IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 June 2024, accepted 14 July 2024, date of publication 23 July 2024, date of current version 1 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432833

RESEARCH ARTICLE

Novel Machine Learning-Based Identification and Mitigation of 5G Interference for Radar Altimeters

ANAS AMAIREH AND YAN ZHANG, (Senior Member, IEEE)
Intelligent Aerospace Radar Team (IART), Advanced Radar Research Center, School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA

Corresponding author: Anas Amaireh (anas@ou.edu)

ABSTRACT As a critical aviation support sensor, radar altimeters face imminent challenges due to the interference caused by the frequency overlap with the developing 5G telecommunications networks. This paper addresses the critical challenge posed by 5G interference with radar altimeter signals, which is crucial for maintaining aviation safety. It introduces a novel machine learning (ML) framework to preserve altimeter accuracy in the presence of 5G signals. This framework first classifies signals into pure or interfered categories, then applies regression models to predict altitudes when interference is detected, effectively quantifying and mitigating the interference impact. Distinguished by using real 5G signals from a Norman, Oklahoma base station, the approach offers a realistic evaluation and demonstrates the ML framework's effectiveness in real-world conditions. Using various advanced ML models, the methodology showcases how critical aviation instruments can be safeguarded against the challenges posed by emerging telecommunications technologies, ensuring air travel safety and efficiency.

INDEX TERMS 5G, aviation safety, machine learning, radar altimeters, interference detection.

I. INTRODUCTION

Radar altimeters (RAs), also known as radio altimeters, are critical commercial and civil aircraft sensors that improve safety and navigation at all phases of flight [1]. These devices enhance situational awareness by measuring clearance height above terrain and potential obstacles, assisting Automatic Flight Guidance and Control Systems (AFGCS) during instrument approaches, and contributing to systems such as the Engine-Indicating and Crew-Alerting System (EICAS), Predictive Wind Shear (PWS), and Electronic Centralized Aircraft Monitoring (ECAM) [2]. In civil and commercial aircraft, up to three radar altimeters can be employed simultaneously to guarantee accurate altitude measurements. These altimeters, which include pulsed and frequency-modulated continuous-wave (FMCW) radars, measure the aircraft's above-ground level (AGL) height by emitting radio frequency (RF) energy to the ground and measuring the round-trip propagation time of the reflected energy [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Cheng Hu.

The introduction of 5G technology marks a major advancement in communication capabilities, providing higher data speeds and connectivity for diverse applications worldwide [4]. As countries rapidly deploy these networks, the use of the C-band spectrum for 5G systems has not fully considered the electromagnetic compatibility issues with other systems using the C-band spectrum. For example, agencies are reallocating a portion of the 3.7-4.2 GHz frequency band for 5G use, with the 3.7-3.98 GHz spectrum being assigned to new licensees since December 2020. This reallocation places the 5G frequencies close to those used by radar altimeters, raising concerns about potential signal interference [5]. Such interference could degrade altimeter functionality, significantly endangering aviation safety. Two types of 5G fundamental radiations might cause interference to the radar altimeter: emissions within the source's operating frequency band (in-band) and spurious emissions falling inside the 4.2-4.4 GHz band [6] (out-of-band).

As part of the existing solutions, band-pass filters (BPF) were incorporated to control the interference in the 3.7 to 3.98 GHz range. These filters are suitable for direct

integration with radar altimeter systems. However, while these static filters provide an immediate solution, they cannot adapt to the rapidly evolving 5G spectrum and unexpected signal variations. Also, another study in [7] discusses using BPFs optimized for minimal propagation delays and sharp transitions, which are effective in controlled scenarios but require precise engineering to match specific interference profiles, making them less adaptable in variable conditions. Furthermore, the study in [8] employs a more dynamic approach by adjusting its frequency response based on altitude measurements. While this is an improvement over static filters, it still cannot handle unexpected conditions. In contrast, the digital-signal-processing method offers significant advantages by continually adapting and learning from data, making it highly effective in environments with variable 5G interference [9]. ML models effectively handle complex, non-linear data relationships, making them suitable for aviation environments where traditional filters may fail [10]. ML enhances radar system efficiency by automating adjustments and leveraging real-time data, which reduces costs and minimizes errors [11]. The initial data and computing infrastructure investment pays off by decreasing the need for manual re-calibrations and lowering maintenance costs. ML's scalability allows for ongoing updates without hardware modifications, offering cost-effectiveness over time [11].

This study uses Machine Learning (ML) models and algorithms to classify radar altimeter signals subjected to 5G NR interference. We also aim to improve the accuracy of altitude measurements in the presence of potential 5G interference in radar altimeters. Figure 1 shows a flow chart detailing all of the steps included in our study. The technique begins with careful signal collection and validation using a low-cost Software-Defined Radio (SDR), resulting in a dataset that supports examining interference risks such as spurious emissions, receiver front-end overloading, receiver desensitization, and inter-modulation products. The study simulates radar altimeter signals and refines the dataset to include around 20,000 signals across a 1.676-kilometer range to ensure a wide variety of data for training. The preprocessing step begins with segmenting the 5G base station signals and re-scaling them to emulate various interference levels. The following part of the methodology is based on an ML classification model that differentiates between the radar altimeter signals and 5G NR interference while also considering the particular challenges presented by different waveform characteristics. In particular, after detecting interference, we use two ML models to estimate flight altitudes for radar waveforms using down and up-frequency sweeps. By combining these ML models' altitude predictions, we can obtain a more accurate and consistent measurement by mitigating outliers. The study concludes with a detailed comparison of the ML models' performance, illustrating their effectiveness and the effect of frequency modulation sweep direction on altitude prediction in the presence of 5G interference.

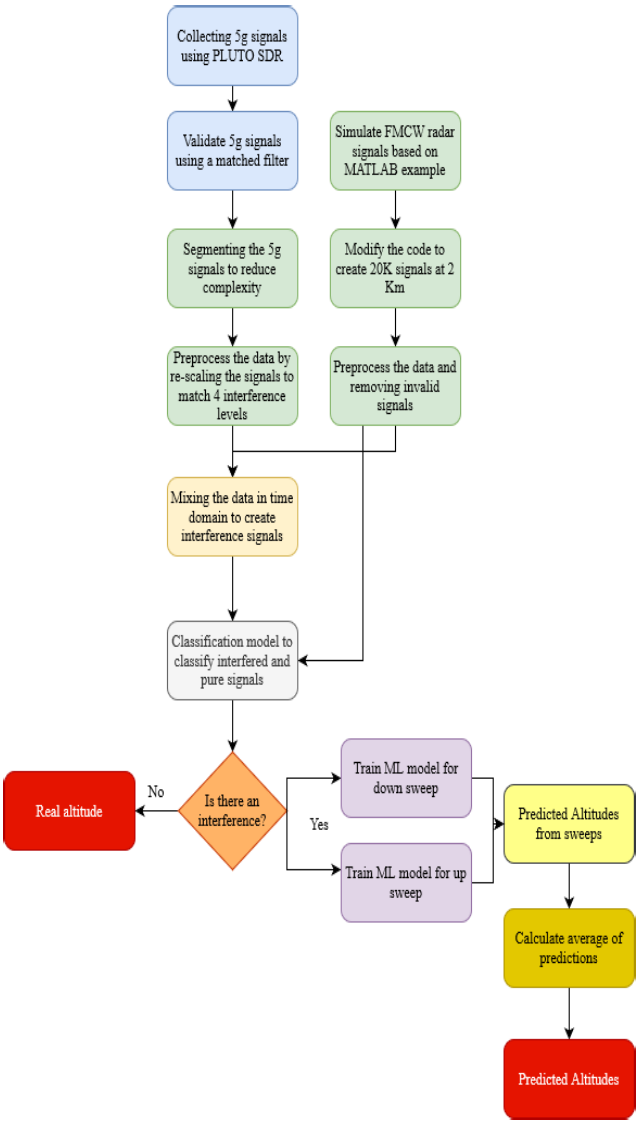


FIGURE 1. Complete processing flow. This diagram shows the steps from signal collection with a Software-Defined Radio (SDR) to machine learning models for classifying and predicting altitudes under 5G interference.

Part of this work has been previously published in the SPIE conference proceedings: ''Improved investigation of electromagnetic compatibility between radar sensors and 5G telecommunications'' [12].

The rest of this paper is organized as follows: Section II describes the data collection process, including the methods for 5G signal collection, verification, and using a radar altimeter simulator to emulate 5G interference. Section III details the methodology, covering feature extraction from time-domain and frequency-domain data, preprocessing steps, performance evaluation metrics, and the machine learning algorithms and methods utilized in this study. Section IV presents the results and discussion, including classification and altitude prediction results, with specific cases for up-sweep and down-sweep scenarios. Finally, Section V concludes the paper and suggests future work.

II. DATA COLLECTION

A. 5G SIGNAL SAMPLING AND DATA COLLECTION

The ADALM-PLUTO Software-Defined Radio (SDR) was used to collect 5G signals at a center frequency of 3.7 GHz, which is adjacent to the aviation radar altimeter range of 4.2-4.4 GHz. The ADALM-PLUTO, which has an operating range of up to 3.8 GHz, was tuned to a central frequency of 3.7 GHz and configured for high-precision capture at 2 million samples per frame. This configuration enabled us to collect a diverse dataset of 1000 signal frames across various times, locations, and weather conditions, providing insights into 5G signal behavior and its potential for interference with adjacent bands. Focusing on the 3.7 GHz band, particularly the n77/n78 5G bands, allowed us to explore possible interference with radar altimeters.

1) VERIFICATION OF 5G SIGNAL

To confirm that the collected signals are from actual 5G base stations, we investigated their cross-correlation with key 5G synchronization signals: the Primary Synchronization Signal (PSS) and the Secondary Synchronization Signal (SSS). These signals, broadcast by base stations, are critical for mobile devices to identify and synchronize with the network, hence initiating the cell search process. The PSS, with its high autocorrelation and specific structure, aids in accurate time synchronization and identifies the cell's Physical Layer Cell Identity Group (PCI Group). Following PSS detection, the SSS refines the cell identification and is required to determine the frame boundary and access broadcast channel information. Cross-correlating the received signal with PSS and SSS sequences effectively confirms their presence. Distinct peaks in the correlation output signal the alignment of the received signal with these sequences, indicating genuine 5G NR signals.

In addition, the power spectrum density plots shown in Figure 2 display a prominent peak at the center, characteristic of the carrier frequency used in 5G communications. The shape of the power spectrum, with its distinct roll-off, further supports the spectral profiles expected of 5G NR transmissions. The cross-correlation analysis using the PSS, as shown in Figure 3 (a), produces noticeable peaks that rise above the noise level. These peaks indicate the PSS within the received signal. Similarly, Figure 3 (b) shows that the cross-correlation with the SSS. The presence of significant peaks at specified lags verifies the detection of the SSS.

