and outward across the graph. While such aggregation captures broader structural context, excessive propagation diminishes discriminative power, leading to a saturation or even degradation in classification accuracy.

Table 4: Precision, Recall, F1-score, training and testing time for decoded ADS-B messages

Class	Precision (%)	Recall (%)	F1 Score (%)	Training Time	Testing Time
Attack	100	99.51	99.75	3.87	0.033
Normal	99.48	100	99.73		

The training and validation loss curves for GCN-1, GCN-2, and GCN-3, using both decoded ADS-B and IQ samples, are displayed in Fig. 4 and Fig.5, respectively. Similarly, Fig. 6 and Fig.7 illustrate the training and validation accuracy. Additionally, the classification summary for the decoded ADS-B data is presented in Table 4.

Bayesian Parameter Search— Estimating uncertainty is crucial in aviation anomaly detection, as incorrect or overconfident predictions can lead to false alarms and missed detections. False alarms may overload air traffic controllers with unnecessary warnings, while missed detections could allow malicious activity and abnormal aircraft behavior to go unnoticed. Both scenarios can cause significant risks to operational safety and decision-making. Therefore, in this work, we focus on epistemic uncertainty, which reflects the model's confidence in its own predictions. Although Bayesian Neural Networks (BNNs) can capture this form of uncertainty, they are computationally expensive and unsuitable for real-time aviation systems. To overcome this, we apply dropout as a Bayesian approximation. This method is computationally lightweight while still providing a meaningful estimate of prediction reliability. We conducted a Bayesian parameter search by testing dropout values between 0.2 and 0.8, aiming to find the best trade-off between accuracy and uncertainty calibration. The most favorable result was obtained at a dropout rate of 0.33, where the model achieved its lowest loss of 0.3081 while maintaining stable and accurate predictions. These results show that robust uncertainty estimation can be integrated into the ADS-B anomaly detection framework while remaining computationally efficient. This makes the system both reliable and practical, qualities that are indispensable for aviation safety and security.

GCN on Raw IQ Samples—We trained a GCN model directly on raw IQ samples, achieving a mean loss of 0.05% and a test accuracy of 99.25%. These results are consistent with those obtained from a GCN trained on decoded ADS-B messages, indicating that the same architecture can be effectively applied to alternative input modalities without requiring modifications and compromising performance. Each IQ sample comprises 480 features extracted from the

in-phase (I) and quadrature (Q) components of the signal values, and the MAE is calculated as the average difference providing a high-dimensional representation suitable for across all time steps. The objective is for the predictions classification. Table 5 presents the average training loss so closely follow the actual values over time. For example, and detection accuracy for different GCN configurations using the latitude feature, the model achieves an MAE of 0.5433, as reported in Table 8, indicating that the model deeper architectures demonstrate substantial improvements. While a single-layer model achieves slightly lower accuracy, effectively captures the underlying sequence dynamics. We with two- and three-layer models surpassing 99% accuracy also extend this evaluation to other features such as rate of This finding highlights the significance of multi-hop feature climb and speed. As illustrated in Fig. 9 and Fig. 10, the aggregation in extracting discriminative patterns from raw predicted curves for RoC and speed follow the general trend IQ data. Table 6 provides the training and inference times of the actual values, indicating that the WaveNet model can of the GCN on IQ samples, together with precision, recall, approximate real aircraft behavior with reasonable accuracy and F1-score. The learning rate and number of epochs were across multiple features.

kept consistent with the experiments on decoded ADS-B data. However, training time is significantly higher for IQ samples Threshold Setting—After training the model, the detection due to their larger feature dimensionality of 480, compared to threshold is determined using the test dataset. Given a set of predictions  $P$  and the corresponding ground-truth values  $V$ , the residual An error is defined as:

Table 5: Graph Convolution Network (Multi-hop layers) Training Loss and Detection Accuracy on Decoded IQ Samples

	1-layer	2-layer	3-layer
Avg. Loss	0.64	0.05	0.05
Accuracy(%)	63.58	99.16	99.25

$$D = |P - V| \quad (11)$$

We then compute the mean and the standard deviation of the residuals. The threshold is set as:

Table 6: Precision, Recall, F1-score, training and testing time for IQ samples

Class	Precision (%)	Recall (%)	F1 Score (%)	Training Time (Seconds)	Testing Time (Seconds)
Attack	99.65	99.16	99.41	86.44	0.014
Normal	98.59	99.42	99.01		

$$T = 3 \quad (12)$$

Any instance where the residual error exceeds this threshold is classified as an anomaly. This approach ensures that the model not only fits the training data but can also generalize to unseen data by identifying deviations that fall outside the normal distribution of prediction errors.

