

Towards an Interpretable AI Framework for Advanced Classification of Unmanned Aerial Vehicles (UAVs)

Ekramul Haque¹, Kamrul Hasan¹, Imtiaz Ahmed², Md. Sahabul Alam³, Tariqul Islam⁴

¹ Tennessee State University, Nashville, TN, USA

² Howard University, Washington, DC, USA

³ California State University, Northridge, CA, USA

⁴ Syracuse University, Syracuse, NY, USA

Email: {ehaque1, mhasan1}@tnstate.edu, imtiaz.ahmed@howard.edu, md-sahabul.alam@csun.edu, mtislam@syr.edu

Abstract—With UAVs on the rise, accurate detection and identification are crucial. Traditional unmanned aerial vehicle (UAV) identification systems involve opaque decision-making, restricting their usability. This research introduces an RF-based Deep Learning (DL) framework for drone recognition and identification. We use cutting-edge eXplainable Artificial Intelligence (XAI) tools, SHapley Additive Explanations (SHAP), and Local Interpretable Model-agnostic Explanations (LIME). Our deep learning model uses these methods for accurate, transparent, and interpretable airspace security. With 84.59% accuracy, our deep-learning algorithms detect drone signals from RF noise. Most crucially, SHAP and LIME improve UAV detection. Detailed explanations show the model's identification decision-making process. This transparency and interpretability set our system apart. The accurate, transparent, and user-trustworthy model improves airspace security.

Index Terms—Drone Detection, RF Signals, Deep Learning, SHAP, LIME, Explainable AI, Airspace Security

I. INTRODUCTION

Several studies have been conducted on small UAV detection and classification systems, each with pros and cons. In their work on UAV classification, Kim et al. demonstrated the usage of a Convolutional Neural Network (CNN) with Doppler images [1], and Choi and Oh are carrying this research forward in 2019 by taking micro-Doppler signatures and spectrogram images into account [2]. Using a deep belief network, the spectral correlation functions of three drones were classified in [3]. Despite their strong detection and categorization skills, these models sometimes lack transparency in their decision-making methods, raising accountability concerns.

This paper introduces deep learning with interpretability for drone identification. We use eXplainable Artificial Intelligence (XAI) and deep learning to create a precision and transparent aircraft security system. This study highlights our methodology's unique contributions and places it in the context of airspace security advances.

II. SYSTEM ARCHITECTURE

A. Interpretable AI-Driven Framework Description

Figure 1 shows the AI-driven UAV categorization system architecture. It uses explainable ML methods like PCA, drones,

LIME [4], and SHAP [5] to boost DL model credibility. Start by transforming RF data from time to frequency domain using DFT. The frequency components and power spectra are the inputs for DNN training. PCA decreases dimensionality and redundancy to simplify datasets. The revised dataset trains the DNN model to classify drones into four groups.

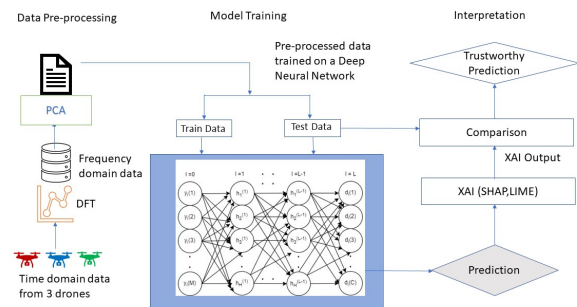


Fig. 1. Illustration of the proposed Interpretable AI Framework for Advanced Classification of UAVs.

We use SHAP, and LIME XAI approaches for transparency, clear decision-making, and user trust in model classifications.

B. Analysis of the Data set

We used the ubiquitous DroneRF dataset [6], which contains RF signals from three drones. CSV segmentation and storage of the DroneRF dataset eliminates memory overflow and maintains program compatibility. The hardest part was converting time-domain data to frequency-domain data for analysis. This conversion used DFT. Our deep learning model used power spectra frequency components.

C. Model Architecture

A multi-class classifier designed with DNNs detects and identifies intrusive drones. The proposed model can detect drones and differentiate among the RF spectra of distinct drones. A DNN comprises an input layer, different hidden layers, and an output layer.

In this study, a DNN is trained and evaluated using the developed RF database to detect the presence of a drone and determine its class.

III. EXPERIMENTAL ANALYSIS

Figure 2 displays the confusion matrix for the DNN's drone classification performance in the test dataset. The matrix displays the model's accurate drone class prediction. The model accurately predicted No Drone cases 4085, Bebop Drones 8241, AR Drones 6000, and Phantom Drones 877 times.

