Wireless Intrusion and Attack Detection for 5G Networks using Deep Learning Techniques

Bayana alenazi[1], Dr. Hala Eldaw Idris[2]
Collage of Computer and Information Science
Jouf University, Al-Jouf, Saudi Arabia

***Abstract*—A Wireless Intrusion Detection System is an important part of any system or company connected to the internet and has a wireless connection inside it because of the increasing number of internal or external attacks on the network. These WIDS systems are used to predict and detect wireless network attacks such as flooding, DoS attack, and evil- twin that badly affect system availability. Artificial intelligence (Machine Learning, Deep Learning) are popular techniques used as a good solution to build effective network intrusion detection. That's because of the ability of these algorithms to learn complicated behaviors and then use the learned system for discovering and detecting network attacks. In this work, we have performed an autoencoder with a DNN deep algorithm for protecting the companies by detecting intrusion and attacks in 5G wireless networks. We used the Aegean Wi-Fi Intrusion dataset (AWID). Our WIDS resulted in a very good performance with an accuracy of 99% for the dataset attack types: Flooding, Impersonation, and Injection.**

Keywords*—Wireless intrusion detection system; 5G; autoencoder; deep learning; attack detection

I. INTRODUCTION

Wireless networks are being developed continuously every day due to the wide range of functionalities and capabilities they could provide, which make our lives easier. It is considered an important topic has a lot of research on it.

One of the recent interesting wireless technologies is the 5G (the 5th generation mobile network), which is "a new global wireless standard after 1G, 2G, 3G, and 4G networks that enables a new kind of network designed to virtually connect everyone and everything together including machines, objects, and devices. The 5G wireless technology has many advantages, such as the ability to deliver higher multi-Gbps peak data speeds, ultra-low latency, more reliability, massive network capacity, increased availability, higher performance, and improved efficiency." (Kei 2015) [1].

As long as 5G wireless technology uses internet protocols and have connectivity features, it means that cyber-attacks could occur on it and cause, and because of the growing number of its functionalities which cause increasing of vulnerabilities on 5G, which resulted increasing in security threats, attacks, and malicious activities on it. This can cause huge damage to devices or lead to data loss.

A lot of cyber-attacks can occur on wireless network technologies like Flooding or packet sniffing, fake authentication, injection, Impersonation, session hijacking, denial of service (DoS), and address spoofing, traffic analysis, and unauthorized access.

The wireless devices relate to each other in one environment, so cyber attackers need to target any software or hardware device or component in this wireless network environment, such as wireless end devices and mobiles, access points, etc. That explains how much necessary to detect these attacks and prevent them from protecting the whole system environment.

The cybersecurity concept, in general, represents techniques and methods that should be used to protect the information system from cyber-attacks, remotely or physically, to avoid destruction and damage that may happen because of it. Several algorithms and technologies have been proposed and used as cybersecurity solutions, but most of these traditional defense solutions fail in detecting or preventing new or zero-day attacks. One of the adaptive and powerful approaches is using machine learning, which is an application of artificial intelligence (AI) that focuses on building applications and systems that can automatically improve themselves from practice and experience by training them on labeled datasets and develop their accuracy over time, in attacks detection, depending on the behavior patterns. And, deep learning, which is "a subset of machine learning, which is essentially a neural network with three or more layers."(IBM 2021) [2]. Using these techniques results in higher performance and accuracy, reduce the workload on security experts. In the context of research and development for information security Wireless Intrusion systems, numerous machine learning techniques have been widely used: Random Forest,Adaboost [3], Naïve bayes and Random Forest (RF) by Jabbar et al [4], Decision Tree(DT) and k-Nearest Neighbor (KNN) by Yerong et al., 2014 [5].

Few studies have used Deep learning techniques: Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) have been implemented (Raj, Akash, et al., 2018) [6], (Bayan et al., 2019) used Auto-encoder[7], (Riyaz, B., and Sannasi Ganapathy,2020) proposed convolutional neural network CNN[8].

This work suggests a methodology for wireless cyber-attacks detection in 5G networks based on deep hybrid learning that overcomes traditional Machine learning techniques and has shown better performance with high accuracy than prior studies.as shown in Table II.