### Anomaly prediction with Wavenet

To enable anomaly detection with the WaveNet model, ADS-B messages are first organized into time-series form. For each aircraft, messages are sorted by timestamp based on their unique ICAO address, and all six features of the ADS-B protocol are included. A min-max normalization is then applied to scale the features uniformly. The input is processed using a sliding window of length 10, corresponding to roughly 5 seconds of ADS-B data, with the task of predicting the next value (the 11th sequence), as shown in Fig. 8.

It is important to note that anomaly thresholds often differ across individual features. Analyzing thresholds at the feature level provides finer granularity in anomaly detection. However, to ensure consistency and simplify the detection process, we aggregate the thresholds from all features and compute their average. This results in a single global threshold, which in our experiments was estimated as  $T = 0.015$ .

Test for Simulated Attacks: To evaluate the robustness of the trained WaveNet model, we conducted several software-simulated attack scenarios, including random noise, jamming, injection, and field modification, as summarized in Table 7. The objective is to test whether the model can correctly identify these anomalies. For example, when the speed value in an ADS-B message is deliberately altered before reaching the ground station, the model, having learned the normal temporal behavior of aircraft, should detect a deviation from expected patterns. In such cases, the residual error between the predicted and manipulated values increases beyond the defined threshold, allowing the event to be flagged as an anomaly.

Abnormal Score Visualization: Fig. 11 and Fig. 12 show anomaly scores for selected features, including rate of climb, speed, altitude, and latitude, compared with their normal baselines. The highlighted spikes (marked with red circles) represent points where the system detected anomalies. Each spike corresponds to an injected and modified attack, and when the anomaly score crosses the threshold of 0.015, the system issues an alert to indicate suspicious activity. This visualization confirms the model's ability to distinguish normal flight behavior from maliciously altered data.

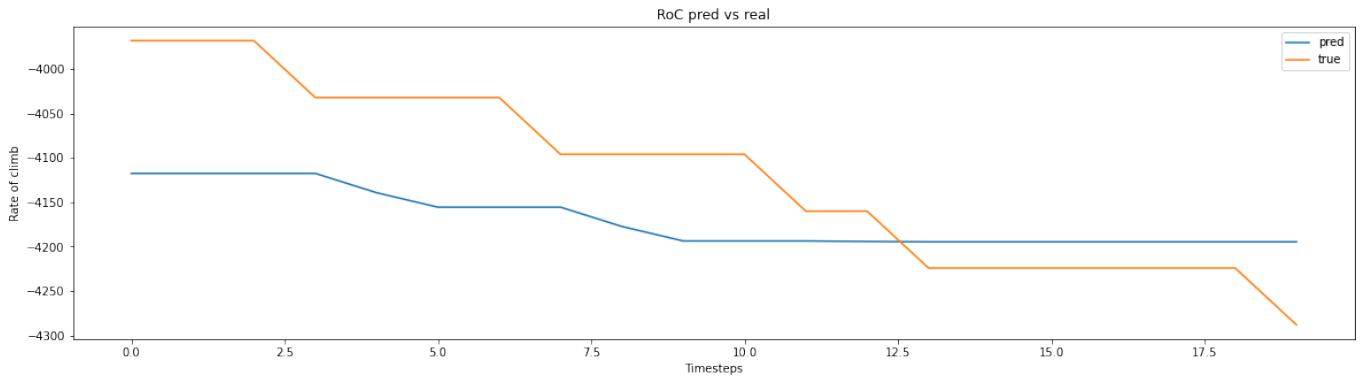
Residual Error: To evaluate how well the WaveNet model learns the temporal patterns of aircraft trajectories, we test it on batches of ADS-B message sequences. Each batch contains 10 consecutive timestamps, and the model predicts the next (11th) value for each feature. The predicted outputs are then compared with the corresponding ground-truth values.

