Accurate and Reliable Methods for 5G UAV Jamming Identification With Calibrated Uncertainty

Hamed Farkhari[1,2,*,†], Joseanne Viana[2,3,*,†], Pedro Sebastiao[2,3], Luis Bernardo[3,4], Sarang Kahvazadeh[5] and Rui Dinis[3,4]

[1]PDMFC, Rua Fradesso da Silveira, n. 4, Piso 1B, 1300-609, Lisboa, Portugal
[2]ISCTE – Instituto Universitario de Lisboa, Av. das Forcas Armadas, 1649-026 Lisbon, Portugal
[3]IT – Instituto de Telecomunicacoes, Av. Rovisco Pais, 1, Torre Norte, Piso 10, 1049-001 Lisboa, Portugal
[4]FCT – Universidade Nova de Lisboa, Monte da Caparica, 2829-516 Caparica, Portugal;
[5]CTTC - Centre Tecnologic de Telecomunicacions de Catalunya (CERCA)

Abstract

This research highlights the negative impact of ignoring uncertainty on DNN decision-making and Reliability. Proposed combined preprocessing and post-processing methods enhance DNN accuracy and Reliability in time-series binary classification for 5G UAV security dataset, employing ML algorithms and confidence values. Several metrics are used to evaluate the proposed hybrid algorithms. The study emphasizes the XGB classifier's unreliability and suggests the proposed methods' potential superiority over the DNN softmax layer. Furthermore, improved uncertainty calibration based on the Reliability Score metric minimizes the difference between Mean Confidence and Accuracy, enhancing accuracy and Reliability.

Keywords

Unmanned Aerial Vehicle, Deep Neural Networks, Calibration, Uncertainty, Reliability, Jamming Identification, 5G, 6G

1. Introduction

Deep Neural Networks (DNNs) have seen extensive deployment due to their recent achievements in several fields. Prediction distributions generated by such models increasingly make decisions in the telecommunications and security sectors [1, 2, 3].

For example, 6G telecommunication systems will incorporate Machine Learning (ML) mechanisms such as DNNs into their standards [1] and there are several studies on how to apply deep learning decision-making in the physical layer [2]. Another promising field for DNN applications is 5G Unmanned Aerial Vehicle (UAV) security [4, 5]. DNNs are interesting to use due to their universal function capabilities, superior logic that allows them to solve complex

Research Projects Track @ RCIS 2023: The 17th International Conference on Research Challenges in Information Science, May 23–26, 2023, Corfu, Greece
*Corresponding author.
†These authors contributed equally.
Hamed_Farkhari@iscte-iul.pt (H. Farkhari); joseanne_cristina_viana@iscte-iul.pt (J. Viana)
0000-0002-2620-260X (H. Farkhari); 0000-0002-4191-3127 (J. Viana); 0000-0001-7729-4033 (P. Sebastiao); 0000-0002-3384-9997 (L. Bernardo); 0000-0001-5607-8120 (S. Kahvazadeh); 0000-0002-8520-7267 (R. Dinis)

time series modeling issues, and depending on their design, the possibility to process data in parallel. However, due to the DNN's iterative data processing, classification applications can provide probabilities with uncertainties in the outputs which raise concerns about the reliability of the true correctness likelihood of its classification decisions.

The authors in [6] discuss the importance of calibrating DNNs in order to guarantee high accuracy and reliable output decisions. They show at least six calibration techniques that increase both parameters in widely recognized datasets (i.e., CIFAR-10 and ImageNet) applied in pretrained DNNs (i.e., RestNet, WideNet, and LeNet). In [7], the authors justify the need to specify the uncertainty especially in critical real-world settings, in which the input distribution deviates from the training distribution because of sample bias and non-stationarity.

Understanding questions of risk, uncertainty, and trust in a model's output becomes increasingly important when augmented techniques are used at the original data preprocessing stage. The authors in [8] suggest that prepossessing and post-processing techniques can improve N inputs and M class DNNs. The authors in [6, 7] also propose methods that increase accuracy while reducing uncertainty in classification tasks and mathematical approaches to calculate the Expected Calibration Error (ECE), the Maximum Calibration Error (MCE), and estimate if the DNN is over-confident or under-confident.

Inspired by the possibility of choosing a tolerable degree of uncertainty and increasing the reliability of DNN outputs used in 5G UAV security, this study presents several new combined prepossessing and post-processing techniques that increase the overall accuracy and reliability of binary classification deep networks by adjusting the uncertainty. We assess these methods using seven key performance metrics related to errors in calibration and in confidence values. Then, we utilize the Reliability Score (RS) that measures the difference between the Mean Accuracy (MA) and Mean Confidence (MC) to measure the degree of uncertainty. Finally, we evaluate the proposed algorithms' impact on the DNN's performance compared to the baseline DNN with no algorithms applied and the DNN added to the eXtreme Gradient Boosting (XGB) classifier. The XGB classifier is selected because of its superior accuracy in comparison to five other classifiers we test with our data [9].