We have used deep learning algorithms such as Autoencoder, DNN, and others [9]. Then developing a deep learning model that will help detect, defend against cyber-attack, and distinguish between normal and suspicious traffic. The AWID dataset [10] will be used to build this model, which comprises a large set of packets. It contains more than 150 different attributes. And it is available as a CSV format file.

The remaining parts of the research are organized as the Second section shows some of the related papers of cyber-attack on 5G and how to prevent them. Section 3 discusses the methodology we applied and the work details with steps, Section 4 comparison our achievement with other works and the result discussion. The last section is the conclusion and the references of the work.

II. Related Work

Recent years have seen increased use of deep learning, a subdiscipline of machine learning. It has been applied to intrusion detection, where tests showed that deep learning was significantly better than conventional methods. In this section, we present some related works that have been made by prominent researchers that have used Machine learning and deep learning.

In term of network intrusion detection system in general, Dawoud et al. have built multi-layered neural network model that achieved an accuracy of 99%, however, they used KDD99 which is 20 years old and doesn't contain recent or current network attacks. [11].

He Fang et al. proposed modern authentication methodologies that use machine learning algorithms and improve the intelligence in the authentication layer to get more efficient security in 5G wireless networks and they achieve reliable, effective, continuous, and situation-aware authentication approach [12]. Fig. 1 presents their framework diagram of the intelligent authentication design which they proposed.

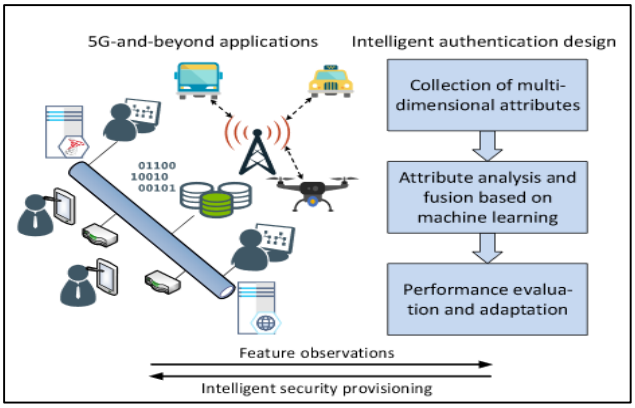

Fig. 1. [12]. Intelligent Authentication Design by (He Fang ,2019).

Jordan Lam et al. have designed an automated end-to-end network defense system that uses CNN machine learning algorithm to design an anomaly detection model for 5G networks, and they have tested the model on a real network environment include malicious and normal traffic flow. The results showed that this system detects the attack with an accuracy of 96% [13]. A high-level view of the End-to-end 5G network threat landscape is shown below in [14]. The idea related to this paper to categorize the threats that could happen on 5G networks.

Lorenzo et al. have developed a novel 5G-oriented cyber defense system to detect cyber threats and attacks in 5G mobile networks fast and efficiently. They used deep learning algorithms for building the model on well-known botnet data set. Good accuracy was achieved in their system. [14].

Jiaqi Li et al. were proposed an intelligent intrusion detection system on Software Defined 5G architecture, taking the advances of artificial intelligence and machine learning, they used the Random Forest technique with the k-mean and Adaboost algorithm. The system was able to detect unknown attacks and the evolution has shown that it has better performance and lower overhead [3].

Fig. 2, [13] End-to-end 5G network threat landscape by (Jordan Lam ,2020) show Core Network Elements" include the following: the Network Function (NF, NFn), the Network Exposure Function (NEF), and the Network Repository Function (NRF).

The approach of this paper tries to solve some issues. That was done by using the AWID dataset that contains real-world wireless network attacks. And, by using deep learning algorithms that give the highest accuracy.

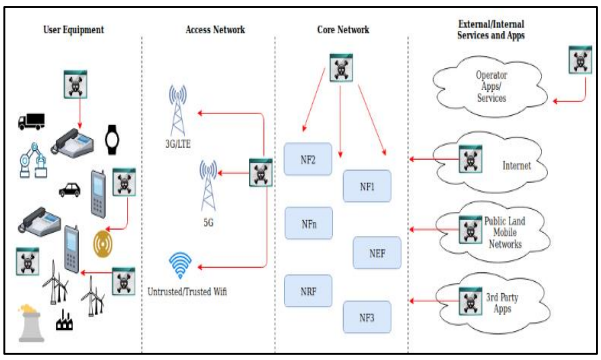

Fig. 2. [13] End-to-end 5G Network Threat Landscape.