(a) M 1 Training and Validation Accuracy (b) M 2 Training and Validation Accuracy (c) M 3 Training and Validation Accuracy  
Figure 6: M1, M2 and M3 Model Training and Validation Accuracy on ADS-B Decoded Messages

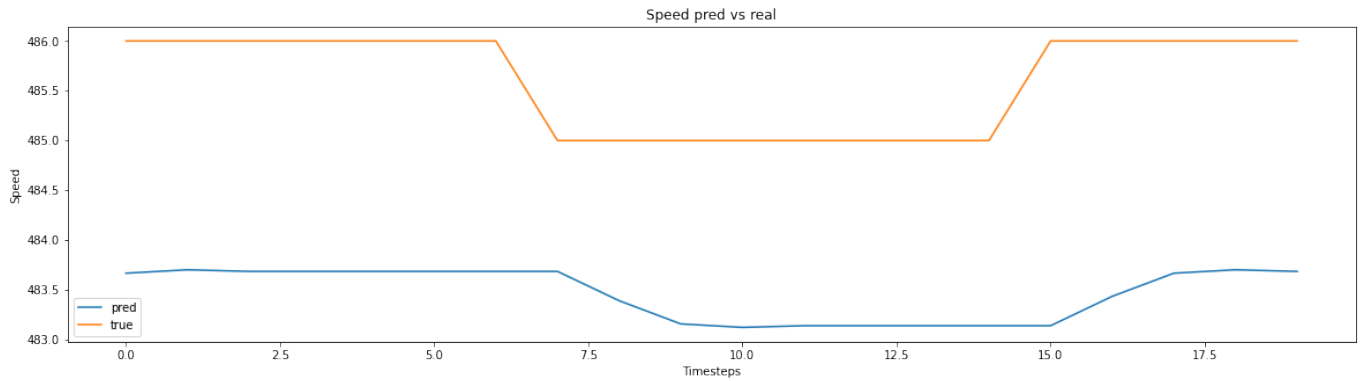
(a) M 1 Training and Validation Accuracy (b) M 2 Training and Validation Accuracy (c) M 3 Training and Validation Accuracy  
Figure 7: M1, M2 and M3 Model Training and Validation Accuracy on IQ samples

Figure 8: Illustration of sliding window

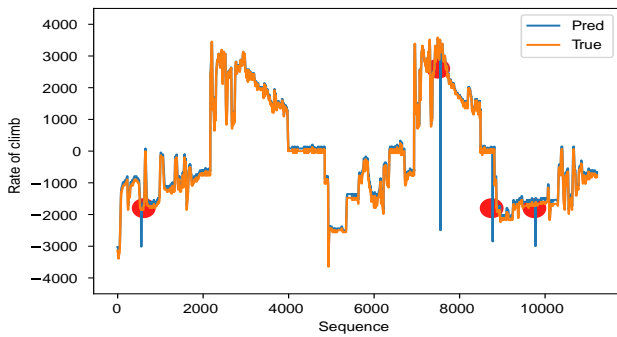
The Table 8 is a comparison between an LSTM and Wavenet score. In most cases, Wavenet has a lower overall MAE for all the ADS-B features, except for Altitude, which is relatively higher, where the LSTM model yields a better prediction. Finally, to counter spoofing attacks in which the ADS-B data impersonates an existing aircraft with its ICAO address, we



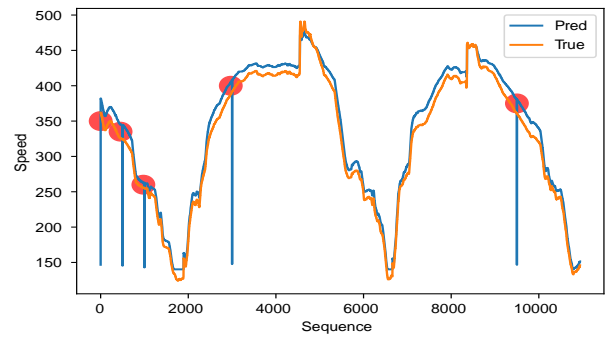
**Figure 9:** Prediction of 20 Time-steps for RoC