2. System Model

2.1. Dataset and Methods

Dataset

Our dataset contains data from the Received Signal Strength Indicator (RSSI) and the Signal to Interference-plus-Noise Ratio (SINR) measurements collected when an authenticated UAV is connected in the small cell through the 5G communication system, and there are power attacks from other UAVs in the network. There are other terrestrial users connected to the network. The measured parameters in the authenticated UAV change as the interference from the other devices increases or decreases. More details on the dataset construction and one possible application for the dataset is available in [10] and in [11].

Method 1

We apply this method on the probabilistic outputs of the DNN for all the augmented samples. Each output is in the one hot encoding form for binary classification. For example, $[\alpha, 1-\alpha]$ in which α is a number between zero and one.

Method 2

In Method 2, we convert outputs from a probabilistic to an integer form and apply a majority voting algorithm to them.

Method 3

Method 3 calculates the confidence of each output. The output with the maximum confidence value is selected as the final result.

2.2. Evaluation Metrics

We use well-known metrics proposed by [6] to measure the model's uncertainty, accuracy, and quality to compare method improvements with each other. These metrics are explained below:

Accuracy per Confidence. This metric is used in its visual form to analyze the calibration and uncertainty of the DNN model.

Mean Confidence and Mean Accuracy. These two metrics are the total weighted average of confidence and accuracy for the number of samples per each confidence interval.

Reliability Score. We define the difference between the MC and MA values by another metric which is denominated the Reliability Score.

Expected and Maximum Calibration Errors. At each confidence interval, the accuracy deviation away from the confidence interval center is considered as the error per each interval. The Expected Calibration Error is defined as the weighted error and the Maximum Calibration Error describes the maximum error per all intervals.

Negative LogLikelihood Loss (NLL). This metric is known as cross-entropy loss and is used as a loss function for DNNs [12]. It is also utilized as a metric to measure the quality of the probabilistic model [13].

Brier Score Loss (BSL). This metric is defined by the square error of the predicted probability vector and ground truth values in one hot encoding form.

3. Experimental Results

In this section, we present the simulations results. We compare the results of the DNN using each of the five prospective methods to the results of the DNN with no method and the DNN with the XGB. We choose XGB because it is the best performing publicly available classifier applied to our dataset in terms of accuracy [11]. We use the Accuracy vs the Confidence Intervals Central Values and we evaluate the performance of each algorithm using the seven metrics previously mentioned in subsection 2.2.

DNN+	ECE (%)	MCE (%)	MC (%)	MA (%)	NLL	BSL (%)	OC or UC	RS = |MA-MC| (%)
No Method	3.71	**7.07**	**89.77**	91.01	**0.2**	**6.25**	UC	**1.24**
XGB	7.22	27.77	92.43	85.21	0.63	12.45	OC	7.22
Method 1 + XGB	4.70	**14.08**	91.59	87.74	0.54	10.72	OC	3.85
Method 2 + XGB	**3.21**	36.74	94.41	**91.19**	2.53	8.32	OC	3.21
Method 3	4.03	15.18	**91.84**	**91.19**	**0.21**	**6.53**	OC	**0.65**
Method 1 + 3	**2.19**	**12.37**	**90.26**	**91.19**	**0.22**	**6.82**	UC	**0.92**
Method 2 + 3	**3.02**	41.16	94.20	91.18	2.52	8.21	OC	3.02

ECE; Expected Calibration Error, MCE; Maximum Calibration Error, MC; Mean Confidence, MA; Mean Accuracy, NLL; Normalized Negative Log Likelihood, BSL; Brier Score Loss, OC; Over-Confidence, UC; Under-Confidence, RS; Reliability Score.

Table 1
Key Performance Parameters for Reliability, Top three results for each metric are highlighted

Table 1 shows the metric details used to define the best algorithm performance. It is difficult to choose the best method based on only one metric or consider the best performance on each metric. There is no method that can satisfy the best performance on all metrics. Therefore, we highlight the top three results in each metric. After indicating the top three metric results in Table 1, we notice that the combination of DNN and Methods 1 and 3 (Method 1+3) satisfies the most metrics. Comparison from No method to Method 1+3 shows an almost double increase in MCE, a considerable decrease in ECE, and minor differences in the remaining variables. Therefore, it is more necessary to lower the ECE, although the MCE error will increase. In second place, Method 3 can not only improve the total accuracy, but also satisfies most of the reliability metrics. Furthermore, this method achieves the closest to zero RS results followed closely by Method 1+3 compared to all the suggested algorithms.