B. SIMULATION OF RADAR ALTIMETERS

A system simulation of an FMCW radar altimeter was developed and used to generate radar I/Q data. The radar parameters are detailed in Table 1. Waveform and sensor parameters are set to simulate the airplane landing process beginning at an altitude of 1,676 meters. The simulation includes the ability to adjust different losses, terrain types, etc. Around 20,000 radar signal samples are simulated for various aircraft altitudes. Standard FMCW radar signal

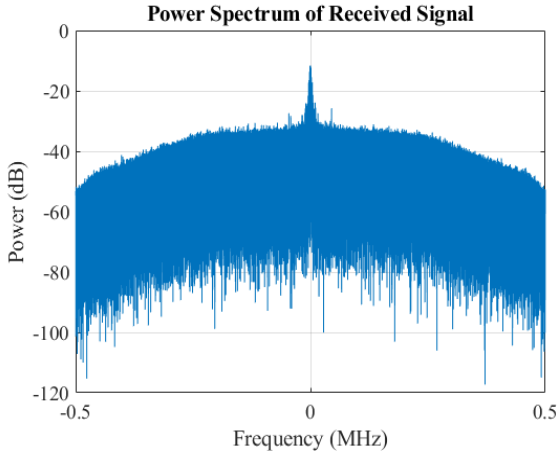
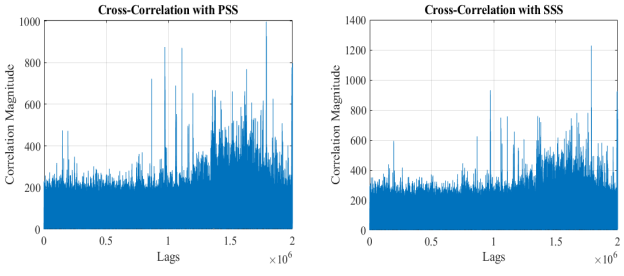


FIGURE 2. Power spectrum of the received 5G signal, showing a peak at the carrier frequency used in 5G communications.



(a) (b)

FIGURE 3. Cross-Correlation Analysis with (a) Primary Synchronization Signal (PSS), and (b) Secondary Synchronization Signal (SSS). Distinct peaks confirm the presence of 5G signals.

TABLE 1. Parameters of the FMCW radar altimeter used in the simulation.

Parameter	Value
Center Frequency	4.3 GHz
Bandwidth	150 MHz
Chirp Repetition Frequency (CRF)	22 KHz
Antenna Beamwidth	40 degrees
Transmitter Power	0.4 W
Maximum Range	1,676 m
Sweep Direction	Triangle
Noise Figure	8 dB

processing [13] is applied for the altitude estimations, which are updated at a 10 Hz rate.

Fig. 4 displays spectrograms of the received radar altimeter signals at altitudes of 1493.29 m, 969.68 m, 371.51 m, and 76.57 m, respectively. Signal strength visibly increases as altitude decreases, marked by the values changing in warmer colors. The pattern structure in the spectrograms at higher altitudes is potentially caused by the radar's antenna beam coverage at higher altitudes. This spreads across frequencies, particularly when the radar approaches its maximum altitude of 1,676 meters. As the altitude decreases, the radar's footprint narrows, and the spectrum spreading decreases, leading to "straight line" type of spectral patterns. This characteristic is one of the possible features used in the ML algorithms to discriminate radar signals from 5G-base station signals from the ground.

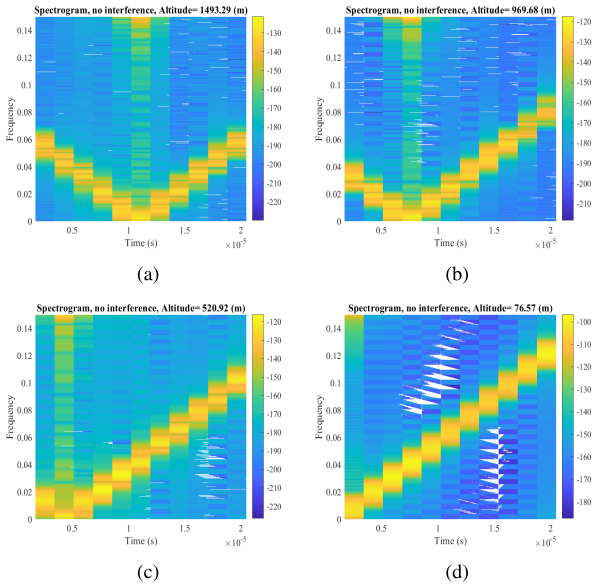


FIGURE 4. Spectrogram of the sample FMCW radar altimeter signals at the following altitudes: (a) 1493.29 meters, (b) 969.68 meters, (c) 520.92 meters, and (d) 76.57 meters.

C. EMULATING RADAR ALTIMETER SIGNALS AND 5G RFI

Integrating real 5G and simulated radar signals through time-domain superposition of In-phase and Quadrature (IQ) data was performed. A segmentation approach was applied to handle the challenge of time sampling in the captured 5G signals. This technique maintained the integrity of 5G interference characteristics while simplifying the number of input features for effective machine learning (ML) training. Following segmentation, the airplane landing process was simulated with two critical factors: the increase in radar altimeter signal power as the aircraft descends and the rise in 5G interference levels caused by the aircraft's proximity to the 5G base stations. For the detection and 5G-RFI classification, the interference was classified into four level categories during descent: low (for altitudes between 1,676 and 1,138 meters), moderate (for altitudes between 1,138 and 782.6154 meters), high (for altitudes between 782.6154 and 427.5399 meters), and extreme (for altitudes between 427.5399 meters and the ground). The raw 5G signal segments were rescaled to match the simulated aircraft altitudes, using statistical power values from radar altimeter signals at each altitude. Furthermore, the re-scaled 5G and radar altimeter signals were randomly combined for each interference category. Mixing the adjusted 5G interference signals with the simulated radar altimeter signals produced a dataset reflecting the full range of interference levels correlated with the aircraft's altitude. Fig. 5 illustrates spectrograms for different signals at different altitudes before (''pure'' signal) and after adding the interference signals to the pure radar altimeter signals. As the simulated spectrum shows, the Radio Frequency Interference (RFI) from the 5G signals mainly affects the radar altimeter through the out-of-band (OOB) spectrum. It can be more significant at

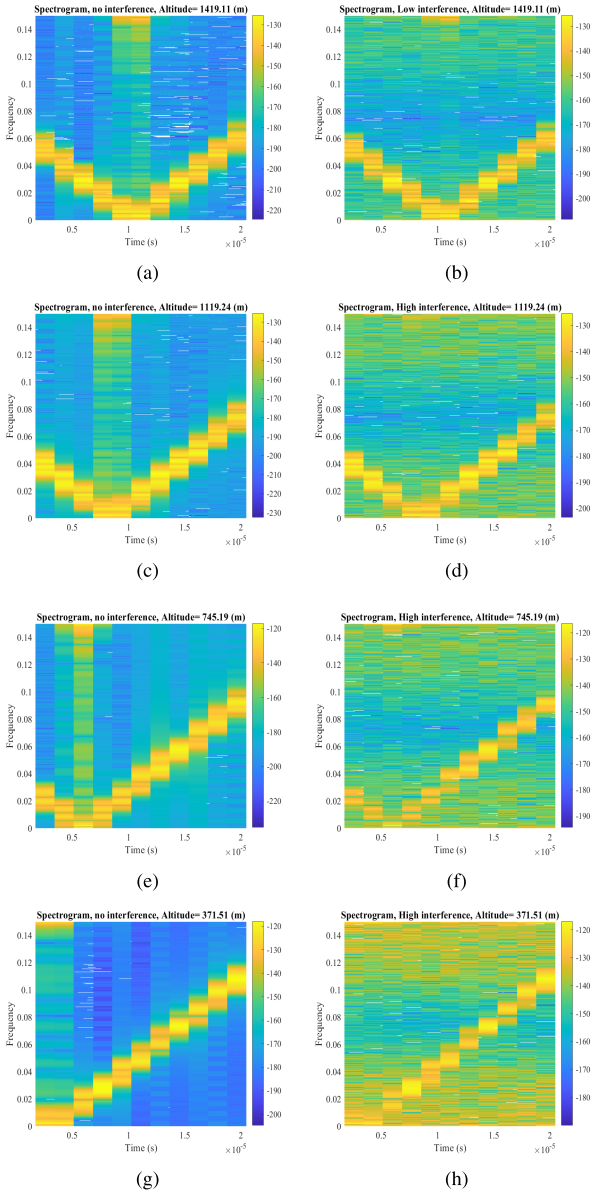


FIGURE 5. Spectrograms Showing the Effect of 5G Interference on Radar Altimeter Signals at Various Altitudes: (a) 1419.11 meters, (b) 1119.24 meters, (c) 745.19 meters, and (d) 371.51 meters.

some time due to its noise-like nature. At a lower altitude of 371.51 meters with high interference, there is a noticeable distortion in the signal's spectral components, illustrating the interference's significant impact. The effect of interference is less significant at higher altitudes. Fig. 6 shows an example of signal segments in the time domain before and after adding the interference. Notably, at the lower altitude of 371.51 meters, the interfered radar signal displays significant fluctuations in amplitudes.

III. DETAILED METHODOLOGY

A. FEATURE EXTRACTION

1) TIME-DOMAIN FEATURES

Based on typical signal feature analysis [14], time-domain signal features include:

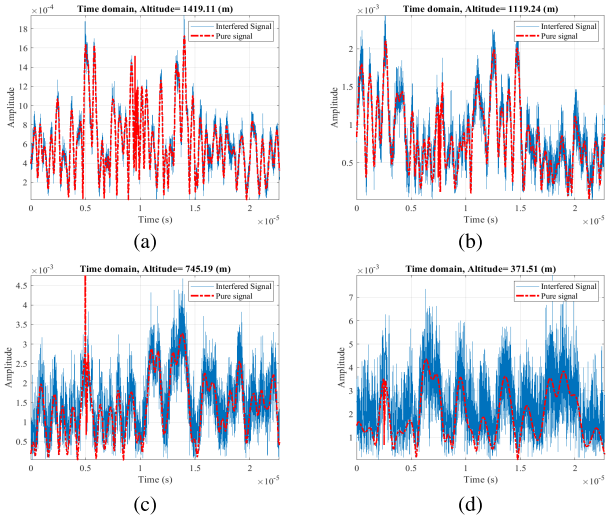


FIGURE 6. Time-Domain Analysis of Radar Altimeter Signals with and without 5G Interference: (a) 1419.11 meters, (b) 1119.24 meters, (c) 745.19 meters, and (d) 371.51 meters.

