Machine Learning Approaches for Flow-Based Intrusion Detection Systems

Razan Abdulhammed, Hassan Musafer, Ali Alessa, Miad Faezipour, and Abdelshakour Abuzneid
Department of Computer Science and Engineering
University of Bridgeport, Bridgeport, CT

Abstract

In cybersecurity, machine/deep learning approaches can predict and detect threats before they result in major security incidents. The design and performance of an effective machine learning (ML) based Intrusion Detection System (IDS) depends upon the selected attributes and the classifier. This project considers multi-class classification for the Aegean Wi-Fi Intrusion Dataset (AWID) where classes represent 17 types of the IEEE 802.11 MAC Layer attacks. The proposed work extracts four attribute sets of 32, 10, 7 and 5 attributes, respectfully. The classifiers achieved high accuracy with minimum false positive rates, and the presented work outperforms previous related work in terms of number of classes, attributes and accuracy. The proposed work achieved maximum accuracy of 99.64% for Random Forest with supply test and 99.99% using the 10-fold cross validation approach for Random Forest and J48.

Introduction

Machine learning approaches using Neural Network (NN) can be classified into three sub-groups as illustrated in Figure 1. In this project, we aim to investigate how to model an intrusion detection system based on machine learning approaches. In this context, we reviewed the existing datasets and their usage in IDS and ML between 2014 and 2018. The review includes different datasets such as AWID, UNSW-NB15 and GPRS. Here, we consider multi-class classification for the Aegean Wi-Fi Intrusion Dataset (AWID).