III. Methodology

The wireless cyber-attack detection system will start collecting and capturing the traffic inside the wireless network for the database. Then preprocessing and analyzing the collected traffic and log files records. After that, the system applies features (attributes) extraction so that each instance is described by a set of features in an organized way and use an unsupervised learning neural network known as an auto-encoder (AE) with deep neural network DNN attempts to reproduce the input data as much as possible. Two training procedures come into play: the L-BFGS algorithm that sets the network weights before training and fine-tuning, in which the network parameters are fine-tuned. AE can perform dimensionality reduction. We can get more abstract high-level and low-dimensional representations of the original feature data by removing irrelevant and redundant features from the feature vector. The majority of DNNs are Feed-Forward

Networks (FFNNs), in which data flows from the input layer to the output layer without going backward. Fig. 3 showing the steps of model implementation and classification data traffic and identify attacks type.

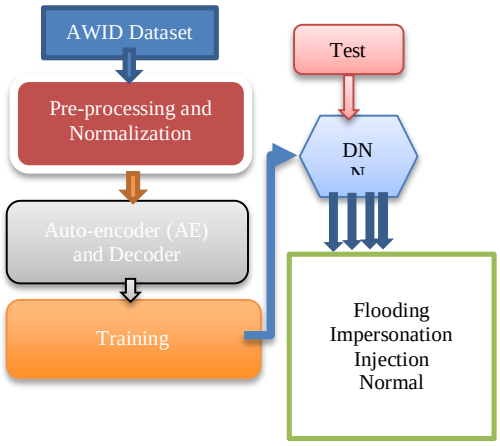

Fig. 3. Model Architecture for our Proposed System.

This model presents the initial design of any 5G network that needs to use for the attack detection system that have been built. However, there are tools that monitor and capture the traffic inside the 5G network such as TCP dump/W Wireshark and CIC Flow meter, then the traffic will be converted as CSV files so that our model could be able to classify it, and will generate a notification when an attack hit the network as shown in Fig. 3.

A. *Aegean Wi-Fi Intrusion Dataset (AWID) Description*

AWID is a public free available dataset[10] that contains Wi-Fi (802.11) network traffic which was captured and collected from physical real-time infrastructure wireless network traffic [15]. The dataset includes about 156 attributes that describe information and data about the MAC layer as numeric or string values. Fig. 4 shows the infrastructure was used to collect the dataset. This dataset is oriented toward being used for IDS especially wireless intrusion detection systems, and it's the first of its type.

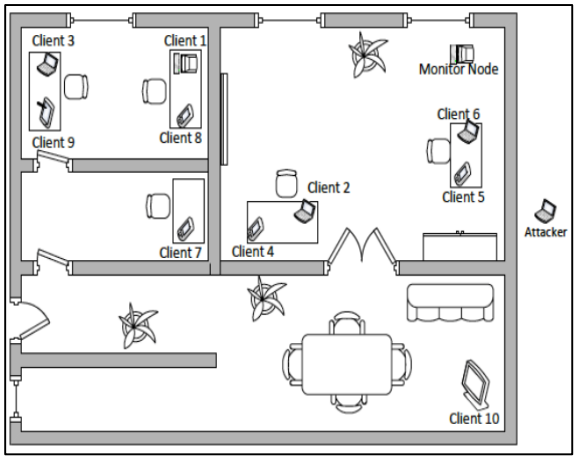

Fig. 4. The Infrastructure used to gather the Dataset [10].

The data collected from network that contains 1 desktop machine, 2 laptops, 2 smartphones, 1 tablet, and 1 smart TV and it was covered by one AP (Netgear N150 WNR1000 v3), protected by the untrusted WEP encryption.