- **Peak amplitude**: Detects important signal reflections, which are critical for identifying altitudes.
- **Root Mean Square (RMS) amplitude**: Distinguishes signals from noise, aiding in accurate altitude determination.
- **Kurtosis and skewness**: Provide insights into the signal's amplitude distribution, showing potential outliers and asymmetry introduced by interference [15].
- **Standard deviation and variance**: Measure amplitude variations.
- **Signal power and Signal-to-Noise Ratio (SNR)**: Indicate the overall strength of the signal relative to noise.

These features collectively present the effects of 5G interference on radar altimeter signals. The statistical distributions of the ''pure'' and interfered signals for down and up sweeps are shown in Fig. 7, in the form of boxplots for peak amplitude, RMS amplitude, signal kurtosis, skewness, standard deviation (STD), and signal power, which collectively present the effects of 5G interference on radar altimeter signals with clear attenuation in features during down sweeps. The peak amplitude boxplots show attenuation in the presence of interference, especially during down sweeps, indicating possible difficulty in identifying ground reflections at lower altitudes, especially important during takeoff and landing. Variations in kurtosis and skewness in the interfered radar signals show clear non-normal distributions with heavier tails and asymmetric amplitude values.

Fig. 8 illustrates the impacts 5G-RFI on the radar signal amplitude and power features at different altitudes. The power profile does not show a consistent reduction across all altitudes. Instead, there is a substantial difference between the original and interfered signals at lower altitudes, which narrows as altitude increases. This pattern indicates a complex relationship between interference strength and

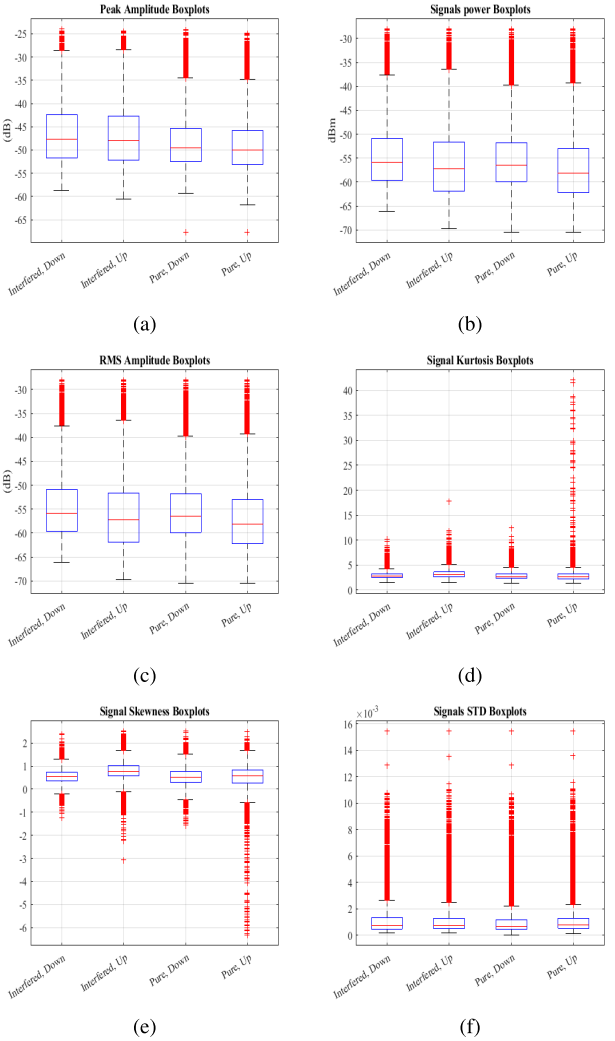


FIGURE 7. Boxplots of the time-Domain Features used by the ML processing for the original and interfered radar altimeter signal for altitude = 317.5 meters. All the features are based on the amplitudes of one time-frame capture. (a) Peak amplitudes, (b) Signals power values, (c) RMS amplitude values, (d) Signal Kurtosis, (e) Signals' skewness, (f) Signal amplitudes' standard deviation (STD).

altitude. Furthermore, the peak amplitude profile shows that interference has the greatest influence in the critical lower altitude zones, affecting important aircraft operations such as takeoff and landing. Integrating these time-domain features into machine learning models improves prediction accuracy, allowing for a more accurate estimate in the presence of interference. This approach assures that machine learning algorithms can handle the complexity of signal variance, keeping the integrity and reliability of radar altimetry in aircraft.

2) FREQUENCY DOMAIN FEATURES

Spectral features have been computed for the original (or ''pure'') and interfered radar altimeter signals across the up and down sweeps. The calculated features include:

- **Spectral Centroid**: Indicates where most of the signal energy is concentrated in the frequency spectrum [16].

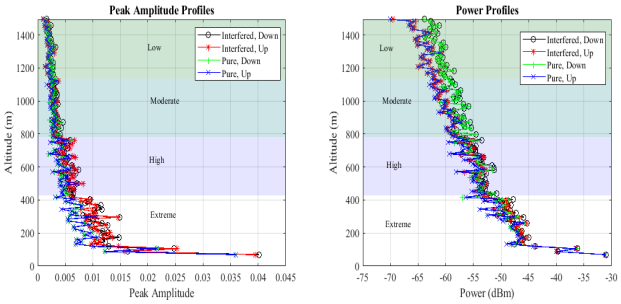


FIGURE 8. Altitude-Dependent Profiling of (a) Peak Amplitudes, and (b) Power levels for the Original Radar Altimeter Signals and the Signals with the 5G-RFI Effects.

- **Bandwidth**: Measures the range of frequencies present in the signal, which can broaden due to interference.
- **Flatness**: Indicates how noise-like the signal is [17].
- **Rolloff**: Shows how quickly the signal's energy drops off [18].
- **Slope**: Describes the tilt of the signal spectrum.
- **Crest**: Measures the peak-to-average ratio of the signal.
- **Decrease**: Evaluates amplitude decline in the spectral distribution [19].
- **Entropy**: Represents the randomness in the spectral distribution, which increases with interference [20].
- **Energy**: Represents the total signal power.
- **Edge Frequency**: Defines the upper-frequency boundary of the signal's energy [21].
- **Zero Crossings**: Indicates the stability of the signal's frequency components [21].

These features are essential to machine learning (ML) frameworks that aim to improve the accuracy of altitude predictions and interference detection. Fig. 9 illustrates the spectral centroid, bandwidth, and entropy features. Interference usually increases spectral bandwidth, as seen by the broadened frequency spread, and introduces variability in the spectral centroid, indicating variations in the energy distribution within the spectrum. Spectral entropy increases to imply that interference causes more uncertainties. This study used time-domain and frequency-domain signal features for RFI detection and possible future mitigation. However, more weighting was placed on the time-domain features based on computational efficiency considerations.

B. PREPOSSESSING

The pre-processing step extracts and combines both time-and-frequency domain features into training and testing datasets. A signal classification approach based on time-domain samples and features is proposed by maintaining the original radar receiver sampling frequency of 149.996 MHz within a 150 MHz bandwidth. The dataset was segmented for altitude estimations based on sweep direction (up or down) before ML model training. This technique makes creating specific ML models for each sweep direction easier, using the unique altitude information

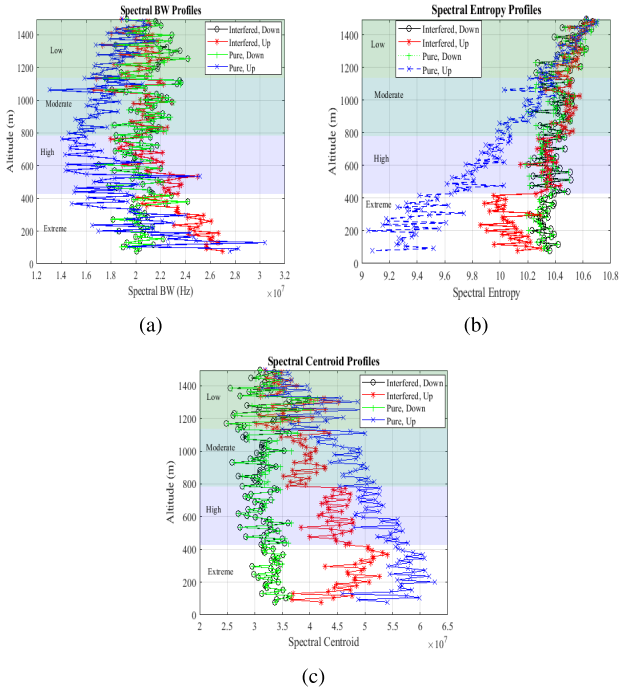


FIGURE 9. Altitude-Dependent Profiling of (a) Spectral BW, (b) Spectral Entropy, and (c) Spectral centroid for up and down sweeps for the original radar altimeter signals and the signals with the 5G-RFI Effects.

within each. Training separate models for up and down sweeps also allowed for capturing distinct altitude aspects more effectively. Next, a five-fold cross-validation across classification and altitude prediction phases was implemented for enhanced model reliability and generalizability. This technique, critical for assessing model performance consistently across various scenarios, was complemented by outlier detection and removal to refine data quality.

C. PERFORMANCE EVALUATION METRICS

Different ML processing performance metrics are used in this study based on their characteristics. For instance, Accuracy is a fundamental metric that measures the proportion of total correct predictions made by the model out of all predictions [22]:

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

The Total Cost (TC) is a metric used to evaluate machine learning models by assigning penalties to incorrect classifications through a cost matrix, where rows represent true classes and columns represent predicted classes. The total misclassification cost is calculated by summing the products of the cost matrix and the confusion matrix [23]:

$$Total\ Cost = \sum (Confusion\ Matrix \times Cost\ Matrix) \quad (2)$$

Positive Predictive Value (PPV), also known as Precision, measures the model's ability to categorize positive instances

properly [24]:

$$PPV = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

The False Discovery Rate (FDR) is the complement of PPV and displays the proportion of false positives within predicted positives [24]:

$$FDR = \frac{\text{False Positives}}{\text{True Positives} + \text{False Positives}} \tag{4}$$

The True Positive Rate (TPR), or Sensitivity, evaluates the model's ability to detect all true positive instances [24]:

$$TPR = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{5}$$

In contrast, False Negative Rate (FNR) is the rate at which actual positives are wrongly predicted as negatives [24].

$$FNR = \frac{\text{False Negatives}}{\text{True Positives} + \text{False Negatives}} \tag{6}$$

On the other hand, various statistical criteria are used to analyze the effectiveness of altitude prediction algorithms, each providing a unique perspective on the algorithms' accuracy. These include Correlation Coefficient (ρ), which examines the linear connection between actual and predicted data, obtained as follows [25]:

$$\rho\left(X_{\text{true}}, X_{\text{predicted}}\right) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{X_{\text{true}}(i) - \mu_{\text{true}}}{\sigma_{\text{true}}}\right) \times \left(\frac{X_{\text{predicted}}(i) - \mu_{\text{predicted}}}{\sigma_{\text{predicted}}}\right) \tag{7}$$

N represents the number of data points. μ_{true} and $\mu_{\text{predicted}}$ represent the mean values of the true and predicted data sets, respectively, while σ_{true} and $\sigma_{\text{predicted}}$ are their corresponding standard deviations. This metric, ranging from -1 to 1, offers insight into the extent to which the predicted values reflect actual value changes, although it doesn't directly quantify prediction accuracy or error magnitude. Following this, the Mean Absolute Error (MAE) serves as a crucial metric, represented by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|X_{true}\ (i) - X_{predicted}\ (i)\right| \tag{8}$$

Here, N is the total sample count, X_{true} the observed values, and $X_{predicted}$ the predicted values, respectively. MAE averages the absolute differences between actual and predicted values, which is valuable for cases where all errors are equally significant.