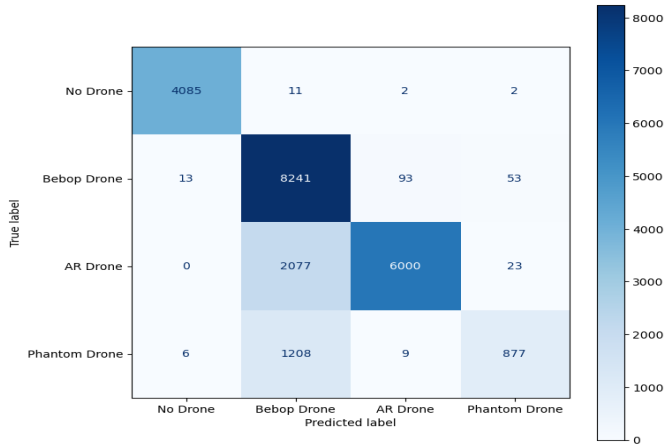


Fig. 2. Confusion Matrix for Drone Classification

SHAP local explanations explain a single instance's feature values that affect positive and negative decisions. Figure 3 shows a local explanation with a 0.66 probability of No Drone output, along with features and their values: Feature 7, 435, 414, 613, 47, 576, and 0.

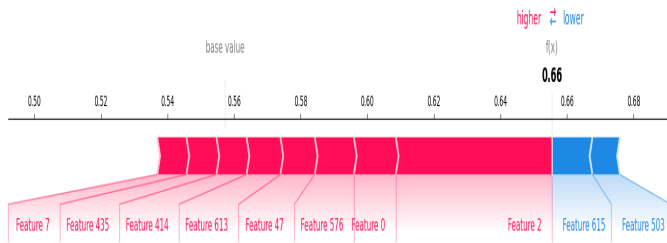


Fig. 3. SHAP force plot used for local explanations

Features that push the prognosis higher are displayed in red, while those that push the prediction lower are displayed in blue. Note that this explanation varies depending on the input instance.

A. Interpreting Predictions with LIME

LIME produces regional explanations. Figure 4 shows how to classify results as No Drone, Bebop, AR, or Phantom,

along with probability and original instance values. We presented two examples. The classifier only shows Bebop and NOT Bebop because it is most important. Colors signify character classes. Blue attributes are Not Bebop, while orange attributes are Bebop.

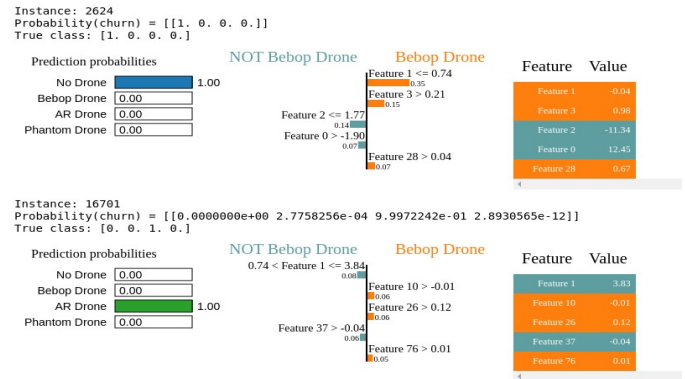


Fig. 4. Explaining individual prediction of deep learning classifier using LIME.

We can see distinct characteristics justify different classifications. In instance 2624, Features 2 and 3 contribute to the prediction that there is no drone, whereas in instance 19701, features 1 and 37 contribute to the prediction that there is an AR Drone.

CONCLUSION

By combining Deep Learning (DL) and Radio Frequency (RF) technologies, this study improves UAV identification and airspace security. Unlike conventional methodologies, our model's revolutionary usage of SHapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) fulfills the need for transparency in AI decision-making. The model excels in performance and provides clear and reliable explanations for its conclusions, recognizing drone signals with 84.59% accuracy. This makes airborne security systems more reliable and sets a precedent for AI-driven security technology.

REFERENCES

- [1] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged doppler images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 38–42, 2016.
- [2] B. Choi and D. Oh, "Classification of drone type using deep convolutional neural networks based on micro-doppler simulation," in *2018 International Symposium on Antennas and Propagation (ISAP)*, pp. 1–2, IEEE, 2018.
- [3] C. Aker and S. Kalkan, "Using deep networks for drone detection," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2017.
- [4] M. T. Ribeiro, S. Singh, and C. Guestrin, "why should I trust you?": Explaining the predictions of any classifier," *CoRR*, vol. abs/1602.04938, 2016.
- [5] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017.
- [6] M. S. Allahham, M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khatib, and A. Erbad, "Dronerf dataset: A dataset of drones for rf-based detection, classification and identification," *Data in brief*, vol. 26, p. 104313, 2019.