This dataset consists of two types, one of them call "CLS" which contains four labels: Impersonation, Flooding, Injection, and Normal. While the other is called "ATK" that contain 16 classes, which are subcategories of the main four classes exist in "CLS": DE-authentication, disassociation, authentication request, amok, probe request, probe response, beacon, CTS, RTS, power saving, power, evil-twin, caffe-. In this research, we used the "CLS" dataset called (AWID-CLS-R-Trn, AWID-CLS-R-Tst) that contains four labels, three extracted from packets, they can be found in the reference [10] attacks scenario (the dataset contains 155).

TABLE I. SHOW THE REDUCED AND USED FEATURES

NO.	Important Attributes name
1	frame.time_delta_displayed
2	frame.time_relative
3	frame.len
4	radiotap.length
5	radiotap.present.tsft
6	radiotap.present.flags
7	radiotap.present.channel
8	radiotap.present.fhs
9	radiotap.present.dbm_antsignal
10	radiotap.present.antenna
11	radiotap.present.rxflags
12	wlan.fc.type_subtype
13	wlan.fc.version
14	wlan.fc.type
15	wlan.fc.subtype
16	wlan.fc.ds,wlan.fc.frag
17	wlan.fc.retry
18	wlan.fc.pwrmgt
19	wlan.fc.moredata
20	wlan.fc.protected
21	wlan_mgt.fixed.capabilities.cfpoll.ap
22	wlan_mgt.fixed.listen_ival
23	wlan_mgt.fixed.status_code
24	wlan_mgt.fixed.time
25	stamp,wlan_mgt.fixed.aid
26	wlan_mgt.fixed.reason_code
27	wlan_mgt.fixed.auth_seq
28	wlan_mgt.fixed.chanwidth
29	wlan_mgt.tim.bmapctl.offset
30	wlan_mgt.country_infoenvironment
31	wlan_mgt.rsn.capabilities.ptksa_replay_counter
32	wlan_mgt.rsn.capabilities.gtksa_replay_counter
33	wlan.qos.ack
34	wlan_mgt.fixed.timestamp
35	class

B. *Preprocessing the Dataset*

The following steps are adopted to preprocess the Dataset:

- Delete missing value.
- Format data into standard data type.
- To reduce the size of the datasets, reduce the unnecessary accuracy of the float numbers by dropping digits after the decimal point.

- Drop samples with "Infinity" and "NaN" values.
- Parse and remove columns that were repeated in the dataset.
- Reduce the unbalance in the dataset.

After Preprocessing, about 36,366 samples were dropped because of the data cleanup process.

We are going to apply feature selection because the dataset has a huge number of attributes and that would cause failure and affect the accuracy badly.

C. Normalization

Classification performance can be affected by features with different scales.so it is important to normalize the values in each attribute and map features onto a normalized range. So that the minimum value in each attribute is zero and the maximum is one. This provides more homogeneous values to the classier while maintaining relativity among the values of each attribute.

D. Feature Selection

All features with zero amount of variation are removed as those features will not influence the prediction of the target variable. We removed the duplicated features using NetMate and WEKA, as we said and removed all single value features. Then, we dropped unrelated features. In the next step, Features having a high correlation amongst each other are removed. Those features won't bring any additional predictability but may introduce noise.

Finally, the remaining number of features in the dataset after feature engineering is 35 features that we will use in building the model as shown in Table I.

E. Data Splitting

After finishing the feature engineering and preprocessing on the dataset, we split the dataset with (80/20).In the 80/20 method, we split the data into two subsets which take 80% from each label of the dataset for training the model and 20% from each label of the dataset for the test.

F. Model Methodology

In our project model, the essential goal is to get a set of statistics information of traffic flow in the 5G network and identify if this traffic is benign or malicious based on learning a set of already labeled data containing both benign and intrusion traffic.

Training of the Hybrid Deep learning model is two-part, which comprises of forwarding Propagation and Back Propagation. Back Propagation (AE) is responsible for propagating values backward used to give compressed reduced features, and the feed-forward is used for classification data traffic. The first Model is Autoencoder ANN for unsupervised pre-training, which will help in decreasing the rebuilding error. The information we'll get from this step will provide good accuracy to provide an efficient learning system. The autoencoder took all attributes without class attribute (unsupervised) as input, and then it will give compressed reduced features at the end of this phase. The purpose of this stage is to discover the perfect parameters for the next step.