The Mean Squared Error (MSE) comes next, defined as [26]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(X_{true}\ (i) - X_{predicted}\ (i)\right)^2 \tag{9}$$

TABLE 2. List of the hyperparameters of the classification models.

Model	Parameters
Fine Tree	Splits: 100, Criterion: Gini, Surrogate: Off
Medium Tree	Splits: 20, Criterion: Gini, Surrogate: Off
Coarse Tree	Splits: 4, Criterion: Gini, Surrogate: Off
Narrow NN	Layers: 1 (10), ReLU, Iter: 1000, λ: 0, Std: Yes
Medium NN	Layers: 1 (25), ReLU, Iter: 1000, λ: 0, Std: Yes
Wide NN	Layers: 1 (100), ReLU, Iter: 1000, λ: 0, Std: Yes
Bi NN	Layers: 2 (10, 10), ReLU, Iter: 1000, λ: 0, Std: Yes
Tri NN	Layers: 3 (10, 10, 10), ReLU, Iter: 1000, λ: 0, Std: Yes
Fine KNN	Neighbors: 1, Metric: Euclidean, Weight: Equal, Std: Yes
Gaussian NB	Distribution: Gaussian
Logistic Reg	None

MSE prioritizes larger errors by squaring the difference between actual and predicted values, which is useful for penalizing significant deviations more severely. The Root-Mean-Square Error (RMSE) provides a nuanced look at prediction accuracy:

$$RMSE = \sqrt{\left[\left(\sum_{i=1}^{N}\left(X_{\text{true}}\ (i) - X_{\text{predicted}}\ (i)\right)^2\right)/N\right]} \tag{10}$$

D. ML ALGORITHMS AND METHODS USED IN THIS STUDY

This study applies various ML algorithms to classification (detecting the presence of 5G-RFI and classifying the interference levels) and regression (estimating the approaching altitudes in the presence of 5G-RFI). Decision Trees, organized into hierarchical decision-making pathways ranging from fine to coarse, adopted various decision depth and complexity levels. The fundamental advantage of Decision Trees is their interpretability since they allow tracing decisions from input features to outcomes, although overfitting is possible in highly complicated cases. The Artificial Neural Network (ANN) models also solve classification and regression problems. They range from narrow to tri-layered networks, each with different levels of complexity and functionality. Table 2 details the machine learning models applied for classifying pure and interfered signals, including their hyperparameters, while Table 3 lists the models used for regression-altitude prediction along with their respective hyperparameters. The methodology was executed on a computer with an i7-2600K CPU and 24 GB of RAM. The integration of these strategies highlights a sophisticated balance between maintaining signal integrity and computational practicality.

IV. RESULTS AND DISCUSSION

A. 5G-RFI DETECTION RESULTS

The signal samples with various levels of RFI and without RFI are divided into testing (20%) and training datasets (80%). The first objective is to discriminate the radar signal samples containing any levels (low, moderate, high, and extreme, called Class 1) of RFI from the "pure" radar altimeter signals (called Class 2). Table. 4 compares

TABLE 3. List of the hyperparameters of the regression models.

Model	Parameters
Fine Tree	Leaf size: 4
Medium Tree	Leaf size: 12
Coarse Tree	Leaf size: 36
Boosted Tree	Leaf size: 8, Learners: 30, Rate: 0.1
Bagged Tree	Leaf size: 8, Learners: 30
Linear Reg.	Terms: Interactions, Robust: Off
Narrow NN	Layers: 1 (10), ReLU, Iter: 1000
Medium NN	Layers: 1 (25), ReLU, Iter: 1000
Wide NN	Layers: 1 (100), ReLU, Iter: 1000
BiNN	Layers: 2 (10 each), ReLU, Iter: 1000
TriNN	Layers: 3 (100, 50, 10), ReLU, Iter: 1000

TABLE 4. Performance of radar signal RFI level classification for different ML models.

Model	Accuracy (%)		Total Cost		Speed	Time
	Val	Test	Val	Test	(obs/sec)	(sec)
F. Tree	98.2	98.4	1138	187	11618	369
M. Tree	97.0	97.3	1926	313	12587	287
C. Tree	95.2	95.7	3096	490	13789	223
N. NN	91.4	93.4	5535	748	9651	9910
M. NN	91.5	92.6	5516	845	9375	8397
W. NN	91.7	92.9	5387	815	4283	21755
BiNN	91.2	91.9	5689	918	9554	10210
TriNN	91.0	92.2	5831	888	3834	28930
F. KNN	21.4	11.0	50755	10148	14	29474
G. NB	79.4	89.1	13303	1244	1120	29685
L. Reg.	90.6	87.2	6061	1459	2987	40387

machine learning models' performance in radar signal classification under interference. The Fine Tree model has remarkable accuracy both in validation (98.24%) and testing (98.36%), as well as the lowest total cost in testing (187), making it the most effective model for exact classifications with few misclassification errors. Its prediction speed (11618.28 observations per second) and comparatively low training time (369.22 seconds) highlight its effectiveness. The Medium Tree and Coarse Tree models exhibit a decreasing pattern in accuracy and an increasing pattern in total cost over the validation and testing phases, Despite its lower accuracy, the Coarse Tree model has the fastest prediction speed (13788.99 obs/sec), Neural Network models perform differently, with the Narrow, Medium, and Wide configurations achieving test accuracies of 93.44%, 92.59%, and 92.85%, respectively. Notably, increasing network size does not directly correspond with higher accuracy, as seen by the Wide Neural Network's lengthy training period (21754.53 seconds), but only modestly better test accuracy than its narrower counterparts. The Bilayered and Trilayered networks underscore the declining results on increasing complexity, with test accuracies of 91.95% and 92.21%, respectively, with the Trilayered Neural Network having the longest training time (28930.28 seconds). The Fine KNN model significantly fails in accuracy and cost parameters, In contrast, the Gaussian Naive Bayes and Logistic Regression models provide feasible alternatives. However, they also demonstrate trade-offs between accuracy and computational demands.

Fig. 10 depicts the confusion matrices of the training dataset for several ML models. The Fine Tree model's confusion matrix shows good accuracy, with true positives for Class 1 and Class 2 stated at 31325 and 32137, respectively, compared to very low false negatives and false positives (248 and 890). In contrast, the KNN model does much worse, with true positives of 4756 for Class 1 and 9089 for Class 2, but misclassification is significantly greater, particularly for false positives in Class 1 (27459). The Gaussian Naive Bayes model performs well, with true positives of 30913 for Class 1 and 20384 for Class 2. Given its performance, the model shows limits with 1,302 false negatives for Class 1 and 12,001 false positives for Class 2, suggesting constraints related to its assumptions of feature independence. Logistic regression performs effectively, with true positives of 28697 for Class 1 and 29842 for Class 2, indicating that it can classify accurately. The false positives and false negatives for Class 1 are 3518 and 2543, respectively, showing an acceptable degree of accuracy. The Narrow Neural Network has true positives of 29329 for Class 1 and 29736 for Class 2, and there are false negatives (2649) and false positives (2886). The Tri-layered Neural Network's confusion matrix shows true positives 29320 for Class 1 and 29449 for Class 2, with fewer false negatives (2936) and false positives (2895). These results demonstrate the advantages of neural network architecture for certain classification tasks and the capability to leverage deeper networks to attain greater classification accuracy.

The confusion matrices in Fig. 11 summarize the ML algorithm performance for the test data sets. The Fine Tree model demonstrates a good classification ability, with 5638 true positives for Class 1 and 5575 true positives for Class 2, suggesting a high accuracy in detecting accurate classifications. The misclassifications are 147 for Class 1 and 40 for Class 2, indicating a low error rate, making the Fine Tree model very dependable for both classes. KNN performed significantly lower, with 391 true positives for Class 1 and 861 for Class 2. The higher misclassification, 5394 false positives for Class 1, emphasizes the model's difficulty in classifying classes. The GNB model has a better result, with 5223 true positives in Class 1 and 4933 in Class 2, with false positives and negatives of 562 and 682, respectively. Although not as good as the FT, the GNB achieves an acceptable performance between accurate classifications and errors. The LR model reveals 4967 true positives for Class 1 and 4974 for Class 2, with a larger number of false positives and negatives, specifically 818 false positives for Class 1. The increasing error rates indicate that LR is effective. Furthermore, the NNN works well, with 5387 true positives for Class 1 and 5265 for Class 2, with slightly higher but still low misclassification counts of 398 and 350 for Class 1 and Class 2, respectively. This illustrates that the model takes a balanced approach to both classes. The TNN has 5311 true positives for Class 1 and 5201 for Class 2, with 474 and 414 mis-classifications, respectively. These results show that the TNN maintains strong predictive performance but slightly higher error rates than the Fine Tree model.