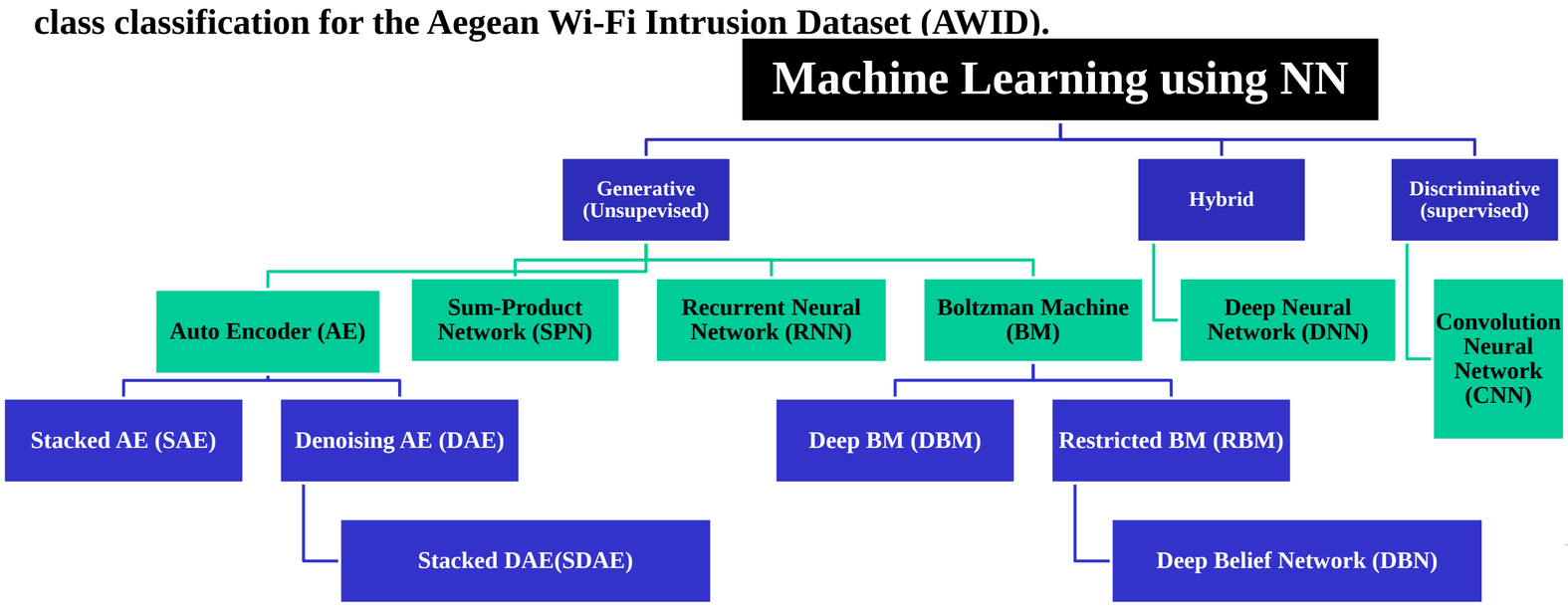

Figure 1. Machine Learning Using NN Classification Approaches

AWID-ATK-R Structure

AWID-ATK-R, a subset of AWID, is a labeled dataset with a total number of 155 attributes (features). AWID-ATK-R was collected based on real traces of normal and intrusion activities of the 802.11 Wi-Fi network and it has a finer grained class labeling corresponding to the attack name. The characteristics of AWID-ATK-R data set are highlighted in Table 1.

Table 1. AWID-ATK-R dataset characteristics

File Name	Classes	Size	Type	Total records	Attack Records
AWID-ATK-R-Trn	10	Reduced	Training	1795575	1633190
AWID-ATK-R-Tst	17	Reduced	Test	575643	44858

Proposed Framework

The procedure of our proposed framework, as presented in Figure 2, mainly includes Preprocessing, Feature selection and Classification. We extracted four different attribute group sets (AGS): 32-AGS, 10-AGS, 7-AGS, and 5-AGS. The selected attributes group sets were able to achieve high accuracy. The key contributions of this work include extracting a new subset of 5 features (attributes) using search based heuristics in the features space that produce high accuracy of 99.99% with minimum false positives (FP) rates and validated using the 10-folds cross validation approach.