The second Model supervised DNN (dense neural network) classification process, where we applied three Dnn layers that took the first step's output as input here. This stage performs the training process with the label feature as it is a supervised process. The output layer is the layer that compares the actual values (classes) and the predicted values of the test part of the dataset, then calculate the error. We used SoftMax activation for that.

After generating a training and testing dataset, we applied an unsupervised Autoencoder algorithm [16]. We supervised Classification with DNN [17] to get a better classification model with better accuracy that suits any system and to be able to detect any new attack (zero-day) to be effective with the continuous modification and development of wireless technologies. Then we evaluate the results using the criteria of precision, recall, F1 score, and accuracy that we described in the background.

G. Experimental Setup

We used python Keras library with TensorFlow [18] back end on Google colab for implementing the classification model. The colab provides us 12 GB RAM, 68 GB disk space and ability to use GPU hardware accelerator, and the ability to mount Google drive cloud storage to use it.

We used batch normalization that helps make the training process faster. Our model achieved 99.8% accuracy performance, as shown in Table II. The model architecture is shown in Fig. 5.

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	[(None, 34)]	0
model_7 (Functional)	(None, 24)	840
dense_12 (Dense)	(None, 34)	850
dense_13 (Dense)	(None, 24)	840
batch_normalization_2 (Batch	(None, 24)	96
dense_14 (Dense)	(None, 4)	100

Fig. 5. DNN Model Summary.

Fig. 8 shows the confusion Matrix for the four Classes, and the model accuracy is shown in Fig. 9, where it presents the high accuracy of our model during all epochs.

IV. Result and Discussion

Free and Open Datasets: Many public datasets are more attractive to demonstrate and compare productivity and affected Attacks using various detection methods. This section includes two excellent data sets used in many research projects, including AWID and KDDCup datasets; Table II contains the relevant details to see how a certain methodology performs and for comparative purposes.

The performance of ANN, CNN, LSTM, naive Bayesian, random forest, multi-layer perceptron, support vector machine, and other machine learning techniques in the Multi classification are studied in the AWID and KDD datasets. This concludes our analysis of the hybrid deep learning model, which compares favorably to traditional methods. This is evident: the dataset features had an imbalanced distribution. We believe our preprocessing step helped and enabled Autoencoder and DNN on the dataset to achieve high accuracy, as shown in Table II.

TABLE II. COMPARISON OUR MODEL WITH OTHER WORK

Dataset	Number of features used	Technique	Accuracy %
KDD Cup[3]	41	RF	98.5
KDD Cup[4]	15	Naive bayes	82.5
KDD 99 Cup [8]	38	CNN	98.88
AWID[19]	41	RF	94.6
AWID[20]	50	Stacked Autoencoder (SAE)	94.81
AWID[21]	25	Weighted-C4.5	99.72
KDD 99 cup[22]	41	hybrid approach	99
AWID on our model	35	Hybrid Model(HM)	99.8
-	35	HM(Adam optimizer)	99.7

A. Experimental Results

The confusion matrix, Accuracy, Precision, Recall, and F1-Score metrics are being used as evaluation metrics to evaluate the performance of any classifier.

Fig. 6 shows the classification report of our classifier, which explains most of the metrics.

	precision	recall	f1-score	support
flooding	0.98	0.98	0.98	14230
impersonation	0.99	1.00	0.99	17072
injection	1.00	1.00	1.00	20486
Normal	1.00	1.00	1.00	541017
accuracy			1.00	592805
macro avg	0.99	0.99	0.99	592805
weighted avg	1.00	1.00	1.00	592805

Fig. 6. Classification Report for the Model.

The exact accuracy result is shown in Fig. 6 and 7. (WIDS) with Deep learning, this system detects attacks in real-time, the system detects attack by using the model that was trained by different deep learning classification algorithms, the model was applied by using autoencoder and DNN algorithms, then we evaluated it. The model was trained by using the AWID wireless dataset which contains different types of attacks that face the wireless network such as (Flooding, Impersonation, Injection, etc.).