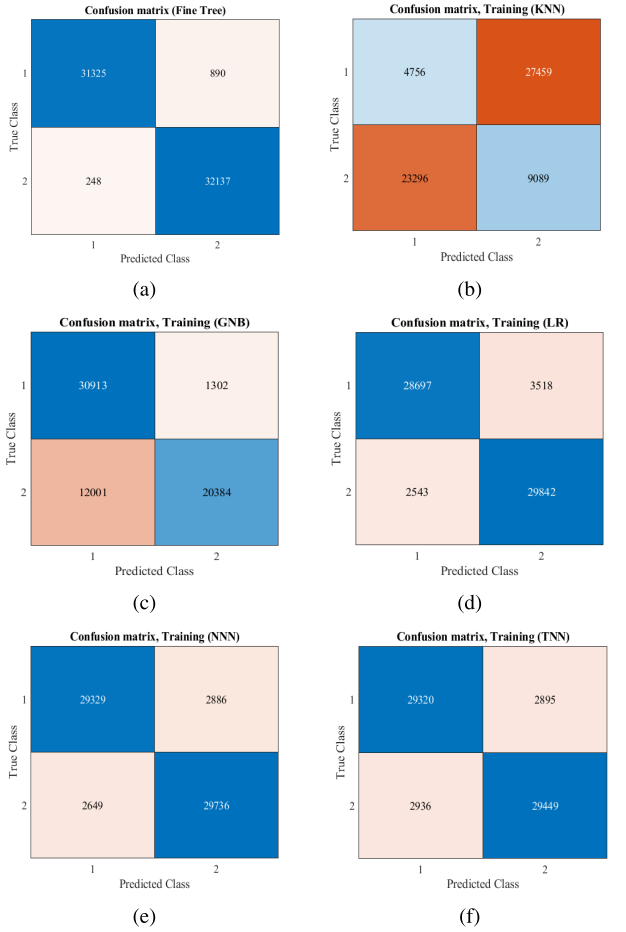


FIGURE 10. Training dataset confusion matrices for (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

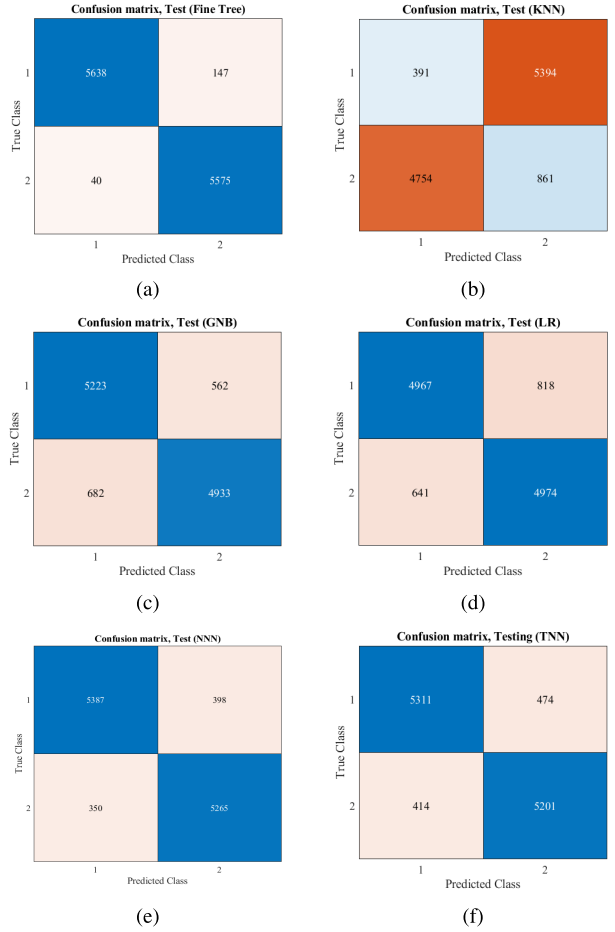


FIGURE 11. Testing dataset confusion matrices for (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

Fig. 12 displays confusion matrices in percentage format, containing essential metrics such as True Positive Rate (TPR) and False Negative Rate (FNR), to illustrate classification performance for the training dataset. In the case of the Fine Tree model, the TPR is an outstanding 97.2% for Class 1 and 99.2% for Class 2, while FNRs remain low at 2.4% and 0.8%. This shows the model's excellent accuracy and consistency in signal categorization for both classes. The KNN model has an extremely low TPR of 14.8% for Class 1 and 28.1% for Class 2 and FNRs of 85.2% and 71.9%, indicating poor performance. This implies a significant level of misclassification, particularly in identifying Class 2 signals. The performance metrics for the GNB model reveal a TPR of 96.0% in Class 1 and a lower TPR of 62.9% in Class 2. The FNR for Class 1 is only 4.0%, while it climbs to 37.1% for Class 2, showing that the model's performance favors Class 1 over Class 2. The NNN shows TPRs of 91.0% for Class 1 and 91.8% for Class 2. The FNRs for this model remain at 9.0% for Class 1 and slightly lower at 8.7% for Class 2, demonstrating a solid balance in detecting both classes without considerable bias. Finally, the TNN scored high TPRs of 91.0% in both classes, demonstrating the model's ability to detect true cases accurately. The FNRs are maintained reasonably low, at 9.0% for both classes, which is good. However, the symmetry of these rates may indicate a balanced classification ability across both classes, with no particular bias toward one class.

In the test dataset shown in Fig 13, the FT performs well, with TPR of 97.5% and 99.3% for Class 1 and Class 2, respectively, showing a good ability to identify positive instances. The FNR for both classes is low, at 2.5% and 0.7%, indicating a few cases where the model incorrectly predicts the negative class. Like the training data results, the KNN struggles with a lower TPR of 6.8% for Class 1 and a very low TPR of 15.3% for Class 2 while having highly high FNRs of 93.2% and 84.7%, respectively. This shows again that the model has significant difficulties in categorizing both positive and negative cases. On the other hand, the GNB model had a TPR of 90.3% for Class 1 and 87.9% for Class 2, and FNRs of 9.7% for Class 1 and 12.1% for Class 2, which show that the model can adequately detect and also misses a substantial proportion of positive events. LR reveals a TPR of 85.9% for Class 1, indicating a strong predictive ability, but with a higher FNR of 14.1%, implying a greater tendency to miss

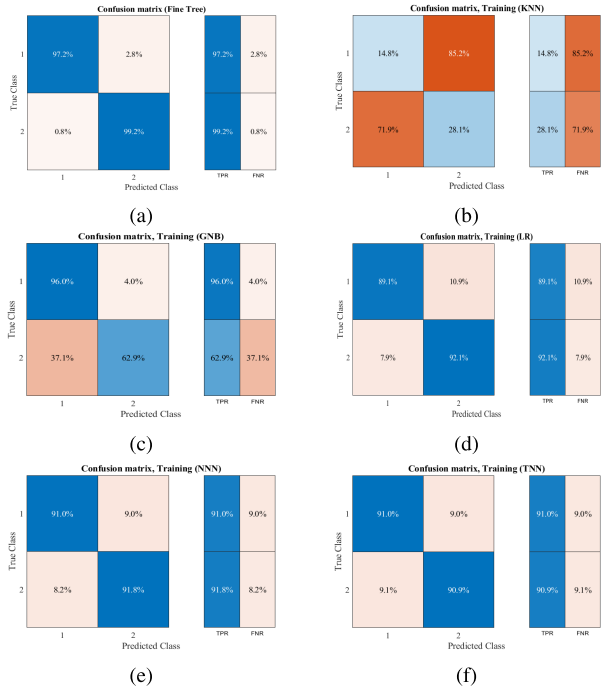


FIGURE 12. Confusion matrices representing the classification performance of machine learning models on the training dataset, showing the True Positive Rate (TPR) and False Negative Rate (FNR) for each class. Models include the (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

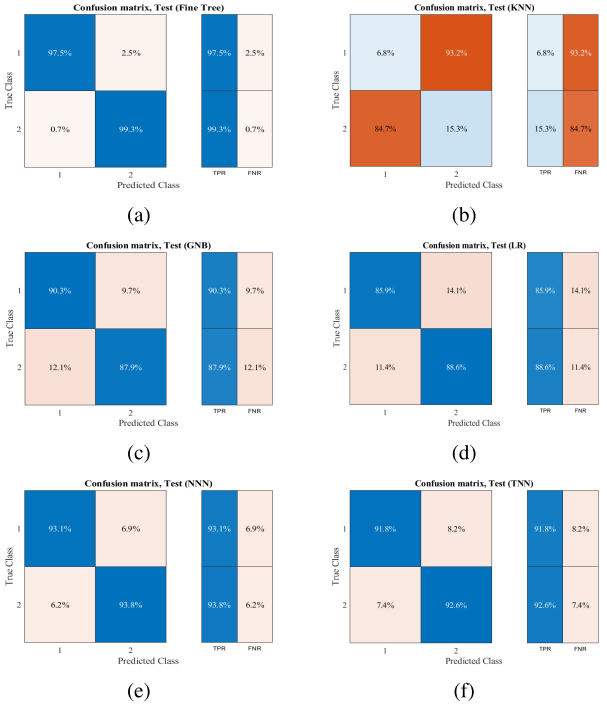


FIGURE 13. Confusion matrices representing the classification performance of machine learning models on the testing dataset, showing the True Positive Rate (TPR) and False Negative Rate (FNR) for each class. Models include the (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

positive Class 1 instances. For Class 2, the TPR is 88.6%, with a corresponding FNR of 11.4%, suggesting the model is better at correctly identifying Class 2 instances than Class 1. The NNN has balanced results, with TPRs of 93.1% for Class 1 and 93.8% for Class 2 and slightly higher FNRs of 6.9% and 6.2% for Class 1 and 2, respectively. The TNN indicates a TPR of 91.8% for Class 1, a slightly lower TPR of 92.6% for Class 2, and FNRs of 8.2% and 7.4%, respectively.

Fig. 14 shows confusion matrices in a percentage format, including Positive Predictive Value (PPV) and False Discovery Rate (FDR) for the training dataset. The FT has an outstanding PPV of 99.2% for Class 1 and 97.3% for Class 2, with extremely low FDRs of 0.8% and 2.7%, respectively. The KNN model has significantly worse performance, The GNB model has a PPV of 72.0% for Class 1 and 94.0% for Class 2, indicating better performance in correctly recognizing Class 2 cases. However, the FDR of 28.0% for Class 1 suggests a high percentage of false detections. Furthermore, the LR model has a high PPV value of 91.9% for Class 1 and 89.5% for Class 2, indicating that it can predict positive cases accurately. Nevertheless, the FDR for Class 2 is 10.5%, implying a significantly greater probability of false positives in Class 2 predictions compared to 8.1% for Class 1. The NNN model has a slightly better PPV for Class 1. The FDR rates are somewhat lower for Class 1. The results from TNN show that the model is equally capable of predicting both classes with the same degree of precision and misclassification rates.

Fig. 15 shows the similar confusion matrices as Fig. 14 for the testing dataset. The FT model performs well in the test dataset as well, with a PPV of 99.3% for Class 1 and 97.4% for Class 2, and very low FDR rates of 0.7% for Class 1 and 2.6% for Class 2, KNN, like earlier results, receives a significantly lower PPV of 7.6% for Class 1 and 13.8% for Class 2 and high FDRs of 92.4% and 86.2%, respectively. Additionally, the GNB model shows a higher PPV for Class 2 at 89.3% compared to 85.5% for Class 1. The FDR for GNB is 11.5% for Class 1 and 10.2% for Class 2, showing that it performs better at Class 2 predictions while producing a moderate number of false detections. LR has a lower PPV of 88.6% for Class 1 and 85.9% for Class 2, with FDRs of 11.4% for Class 1 and 14.1% for Class 2. This suggests that, while LR is effective, it is more likely to produce false positives, especially for Class 2 predictions. The confusion matrix of NNN demonstrates a PPV of 93.9% for Class 1 and 93.0% for Class 2, demonstrating an ability for RFI detection but with a little higher FDR for Class 2 (7%) than Class 1 (6.1%). TNN results exhibit comparable performance and indicate a slightly higher misclassification rate for predicting Class 2.