**Figure 10:** Prediction of 20 Time-steps for Speed



(a) Anomaly Detection in RoC



(b) Anomaly Detection in Speed

**Figure 11:** Anomaly Detection in RoC and Speed

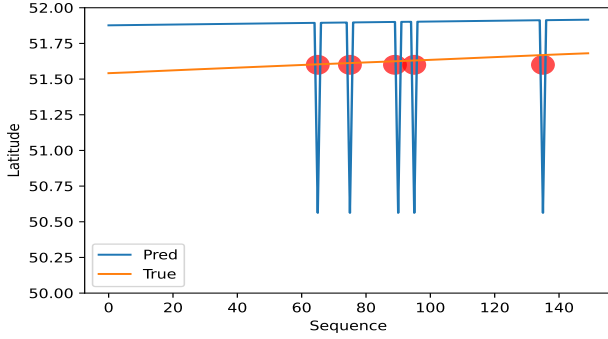
need to utilize transmitter signal characteristics for accurate aircraft identification. We cannot use an attack classification and an Anomaly detection framework in this case, since the ADS-B message is not altered or harmed; instead, the intruder is trying to impersonate and fake its existence as a real aircraft.

Thus, if we identify an aircraft based on its unique RF signal fingerprint, we can classify which aircraft it is originating from and whether an entity is trying to fake its identity. We utilize the raw IQ signal data to achieve this, extracting two essential features, phase and magnitude, from the IQ samples using signal processing techniques. The phase

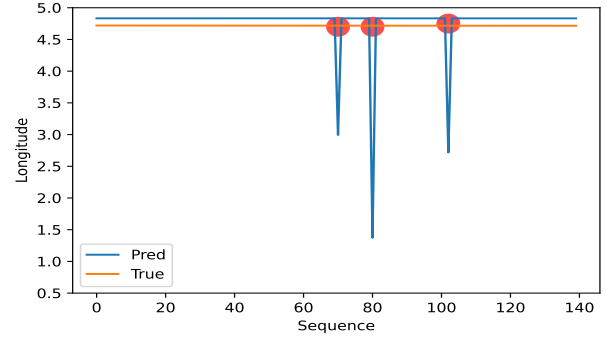
and magnitude act as a unique fingerprint to classify which aircraft it belongs to. Our input to DNN is a multi-modal input (Phase+Magnitude), and outputs are the class labels for each unique aircraft or ICAO. DNN is 4 layers deep with 500 hidden units. The detection accuracy for the spoofing attack is shown in Table 9.

*Final Summary: Best performing model*

Table 10 provides the comparison of the proposed framework with the previous work [23] [26] [27] [28] in terms of DS and FAR. We show that our proposed solution outperforms all other models in terms of high DS and very low FAR, which means the model predicts false attacks on infrequent



(a) Anomaly Detection in Latitude



(b) Anomaly Detection in Longitude

**Figure 12:** Anomaly Detection in Latitude and Longitude**Table 7:** ADS-B Data simulated attacks [21]

Attack	Simulation	Method	Result
Jamming	Random noise	Multiply the flight value obtained in the original ADS-B message by a random value between 0 and 1.	Success
Injection	feature replacement	Given certain feature information, inject different correct feature values to replace the sequence for the selected ADSB sequence segment.	Success
Modification	speed offset(+)	Use 20 knots as multiples to gradually change the speed characteristics of ADS-B messages by increasing speed by 20 knots each time.	Success
	Speed offset(-)	Decrease speed by 20 knots each time.	Success
	altitude offset(+)	Use 400 ft as multiples to gradually change the altitude characteristics of ADS-B messages by increasing altitude by 400 ft s each time.	Success
	Altitude offset(-)	Decrease altitude by 400 ft each time.	Success (0.33 $\geq$ $Th$ )
	Heading change	Change the value of the heading info contained in the ADS-B message to the opposite of the original value.	Success
	Climb rate change	Change the value of the climb rate contained in the ADS-B message to the opposite of the original value	Success

occasions. This can help us save costly time in analysing ADS-B data if there is no attack, leading to less panic during air-ground communication.