In conclusion, the system ensures the accuracy of the results and the speed of performance, and ease of use.

```
loss: 0.0048 - accuracy: 0.9989
```

Fig. 7. The obtained Accuracy and Loss.

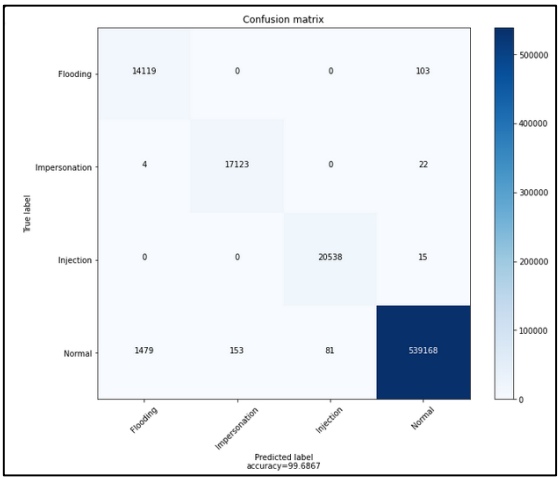

Fig. 8. Confusion Matrix for the 4 Classes.

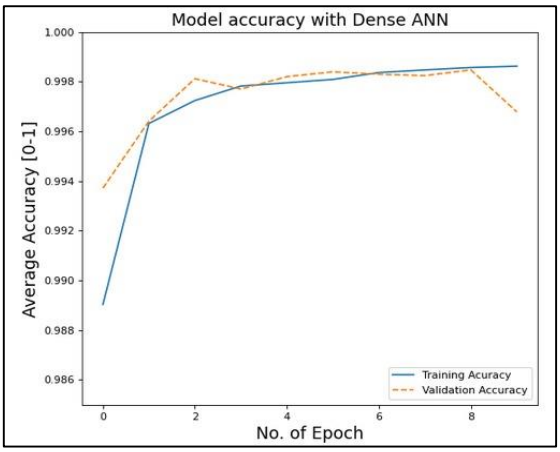

Fig. 9. The overall Accuracy vs Validation.

V. CONCLUSION

Detection Attack facing wireless networks is very important because the network is the main pillar of any 5G wireless network. If this network is compromised, the integrity, safety, and data availability for this architecture do not exist.

Through this project, we implemented a 5G wireless Intrusion Detection System (WIDS) with Deep learning; this system detects attacks in real-time, the system detects attacks by using the model that was trained by different deep learning classification algorithms, the model was applied by using autoencoder and DNN algorithms, then we evaluated it. The model was trained using the AWID wireless dataset, which contains different types of attacks that face the wireless network (Flooding, Impersonation, Injection, etc.). In conclusion, the system ensures the accuracy of the results, speed of performance, and ease of use.

VI. Future Work

Much more can be added to this system. Future work concerns building a complete 5G network in one simulation software to build a complete system that starts with capturing the traffic and classifying it. And make the system easy to use by admin and other users. We can also make the system provides alerts when attacks happen on the 5G network and generate a log for this attack and store it in the database as Archived data.

References

[1] K. Sakaguchi et al., "Millimeter-wave wireless LAN and its extension toward 5G heterogeneous networks," IEICE Trans. Commun., vol. E98B, no. 10, pp. 1932–1948, Oct. 2015, doi: 10.1587/TRANSCOM.E98.B.1932.

[2] "AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? | IBM." https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks (accessed Jul. 31, 2021).

[3] J. Li, Z. Zhao, and R. Li, "A Machine Learning Based Intrusion Detection System for Software Defined 5G Network," 2017, Accessed: Jul. 31, 2021. [Online]. Available: www.ietdl.org.

[4] M. A. Jabbar, R. Aluvalu, and S. S. Reddy, "RFAODE: A Novel Ensemble Intrusion Detection System," Procedia Comput. Sci., vol. 115, pp. 226–234, 2017, doi: 10.1016/j.procs.2017.09.129.

[5] Y. Tao, S. Sui, K. Xie, and Z. Liu, "Intrusion detection based on support vector machine using heuristic genetic algorithm," Proc. - 2014 4th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2014, pp. 681–684, 2014, doi: 10.1109/CSNT.2014.143.