Fig. 16 displays the Receiver Operating Characteristic (ROC) curves for the models in the test dataset. The Fine Tree has a remarkable Area Under the Curve (AUC) of 0.9923, indicating excellent classification performance. KNN has a lower AUC of 0.1105 than the other models, implying this dataset's worst classification performance. GNB achieves an

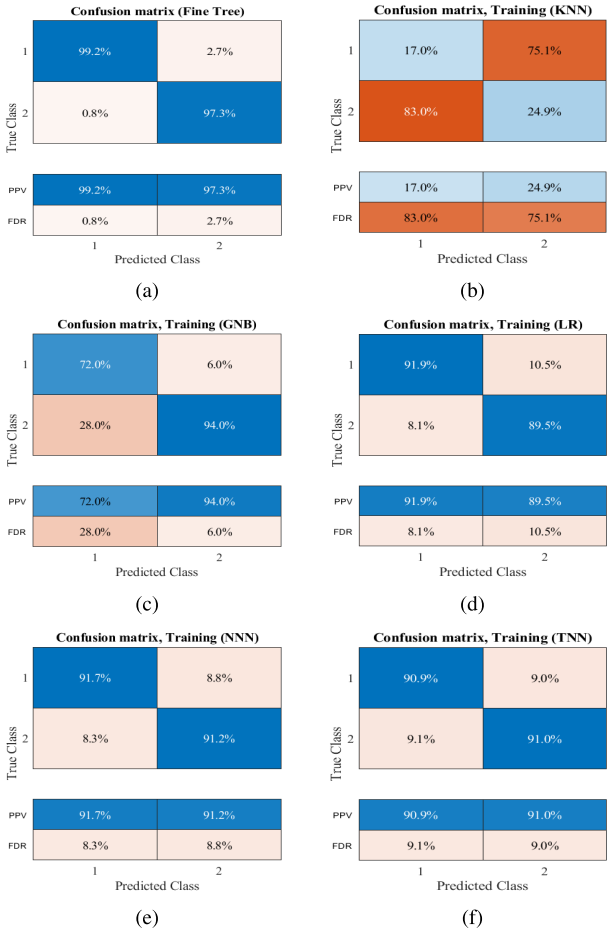


FIGURE 14. Confusion matrices representing the classification performance of machine learning models on the training dataset, showing the Positive Predictive Value (PPV) and False Discovery Rate (FDR) for each class. Models include the (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

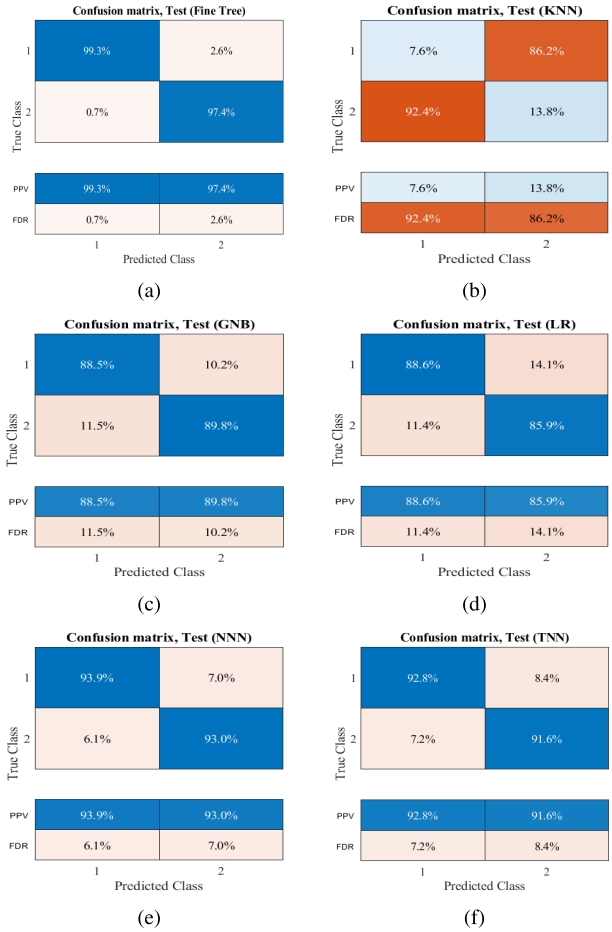


FIGURE 15. Confusion matrices representing the classification performance of machine learning models on the testing dataset, showing the Positive Predictive Value (PPV) and False Discovery Rate (FDR) for each class. Models include the (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

AUC of 0.947, showing that it is a competent classifier. The LR model, with an AUC of 0.955, shows strong predictive performance despite being slightly lower than NN. The NNN also has a high AUC of 0.973, the TNN has a robust AUC of 0.959.

B. RESULTS OF ALTITUDE ESTIMATION

Once the 5G-RFI is detected in the radar signals, the interfered signals proceed to a second stage involving regression ML models, which use the I/Q signals from radar and the additional features to estimate the radar altitudes. We calculate altitudes by dividing the sampled FMCW signal recording into down and up frequency sweeps, creating unique models for each direction, and then averaging the results of the radar signals' up and down sweeps. This approach ensures models are finely tuned to the specific characteristics of each sweep direction. Note that the results for not using ML algorithms for mitigation are only briefly listed in the following sections since they are unacceptable for navigational functions ($>$ 500 meters in MAE). The purpose is to examine the improvements in altitude estimations through ML processing. Again, the training and testing data sets comprise RFI levels from low to extreme that are randomly mixed.

1) UP-SWEEP CASES

Tables 5 and 6 compare the performance of altitude estimations for training and testing datasets, respectively. The units for RMSE and MAE are in meters; the units for MSE are m^2, and R^2 and Corr are unitless metrics. The Fine Tree (FT) model has a significant advantage with a validation RMSE of 39.0623 m and a slightly higher R-squared of 0.9909 over the Medium Tree (MT). The Coarse Tree (CT) model's validation RMSE of 40.2454 shows a balance between model simplicity and accuracy in variance capture. The Bagged Tree model shows how ensemble methods may improve prediction, as indicated by its excellent stability (validation RMSE = 24.6151 m, test RMSE = 24.1991 m). It shows the benefit of aggregating multiple models to minimize variance and increase generalization.

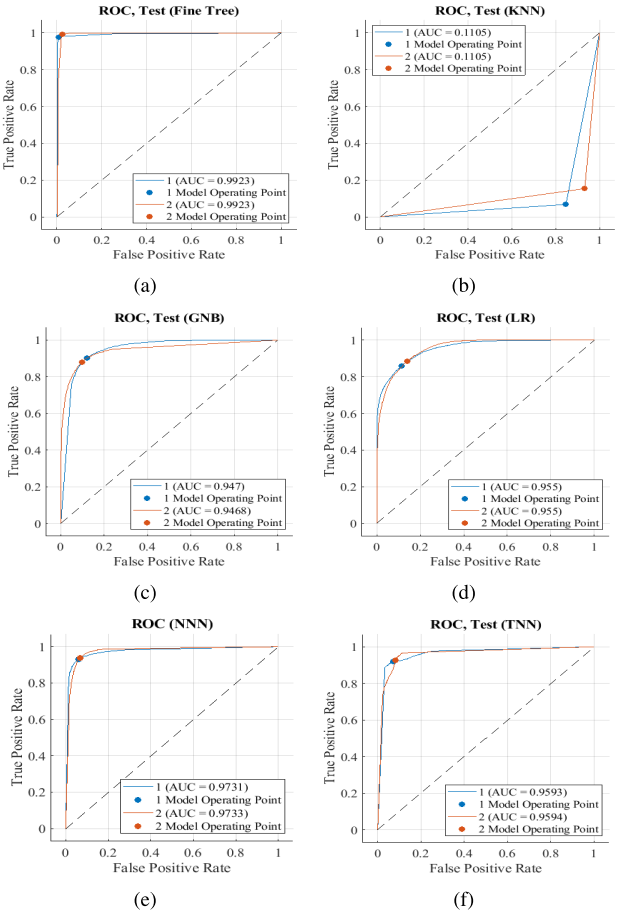


FIGURE 16. Receiver Operating Characteristic (ROC) curves for various ML algorithms evaluated with the testing dataset, with Area Under the Curve (AUC) metrics. Models displayed include the (a) Fine Tree, (b) Fine KNN, (c) Gaussian Naive Bayes, (d) Logistic Regression, (e) Narrow Neural Network, and (f) Trilayered Neural Network.

Linear Regression (LR)'s advantage of having a clear model structure is offset by its relatively high RMSE. The Wide Neural Network (WNN), with lower RMSE scores than the Narrow Neural Network (NNN), demonstrates the benefits of employing many neurons in the network for better variance modeling. The Medium Neural Network (MNN) achieves an excellent mix of depth and variance, while the Tri-layered Neural Network (TriNN) outperforms the Bi-layered Neural Network (BiNN). The Boosted Tree, with a larger RMSE of 49.9384 m, still presents a solid R-squared of 0.9851and shows its ability to represent variance at the expense of more complexity.

Fig. 17 shows scatter density plots with color gradients for the testing dataset based on up-sweep signals. The Bagged Tree (BT) model's graphic displays dense clusters along the line of equality, with minor deviations reflected by cooler colors. Again, the FT plot shows a similar behavior. The NNN and TNN scatter plots show larger color density distributions. While the major concentration remains along the line of equality.

TABLE 5. ML models training performance for altitude prediction from up sweep radar signals with 5G RFI.

Model	RMSE	MSE	$R\hat{2}$	MAE	Corr	Train time (sec)
F. Tree	39.1	1526	0.991	18.0	1.000	10855.50
M. Tree	39.4	1551	0.991	18.9	0.999	3744.24
C. Tree	40.2	1620	0.990	20.2	0.998	501.64
Boosted T.	49.9	2494	0.985	42.6	0.998	488.24
Bagged T.	24.6	606	0.996	12.6	1.000	2318.00
LR	104.8	10982	0.935	74.7	0.989	33.40
NNN	54.4	2956	0.982	40.7	0.995	5036.57
MNN	47.0	2211	0.987	35.3	0.999	9698.42
WNN	43.9	1934	0.989	32.2	0.998	16737.86
BiNN	52.3	2739	0.984	39.2	0.997	8713.75
TriNN	43.6	1900	0.989	32.5	0.998	11904.60

TABLE 6. ML models testing performance for altitude prediction from up sweep radar signals with 5G RFI.