In Fig.13, we depicted the Lambert Conformal Conic (LCC) projection of the entire airspace for different aircraft's positions in a region for a fixed time. We can observe in the Fig.13 that the area circled in red with aircraft A10 is a fake aircraft, whereas other aircraft are legitimate.

#### Operational Considerations and Interpretability

In an operational ATC environment, even a low false positive rate can translate into a non-negligible number of alerts per day. Therefore, the proposed system is intended to function as a decision-support tool rather than an autonomous authority. Alerts generated by the spatial, temporal, or RF

**Table 8:** Absolute mean error for ADS-B features for 20 Timestamps

Model	Lat (deg)	Lon (deg)	Alt (m)	Spd (knots)	Hdg (deg)	Roc
<b>LSTM</b>	0.71	0.55	917.55	20.400	142.30	2734.67
<b>Wavenet</b>	0.54	0.44	5379.98	2.133	13.11	51.06

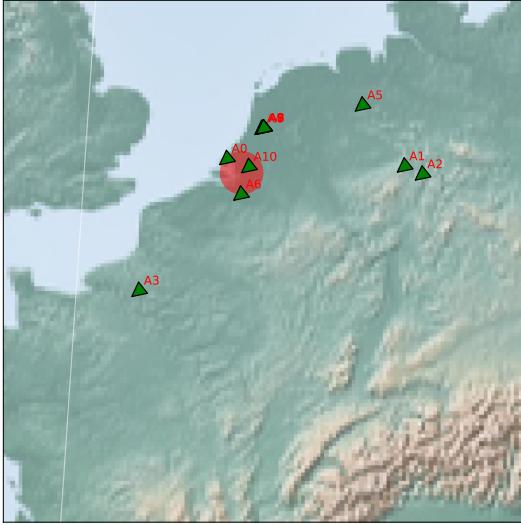
**Table 9:** Detection accuracy for aircraft classification

Model	Phase %	Magnitude %	Phase Magnitude % <sup>+</sup>
<b>DNN</b>	69.79	70.3	<b>87.20</b>

**Table 10:** Comparison of Proposed Framework With Existing Research

Models	Detection score (%)	FAR (%)
ADS-B (SODA) DNN [23]	99.34	0.43
GAN [26]	98.74	7.1
SVM [27]	80.29	25.67
NN (1 hidden, 10 nodes) [28]	91.4	13.8
<b>Our model (GCN on IQ)</b>	<b>99.74</b>	<b>0.25</b>
<b>Our model (GCN on ADS-B)</b>	<b>99.99</b>	-

fingerprinting stages can be cross-validated using additional surveillance sources such as multilateration, primary radar, or flight plan consistency checks before any operational action is taken. This layered verification approach reduces controller workload and limits the operational impact of occasional false positives. Another important aspect is model interpretability. To support controller trust and facilitate integration into safety-critical ATC workflows, future work will incorporate explainability techniques, such as feature attribution, saliency maps, and attention-weight visualizations. These methods can help identify which aircraft, kinematic features, or time intervals contributed most to an anomaly decision, enabling operators to more clearly distinguish between benign deviations (e.g., weather-related maneuvers) and genuine security threats. Enhancing interpretability is essential to ensuring that machine-learning-based surveillance aids can be adopted responsibly within real-world ATC systems.