[6] A. R. Narayanadoss, T. Truong-Huu, P. M. Mohan, and M. Gurusamy, "Crossfire attack detection using deep learning in software defined its networks," IEEE Veh. Technol. Conf., vol. 2019-April, pp. 1–6, 2019, doi: 10.1109/VTCSpring.2019.8746594.

[7] B. Alsughayyir, A. M. Qamar, and R. Khan, "Developing a network attack detection system using deep learning," 2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019, pp. 1–5, 2019, doi: 10.1109/ICCISci.2019.8716389.

[8] B. Riyaz and S. Ganapathy, "A deep learning approach for effective intrusion detection in wireless networks using CNN," Soft Comput., vol. 24, no. 22, pp. 17265–17278, 2020, doi: 10.1007/s00500-020-05017-0.

[9] A. Di Ciaccio and G. M. Giorgi, "DEEP LEARNING FOR SUPERVISED CLASSIFICATION," 2016.

[10] "AWID - Aegean Wi-Fi Intrusion Dataset." https://icsdweb.aegean.gr/awid/awid2 (accessed Jul. 31, 2021).

[11] A. Dawoud, S. Shahristani, and C. Raun, "Deep learning for network anomalies detection," Proc. - Int. Conf. Mach. Learn. Data Eng. iCMLDE 2018, pp. 117–120, 2019, doi: 10.1109/iCMLDE.2018.00035.

[12] H. Fang, X. Wang, and S. Tomasin, "Machine learning for intelligent authentication in 5G and beyond wireless networks," IEEE Wirel. Commun., vol. 26, no. 5, pp. 55–61, 2019, doi: 10.1109/MWC.001.1900054.

[13] J. Lam, "Machine Learning based Anomaly Detection for 5G Networks."

[14] L. Fernandez Maimo, A. L. Perales Gomez, F. J. Garcia Clemente, M. Gil Perez, and G. Martinez Perez, "A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks," IEEE Access, vol. 6, no. February, pp. 7700–7712, 2018, doi: 10.1109/ACCESS.2018.2803446.

[15] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset," IEEE Commun. Surv. Tutorials, vol. 18, no. 1, pp. 184–208, 2016, doi: 10.1109/COMST.2015.2402161.

[16] H. Choi, M. Kim, G. Lee, and W. Kim, "Unsupervised learning approach for network intrusion detection system using autoencoders," J. Supercomput., vol. 75, no. 9, pp. 5597–5621, 2019, doi: 10.1007/s11227-019-02805-w.

[17] Y. Yang, K. Zheng, C. Wu, and Y. Yang, "Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network," Sensors (Switzerland), vol. 19, no. 11, 2019, doi: 10.3390/s19112528.

[18] N. K. Manaswi, "Deep Learning with Applications Using Python," Deep Learn. with Appl. Using Python, pp. 31–43, 2018, doi: 10.1007/978-1-4842-3516-4.

[19] U. S. K. P. M. Thanthrige, J. Samarabandu, and X. Wang, "Machine learning techniques for intrusion detection on public dataset," Can. Conf. Electr. Comput. Eng., vol. 2016-Octob, pp. 7–10, 2016, doi: 10.1109/CCECE.2016.7726677.

[20] M. E. Aminanto and K. Kim, Improving Detection of Wi-Fi Impersonation by Fully Unsupervised Deep Learning, vol. 10763 LNCS. Springer International Publishing, 2018.

[21] M. E. Aminanto, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Wi-Fi intrusion detection using weighted-feature selection for neural networks classifier," Proc. - WBIS 2017 2017 Int. Work. Big Data Inf. Secur., vol. 2018-Janua, pp. 99–104, 2018, doi: 10.1109/IWBIS.2017.8275109.

[22] N. Araújo, R. De Oliveira, E. Ferreira, A. A. Shinoda, and B. Bhargava, "Identifying important characteristics in the KDD99 intrusion detection dataset by feature selection using a hybrid approach," ICT 2010 2010 17th Int. Conf. Telecommun., pp. 552–558, 2010, doi: 10.1109/ICTEL.2010.5478852.