Model	RMSE	MSE	$R\hat{2}$	MAE	Corr
F. Tree	40.4	1635	0.990	18.1	0.995
M. Tree	37.3	1386	0.992	17.5	0.996
C. Tree	39.4	1596	0.991	19.7	0.995
Boosted T.	50.3	2530	0.985	42.4	0.997
Bagged T.	24.2	586	0.997	12.1	0.998
LR	96.7	9349	0.946	70.3	0.973
NNN	50.6	2559	0.985	38.4	0.993
MNN	42.5	1810	0.990	31.3	0.995
WNN	41.4	1712	0.990	31.0	0.995
BiNN	47.6	2266	0.987	35.5	0.993
TriNN	40.2	1613	0.991	29.1	0.995

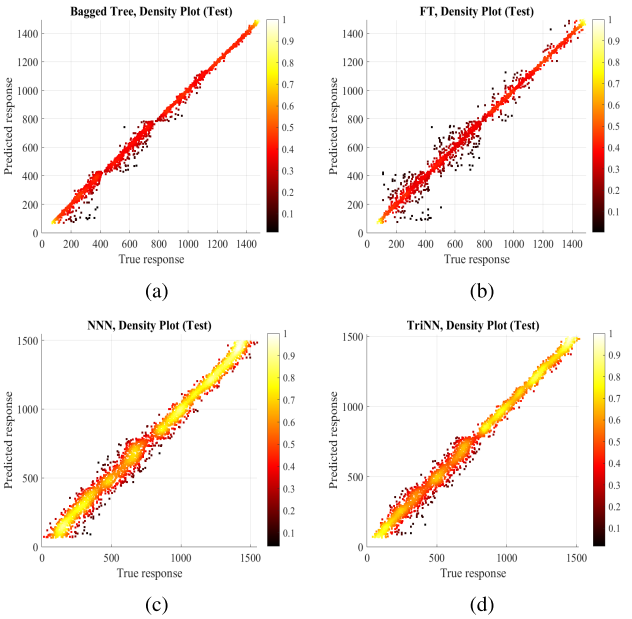


FIGURE 17. Comparative scatter plots with color gradients for Altitude Prediction on up-sweep Test data of the following models: (a) Bagged Tree, (b) Fine Tree, (c) Narrow neural network, and (d) TriLayred neural network.

2) DOWN SWEEP CASE

Tables 7 and 8 evaluate the altitude estimation performance based on down-sweep radar signals. The Fine Tree (FT) model has a notable prediction accuracy with an RMSE of 94.3654 m during validation, suggesting a small decrease in accuracy compared to its up-sweep performance. This trend is also reflected in its test results, with an RMSE

TABLE 7. Training performance of ML models for altitude prediction from down sweep radar signals with 5G RFI.

Model	RMSE	MSE	$R\hat{2}$	MAE	Corr
Fine Tree	94.4	8904.8	0.947	70.5	0.958
Medium Tree	87.8	7715.3	0.954	66.2	0.967
Coarse Tree	81.6	6654.6	0.960	61.9	0.969
Boosted Tree	83.1	6911.4	0.959	65.2	0.976
Bagged Tree	68.4	4674.7	0.972	52.3	0.976
LR	161.9	26237.5	0.844	114.5	0.916
NNN	83.9	7047.6	0.958	64.2	0.943
MNN	75.4	5685.9	0.966	57.3	0.955
WNN	72.9	5312.6	0.968	54.3	0.954
BiNN	81.2	6593.8	0.961	61.8	0.967
TriNN	63.5	4031.7	0.976	47.6	0.957

TABLE 8. Testing performance of ML models for aircraft altitude estimation from down sweep radar signals with 5G RFI.

Model	RMSE	MSE	$R\hat{2}$	MAE	Corr
Fine Tree	90.2	8140.4	0.953	66.9	0.960
Medium Tree	86.2	7426.5	0.957	64.7	0.969
Coarse Tree	79.5	6314.7	0.963	60.4	0.970
Boosted Tree	83.9	7050.0	0.959	65.5	0.977
Bagged Tree	67.2	4509.5	0.974	51.3	0.977
LR	152.8	23357.2	0.865	109.6	0.920
NNN	75.2	5658.2	0.967	56.8	0.945
MNN	71.1	5050.0	0.971	54.7	0.956
WNN	68.9	4754.7	0.972	52.7	0.955
BiNN	98.3	9655.9	0.944	76.1	0.966
TriNN	60.5	3658.3	0.979	45.8	0.958

of 90.2243 m, which, though greater than the up-sweep analysis, still demonstrates significant improvement from non-mitigation results. The Medium and Coarse Tree models support the trend that greater simplicity in the tree model's structure leads to improved management of the down sweep signal variance. This is especially obvious in the Coarse Tree's higher performance on test data, which has an RMSE of 79.4652 m. The Bagged Tree (BT) model outperforms its peers, and Linear Regression (LR) reveals its decreased effectiveness. This issue becomes more apparent when comparing its performance on down and up sweeps. The NNN achieved an RMSE of 83.9502 on validation, indicating a noticeable but insignificant decrease from its up-sweep analysis. The WNN and the TriNN both perform excellently consistently.

Fig. 18 depicts scatter plots with color gradients from the testing dataset, demonstrating the predictive ability of four models when applied to down-sweep radar signals. The Bagged Tree model has a dense region of warm colors that matches the line of perfect prediction, indicating a good connection between actual and predicted responses. This plot shows that the model is good at generalizing to new, unseen data. The FT model produces a slightly more dispersed pattern, particularly at higher response values. Nevertheless, most data points are near the ideal prediction line, consistent with their high-performance scores. Similar to the previous figures, the LR plot exhibits a broader dispersion of data points as we move along the response values. The spreading of warmer colors into cooler ones, particularly in the lower and higher ends of the response spectrum, indicates the

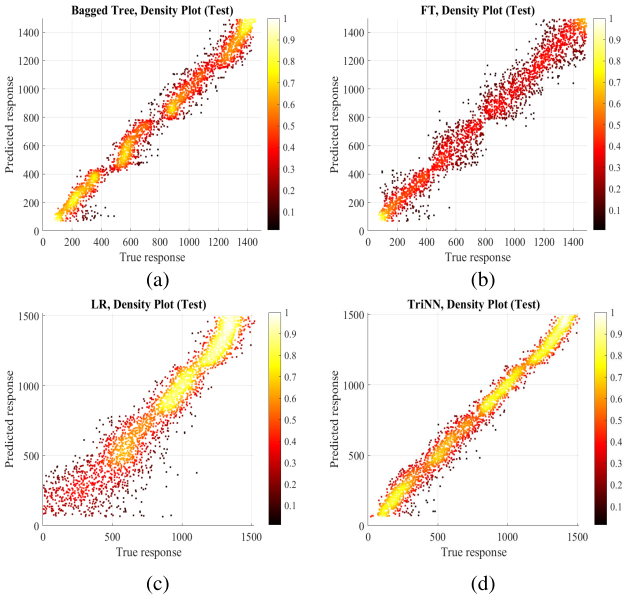


FIGURE 18. Comparative scatter plots with color gradients for Altitude Prediction on down-sweep Test data of the following models: (a) Bagged Tree, (b) Fine Tree, (c) Linear regression, and (d) TriLayred neural network.

model's poorer accuracy, consistent with the larger RMSE and lower R-squared values from the results table. Finally, the TriNN produces a dense line of high-density predictions nearly matching the ideal diagonal. This pattern demonstrates the model's ability to capture the complex relationships in the data, resulting in a strong predictive performance as supported by its test metrics.

The summary of altitude estimation performance is provided in Table 9. It shows the performance of altitude estimation utilizing down-sweep and up-sweep and their average estimation compared to not applying any ML-based RFI mitigation techniques. For down sweep predictions using the TriNN model, validation metrics show a robust R-squared value of 0.976 and an MSE of 4031.706, indicating a strong model fit to the data without overfitting. This performance carries into the testing phase with an R-squared of 0.9788. In contrast, the best up-sweep model (Bagged Trees) outperforms the down-sweep with a notably higher R-squared of 0.9964 in validation and 0.9966 in testing, along with lower RMSE values. The averaged altitude estimations based on optimal up and down sweep estimations produce the most balanced results during validation, with an R-squared of 0.99762 and the lowest MSE (398.2082) among the setups. In the testing phase, the average approach maintains high accuracy with an R-squared of 0.9907. However, it does not reach the accuracy of the best up-sweep predictions alone, evidenced by an RMSE of 40.0318. While this is an improvement over the down sweep's RMSE of 60.4839, averaging the results of multiple models may slightly reduce some of the model's strengths. The last row of the table, which represents altitude estimations derived directly from signals with 5G interference without applying any ML methods, shows severe performance degradation. The RMSE increases

TABLE 9. Summary of performance of aircraft altitude estimation under 5G RFI.

Training Results					
Model	**RMSE**	**MSE**	**R2̂**	**MAE**	**Corr**
Best Down (TriNN)	63.5	4032	0.976	47.6	0.957
Best Up (Bagged)	24.6	606	0.996	12.5	1.000
Mean	19.9	398	0.998	14.2	0.999
With Interf.	821.3	674606	-3.02	553.8	0.440
Testing Results					
Model	**RMSE**	**MSE**	**R2̂**	**MAE**	**Corr**
Best Down (TriNN)	60.5	3658	0.979	45.8	0.958
Best Up (Bagged)	24.2	586	0.997	12.1	0.998
Mean	40.0	1603	0.991	29.5	0.995
With Interf.	808.1	653004	-2.80	552.2	0.445

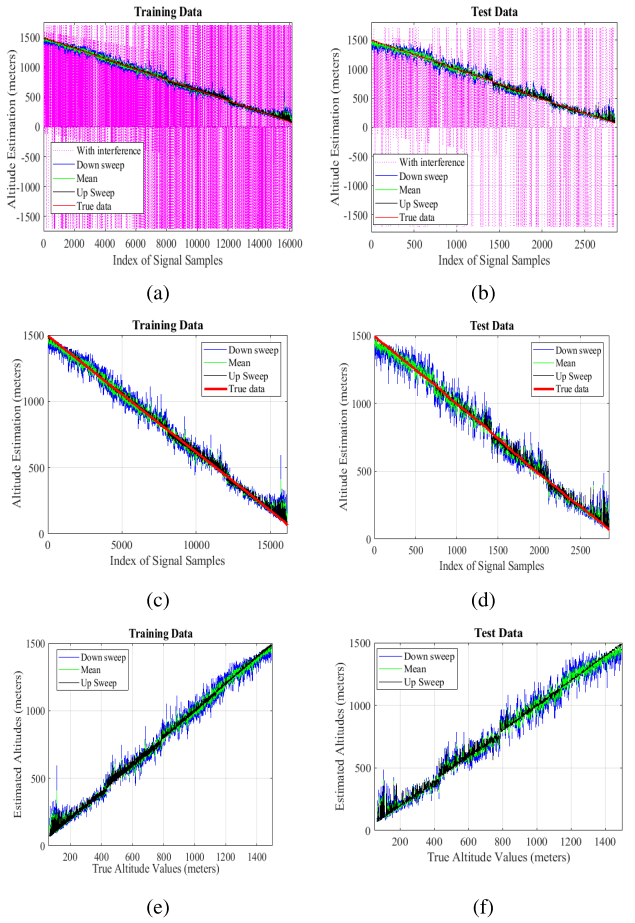


FIGURE 19. Comparative analysis of altitude estimations and actual vs. predicted altitudes: (a) Training dataset including the impact of 5G interference on altitude predictions, (b) Testing dataset including the impact of 5G interference on altitude predictions, (c) Training dataset without the impact of 5G interference, (d) Testing dataset without the impact of 5G interference, (e) Actual vs. predicted altitudes from downs-weep, upsweep, and their mean for the training dataset without the impact of 5G interference, and (f) Actual vs. predicted altitudes from down-sweep, upsweep, and their mean for the testing dataset without the impact of 5G interference.

to 808.1 while the R-squared falls to -2.8, emphasizing the significant impact of 5G interference on measurement accuracy.