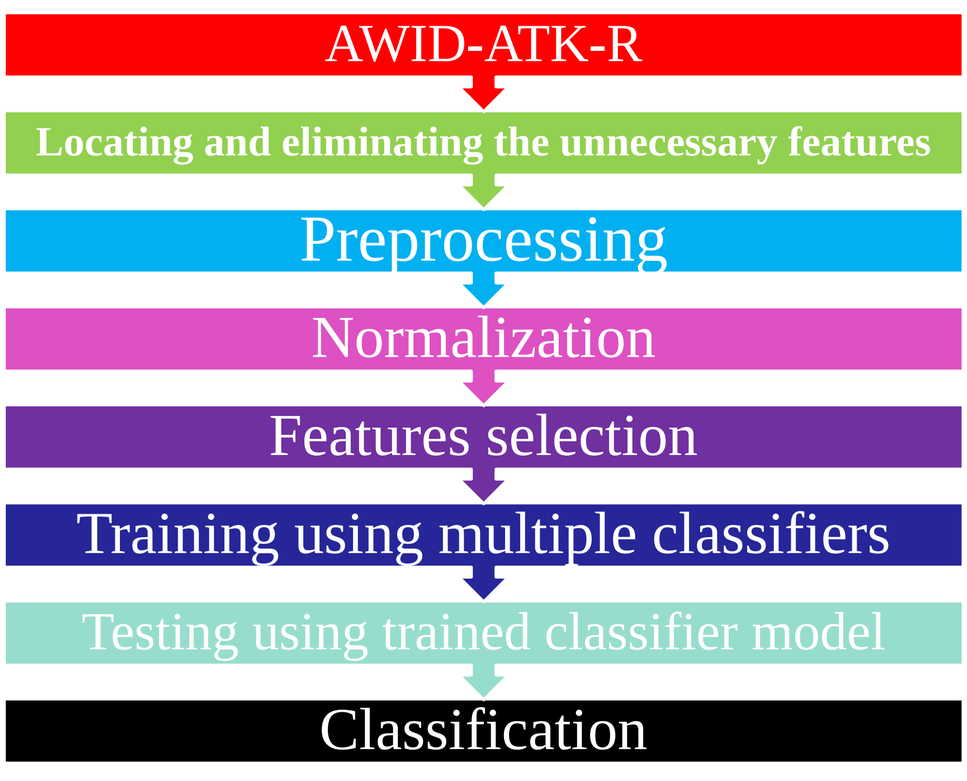

Figure 2. Proposed Machine Learning IDS Framework

Simulation Results

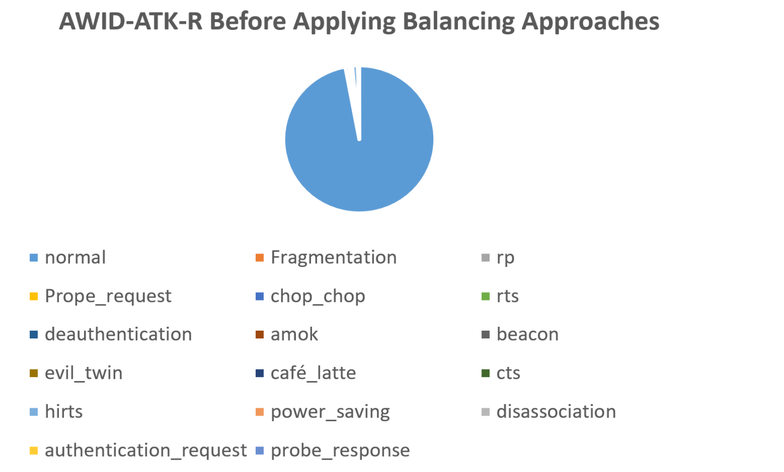

Figure 3. Class distribution of the dataset before applying balancing techniques

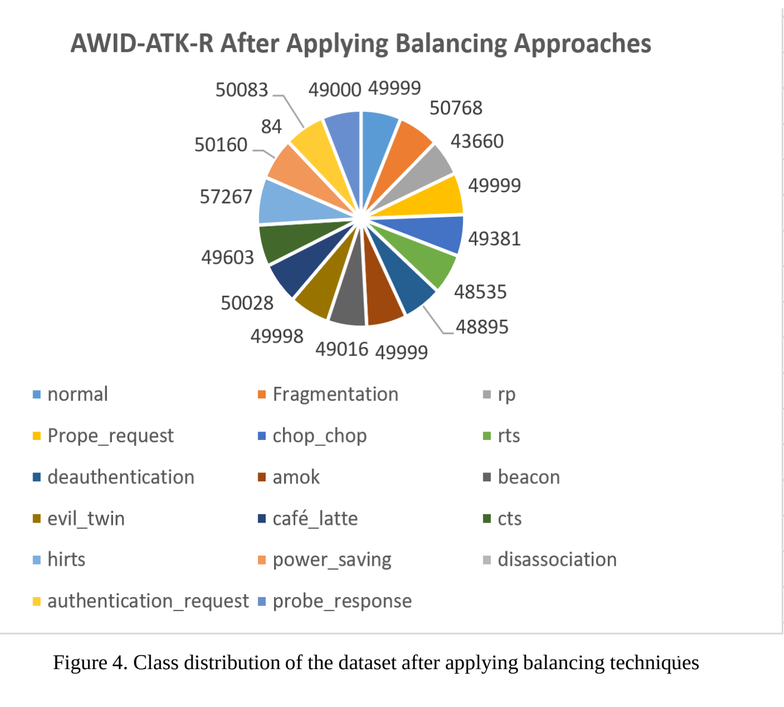

Figure 4. Class distribution of the dataset after applying balancing techniques

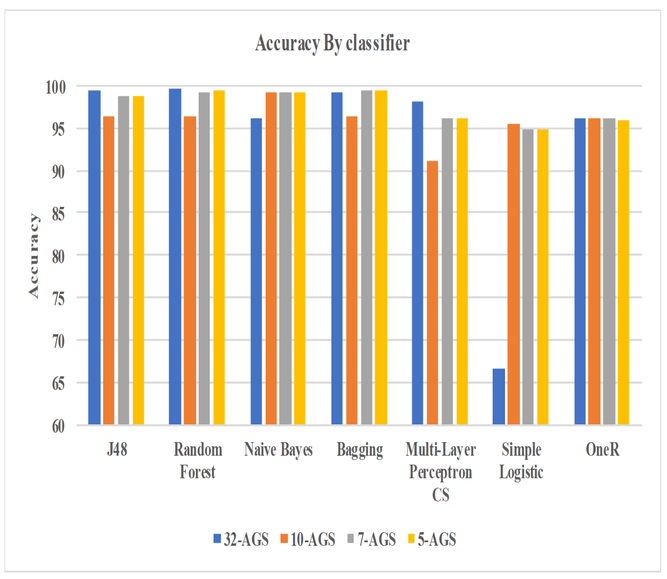

Figure 5. Accuracy by classifier

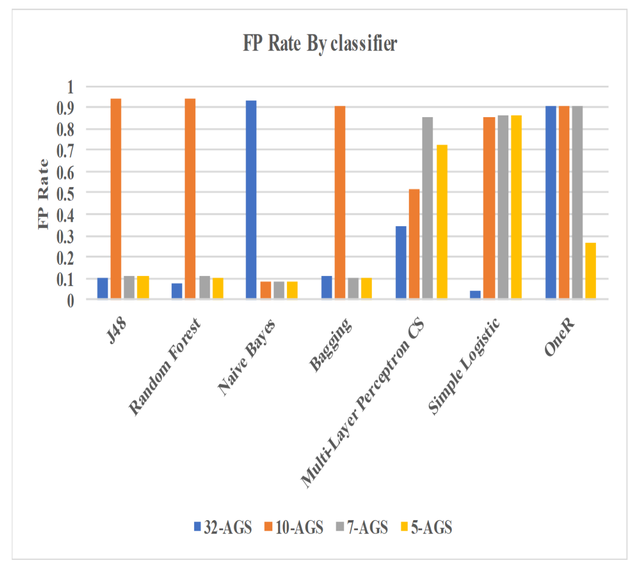

Figure 6. FPR by classifier

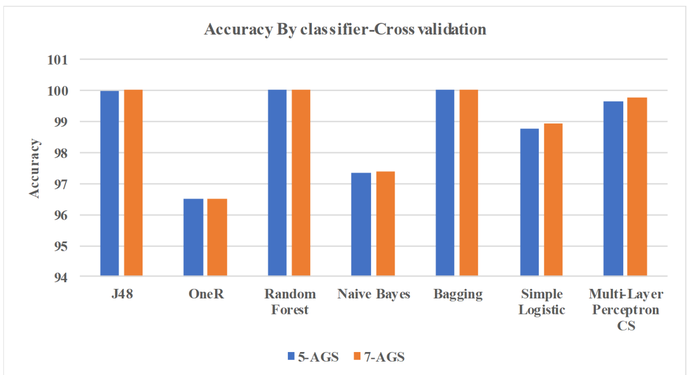

Figure 7. Accuracy by classifier-Cross Validation

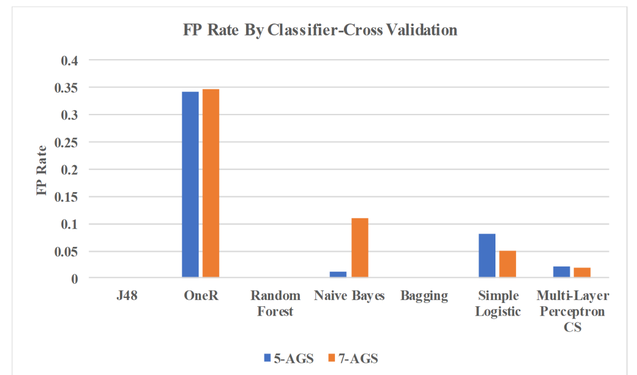

Figure 8. FPR by classifier-Cross Validation

Conclusion

Seven well known classifier algorithms; which are J48, OneR, Naive Bayes, Random Forest, Simple Logistic, Bagging and Multi-Layer Perceptron CS; were evaluated through 4 features (attributes) group sets. The results confirm that optimum features selection/reduction can lead to better results in terms of accuracy and false positive rates (FPR). Future studies on the current topic are therefore recommended. We will continue our work with special emphasis on designing a deep learning architecture that can handle imbalanced class distribution efficiently and effectively.

References

[1] Kolias, Constantinos and Kambourakis, Georgios and Stavrou, Angelos and Gritzalis, Stefanos, "Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset," IEEE Communications Surveys & Tutorials , vol. 18, no. 1, pp. 184–208, 2016.
[2] M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Deep abstraction and weighted feature selection for wi-fi impersonation detection," IEEE Transactions on Information Forensics and Security , vol. 13, no. 3, pp. 621–636, March 2018.
[3] Vrizlynn LL Thing. Ieee 802.11 network anomaly detection and attack classification: A deep learning approach. In Wireless Communications and Networking Conference (WCNC), 2017 IEEE , pages 1–6. IEEE, 2017.
[4] Mateusz Lango and Jerzy Stefanowski.Multi-class and feature selection extensions of roughly balanced bagging for imbalanced data. Journal of Intelligent Information Systems , 50(1):97–127, 2018