The results of the M2+XGB indicates that the XGB can be used as a complementary algorithm to improve the accuracy of the DNN results (MA = 91.19 and ECE = 3.21). Even though the accuracy results of XGB algorithm alone was inferior (MA = 85.21). An comparison of M2+XGB with M1+XGB reveals that using class label outputs instead of probability values to calculate majority voting is recommended when we want to combine the XGB result with DNN.

The results of most of the combined algorithms placed the DNN in the over-confidence region (OC) except for the Method 1+3 algorithm and the DNN with No Methods applied. Both cases were in the under-confidence region (UC). The accuracy results of all the algorithms were similar except for the XGB and the Method 1+XGB. For example, the difference between the highest (M2 + XGB = 91.19) and the lowest values (No Method = 91.01) is 0.18%. However, the difference between both algorithms for the MCE and NLL indicators is 29.67 and 2.33, respectively. These differences highlight the accuracy discrepancies between the confidence interval values and decreases the reliability of the DNN. Therefore, it is fundamental to have DNN reliability evaluation results prior defining best performing architectures.

4. Conclusion

It is expected to have ML mechanisms in 5G and 6G UAV communication systems. Therefore, it is fundamental to understand the uncertainties of the deep networks used in those systems and how reliable they are. In this study, we proposed five combined methods to increase accuracy

and reliability concomitantly in binary classification deep networks applied to UAV security scenarios. By analyzing seven reliability metrics and the accuracy per confidence, Method 1 combined with Method 3 presented the best overall performance that satisfied most of the metrics by achieving the top three in each one. This algorithm reached an ECE of 2.19 and was closer to all ideal levels' values.

Method 3 was the second-best performing algorithm in terms of reliability. With Method 2 + XGB, we showed that a lower performing ML algorithm can be combined with one of the proposed methods to increase the total DNN accuracy, but in terms of the reliability, this might not be a good option.

Finally, four of the five methods presented were able to increase accuracy, but not all of them increased the reliability. As a result, network engineers and developers must take extra precaution when proposing DNN architectures and analyze them in terms of accuracy and reliability.

Acknowledgment

This research received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Project Number 813391. Also, this work was partially supported by Fundação para a Ciência e a Tecnologia and Instituto de Telecomunicações under Project UIDB/50008/2020.

References

[1] X. Lin, An overview of 5g advanced evolution in 3gpp release 18, IEEE Communications Standards Magazine 6 (2022) 77–83.

[2] Z. Qin, H. Ye, G. Y. Li, B.-H. F. Juang, Deep learning in physical layer communications, IEEE Wireless Communications 26 (2019) 93–99.

[3] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity, IEEE Access 6 (2018) 35365–35381.

[4] Y. Li, J. Pawlak, J. Price, K. Al Shamaileh, Q. Niyaz, S. Paheding, V. Devabhaktuni, Jamming detection and classification in ofdm-based uavs via feature- and spectrogram-tailored machine learning, IEEE Access 10 (2022) 16859–16870.

[5] A. Krayani, A. S. Alam, L. Marcenaro, A. Nallanathan, C. Regazzoni, Automatic jamming signal classification in cognitive uav radios, IEEE Transactions on Vehicular Technology (2022) 1–17.

[6] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, On calibration of modern neural networks, Proceedings of the 34th International Conference on Machine Learning - Volume 70 (2017) 1321–1330.

[7] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, J. Snoek, Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift, Proceedings of the 33rd International Conference on Neural Information Processing Systems (2021).

[8] D. Hafner, D. Tran, T. P. Lillicrap, A. Irpan, J. Davidson, Noise contrastive priors for functional uncertainty, in: Conference on Uncertainty in Artificial Intelligence, 2018.

[9] J. Viana, H. Farkhari, P. Sebastiao, L. M. Campos, K. Koutlia, B. Bojovic, S. Lagen, R. Dinis, Deep attention recognition for attack identification in 5g uav scenarios: Novel architecture and end-to-end evaluation, 2023. arXiv:2303.12947.

[10] J. Viana, H. Farkhari, P. Sebastiao, S. Lagen, K. Koutlia, B. Bojovic, R. Dinis, A synthetic dataset for 5g uav attacks based on observable network parameters, 2022. arXiv:2211.09706.

[11] J. Viana, H. Farkhari, L. M. Campos, P. Sebastião, K. Koutlia, S. Lagén, L. Bernardo, R. Dinis, A convolutional attention based deep learning solution for 5g uav network attack recognition over fading channels and interference, in: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), 2022, pp. 1–5. doi:10.1109/VTC2022-Fall57202.2022.10012726.

[12] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer series in statistics, Springer, 2009.

[13] M. P. Naeini, G. F. Cooper, M. Hauskrecht, Obtaining well calibrated probabilities using bayesian binning, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, AAAI Press, 2015, p. 2901–2907.