**Figure 13:** Lambert conformal conic projection of airspace

## 6. CONCLUSION AND FUTURE WORK

The increasing reliance on ADS-B for global air traffic surveillance presents a critical challenge for securing its communications, which is essential for aviation safety. To address this, we presented a comprehensive, deep-learning-based framework that detects and mitigates multiple categories of ADS-B attacks without requiring changes to the protocol or additional infrastructure. The framework combines spatial graph convolutional networks for message-level intrusion detection, a WaveNet-based sequential model for temporal anomaly analysis, and an RF fingerprinting module that leverages IQ signal characteristics to identify spoofed or impersonated aircraft. This layered design ensures that abnormal behavior can be detected at the message, feature, and transmitter levels, providing resilience against message modification, injection, replay, jamming, and spoofing attacks. Despite its robust performance, achieving a globally deployable solution remains a challenge, as the system must adapt to diverse operational environments and increasingly sophisticated adversarial strategies. Refinement of the RF fingerprinting component and large-scale testing with real-world flight data are essential next steps to enhance scalability and robustness. Our experimental results confirm the effectiveness of the proposed framework, achieving 99.99% detection accuracy on decoded ADS-B data, 99.25% on raw IQ samples, and 87.2% accuracy in spoofing detection, with a FAR as low as 0.25%. These findings establish the framework as a promising direction for securing next-generation aviation networks and supporting safer air-ground communication worldwide.

## ACKNOWLEDGEMENTS

Trafikverket and Luftfartsverket supported this work under the MONAD project.

## REFERENCES

- [1] “FAA Aerospace Forecast FY 2022-2042.” [Online]. Available: <https://www.faa.gov/dataresearch/aviation/faa-aerospace-forecast-fy-2022-2042>
- [2] “Supporting european aviation,” Jan 2023. [Online]. Available: <https://www.eurocontrol.int/>
- [3] S. Khandker, H. Turtiainen, A. Costin, and T. Hämäläinen, “On the (in) security of 1090es and uat978 mobile cockpit information systems—an attacker perspective on the availability of ads-b safety- and mission-critical systems,” *IEEE Access*, vol. 10, pp. 37 718–37 730, 2022.
- [4] S. Akerman, E. Habler, and A. Shabtai, “Vizads-b: Analyzing sequences of ads-b images using explainable convolutional lstm encoder-decoder to detect cyber attacks,” 06 2019.
- [5] E. Habler and A. Shabtai, “Using lstm encoder-decoder algorithm for detecting anomalous ads-b messages,” *Computers & Security*, vol. 78, pp. 155–173, 2018.
- [6] M. Schäfer, V. Lenders, and I. Martinovic, “Experimental analysis of attacks on next generation air traffic communication,” in *International Conference on Applied Cryptography and Network Security*. Springer, 2013, pp. 253–271.
- [7] A. Costin and A. Francillon, “Ghost in the air (traffic): On insecurity of ads-b protocol and practical attacks on ads-b devices,” *black hat USA*, vol. 1, pp. 1–12, 2012.
- [8] S. Eskilsson, H. Gustafsson, S. Khan, and A. Gurtov, “Demonstrating ads-b and cpdlc attacks with software-defined radio,” in *2020 Integrated Communications Navigation and Surveillance Conference (ICNS)*, 2020, pp. 1B2–1–1B2–9.
- [9] A. Blåberg, G. Lindahl, A. Gurtov, and B. Josefsson, “Simulating ads-b attacks in air traffic management,” in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 2020, pp. 1–10.
- [10] A. Braeken, “Holistic air protection scheme of ads-b communication,” *IEEE Access*, vol. 7, pp. 65 251–65 262, 2019.
- [11] T. Kacem, D. Wijesekera, P. Costa, J. Carvalho, M. Monteiro, and A. Barreto, “Key distribution mechanism in secure ads-b networks,” in *2015 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. IEEE, 2015, pp. P3–1.
- [12] Z. Feng, W. Pan, and Y. Wang, “A data authentication solution of ads-b system based on x. 509 certificate,” in *27th International Congress of the Aeronautical Sciences, ICAS*, 2010, pp. 1–6.
- [13] A. Darabseh, H. AlKhazimi, and C. Pöpper, “Mavpro: Ads-b message verification for aviation security with minimal numbers of on-ground sensors,” in *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2020, pp. 53–64.
- [14] A. Melo, F. Souza, R. Albarello, R. Nunes, A. Gurtov, C. Ribeiro, C. Marcondes, and L. P. Junior, “A3p: Advanced airspace availability protocol for dynamic and trusted drone operations in smart cities,” in *Anais Estendidos do XXV Simpósio Brasileiro de Cibersegurança*. Porto Alegre, RS, Brasil: SBC, 2025, pp. 380–387.