Fig. 19 presents altitude profiles for both the training and testing datasets, offering a detailed examination of model performance in altitude estimation using the altimeter signals with 5G RFI, up sweep signals, down sweep signals, and their average. The plots show that the up-sweep predictions align best with the truth data, as indicated by the red line, especially at higher altitudes. However, some variability in predictions emerges at lower altitudes. On the other hand, the down-sweep predictions display a slightly reduced accuracy, with notable variations across different altitudes and some peaks at lower altitudes. The averaged predictions further improved the accuracy, suggesting that integrating up and down-sweep predictions leads to a more precise estimation. For the result plots based on testing datasets, although there is a slight increase in the variation of altitude predictions across all methods compared to the training dataset, the overall performance closely aligns with the training results. The up-sweep method achieves the best performance with the least variation in altitude predictions, while the down-sweep shows increased variations across different altitudes. The averaged predictions maintain a balance between the down-sweep and the up-sweep estimations. These observations suggest that while the down-sweep and average predictions offer a smoother predictive trend, they might not entirely address the overfitting or specific biases in the testing data environment. On the other hand, 5G interference significantly impacts altitude measurements, as seen in the pink-colored traces in training and testing datasets that consistently deviate from the true data.

V. CONCLUSION

This study addresses the issue of 5G base station interference with radar altimeters during approaching flights. A two-step machine learning (ML) approach is developed to detect RF interference and mitigate the impact of interference on altitude estimations. The method begins by classifying altimeter signals to distinguish between the ''pure'' and interfered states using computationally effective ML algorithms, then uses various regression models for altitude estimation. The study utilizes three types of data: real 5G signals from a Norman, Oklahoma base station, emulated FMCW radar altimeter signals during an airplane's landing process, and a combination of them to simulate radar altimeter signals with 5G interference. A novel approach estimates altitudes by models trained separately from up-sweep and down-sweep radar signals and then averages the estimations from these models. The performance results validate the two-step and dual-model approach: The FT model achieved 98.3596% accuracy in classification. In altitude prediction, the BT model applied to up-sweep signals demonstrated a correlation coefficient 0.9983. The best results for down-sweep signals were obtained with TriNN, achieving a correlation coefficient 0.9576. Future work will incorporate real-world testing to support the findings and evaluate model performance under actual conditions. The scalability of the models will be explored in more complex environments by including additional environmental factors and larger datasets. A detailed analysis of the energy consumption of the machine learning models will be conducted to optimize resource usage.

Additionally, ethical and privacy concerns will be addressed by developing robust data privacy protocols and ethical standards in machine learning for radar altimetry.

ACKNOWLEDGMENT

The authors would like to thank Dr. Dexiang (John) Xu and David Bate from the Navy Information Warfare Center (NIWC) Atlantic for their support, encouragement, and feedback, support and inspiration provided by Robert Loveland and the team from FAA tech-op AJW-124, as well as the collaborations with industry partners from the Radio Technical Commission for Aeronautics (RTCA) through special committee 230.

An earlier version of this paper was presented in part at the SPIE Conference, [DOI: 10.1117/12.3013749].

REFERENCES

[1] *Sharing Between the Earth Exploration-Satellite (Passive) and Airborne Altimeters in the Aeronautical Radionavigation Service in the Band 4 200–4 400 MHz*, document ITU Rec. ITU-R RS. 1624, 2003.

[2] J. Benveniste, "Radar altimetry: Past, present and future," in *Coastal Altimetry*. Berlin, Germany: Springer, 2011, pp. 1–17.

[3] R. Coleman, *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*. Amsterdam, The Netherlands: Elsevier, 2001.

[4] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[5] A. A. Amaireh and Y. R. Zhang, "Investigation of potential 5G RF interference with C-band radar operations and mitigation solutions," *Proc. SPIE*, vol. 12535, pp. 206–217, Jun. 2023.

[6] *Assessment of C-Band Mobile Telecommunications Interference Impact on Low Range Radar Altimeter Operations*, RTCA, Washington, DC, USA, 2020.

[7] Bench Electronics. (2023). *Filtering the Way to Safer 5G Near Airports*. [Online]. Available: https://www.bench.com/setting-the-benchmark/filtering-the-way-to-safer-5g-near-airports

[8] C. S. Kyriakos, "Tracking filter for radio altimeter," U.S. Patent 4 427 981 A, Jan. 24, 1984.

[9] W. Jiang, Y. Wang, Y. Li, Y. Lin, and W. Shen, "Radar target characterization and deep learning in radar automatic target recognition: A review," *Remote Sens.*, vol. 15, no. 15, p. 3742, Jul. 2023.

[10] X. Cai, M. Giallorenzo, and K. Sarabandi, "Machine learning-based target classification for MMW radar in autonomous driving," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 4, pp. 678–689, Dec. 2021.

[11] P. Lang, X. Fu, M. Martorella, J. Dong, R. Qin, X. Meng, and M. Xie, "A comprehensive survey of machine learning applied to radar signal processing," 2020, *arXiv:2009.13702*.

[12] A. Amaireh, Y. R. Zhang, D. J. Xu, and D. Bate, "Improved investigation of electromagnetic compatibility between radar sensors and 5G-NR radios," *Proc. SPIE*, vol. 13048, pp. 85–98, Jun. 2024.

[13] B. Atayants, V. Davydochkin, V. Ezerskiy, V. Parshin, and S. Smolskiy, *Precision FMCW Short-Range Radar for Industrial Applications*. Boston, MA, USA: Artech House, 2014.

[14] S. Houidi, D. Fourer, and F. Auger, "On the use of concentrated time–frequency representations as input to a deep convolutional neural network: Application to non intrusive load monitoring," *Entropy*, vol. 22, no. 9, p. 911, Aug. 2020.

[15] M. J. Blanca, J. Arnau, D. López-Montiel, R. Bono, and R. Bendayan, "Skewness and kurtosis in real data samples," *Methodology*, vol. 9, no. 2, pp. 78–84, May 2013.

[16] E. Schubert, J. Wolfe, and A. Tarnopolsky, "Spectral centroid and timbre in complex, multiple instrumental textures," in *Proc. Int. Conf. Music Perception Cognition*, 2004, pp. 112–116.

[17] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 314–323, Feb. 1988.

[18] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Apr. 1997, pp. 1331–1334.

[19] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," *CUIDADO 1st Project Rep.*, vol. 54, pp. 1–25, Apr. 2004.

[20] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004.

[21] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, vol. 452. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

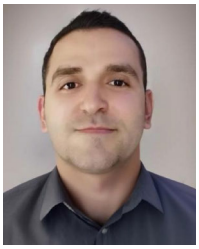
[22] Y. Al-Issa and A. M. Alqudah, "A lightweight hybrid deep learning system for cardiac valvular disease classification," *Sci. Rep.*, vol. 12, no. 1, p. 14297, Aug. 2022.

[23] MathWorks. (2024). *Misclassification Costs in Classification Learner App*. Accessed: Jun. 19, 2024. [Online]. Available: https://www.mathworks.com/help/stats/misclassification-costs-in-classification-learner-app.html

[24] M. A. Al-Hashem, A. M. Alqudah, and Q. Qananwah, "Performance evaluation of different machine learning classification algorithms for disease diagnosis," *Int. J. E-Health Med. Commun.*, vol. 12, no. 6, pp. 1–28, Aug. 2021.

[25] A. Amaireh, Y. Zhang, and P. W. Chan, "Atmospheric humidity estimation from wind profiler radar using a cascaded machine learning approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6352–6371, 2023.

[26] A. Amaireh, Y. R. Zhang, P. W. Chan, and D. Zrnic, "A novel approach for improving cloud liquid water content profiling with machine learning," *SSRN Electron. J.*, Jan. 2023, doi: 10.2139/ssrn.4744644. [Online]. Available: https://ssrn.com/abstract=4744644



ANAS AMAIREH received the bachelor's and master's degrees in communication engineering from Yarmouk University, Jordan, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the University of Oklahoma (OU), Norman. He is also a Research and Teaching Assistant with the OU's Advanced Radar Research Center. He specializes in radar, machine learning in radar, radar signal processing, 5G communication, antenna arrays, and metaheuristic algorithms. His research interests include leveraging machine learning to enhance the functionality and performance of modern radar and communication systems.

YAN (ROCKEE) ZHANG (Senior Member, IEEE) received the Ph.D. degree from the University of Nebraska, in 2004. He is currently a Presidential Professor with the School of Electrical and Computer Engineering, University of Oklahoma (OU). He was one of the founding faculty members of OU's Advanced Radar Research Center (ARRC), served as the Technical Lead of the ARRC's multi-functional phased array radar development, from 2008 to 2011. He is also the Faculty Leader of the Intelligent Aerospace Radar Team (IART). He has extensive research in ultra-wideband (UWB) radar and applications to airborne remote sensing. He is the principal investigator for the OU's polarimetric airborne radar operating at X-band (PARADOX) development and leading the deployment missions related to radar and radios supporting advanced air mobility (AAM), autonomous vehicle systems, and air-surveillance services. He is one of the Faculty Fellow Members of the Cooperative Institute for Severe and High Impact Weather Research and Operations (CIWRO), OU; and a Representative of OU at multiple RTCA special committees supporting industry standard developments for avionics.