A Comparative Study for SDN Security Based on Machine Learning

https://doi.org/10.3991/ijim.v17i11.39065

Khattab M Ali Alheeti[1(✉)], Abdulkareem Alzahrani[2], Maha Alamri[2], Aythem Khairi Kareem[3], Duaa Al_Dosary[1]

[1] Computer Networking Systems Department, University of Anbar, Anbar, Iraq

[2] Faculty of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia

[3] Department of Heet Education General Directorate of Education in Anbar, Ministry of Education, Heet, Anbar, Iraq

co.khattab.alheeti@uoanbar.edu.iq

Abstract—In the past decade, traditional networks have been utilized to transfer data between more than one node. The primary problem related to formal networks is their stable essence, which makes them incapable of meeting the requirements of nodes recently inserted into the network. Thus, formal networks are substituted by a Software Defined Network (SDN). The latter can be utilized to construct a structure for intensive data applications like big data. In this paper, a comparative investigation of Deep Neural Network (DNN) and Machine Learning (ML) techniques that uses various feature selection techniques is undertaken. The ML techniques employed in this approach are decision tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM). The proposed approach is tested experimentally and evaluated using an available NSL–KDD dataset. This dataset includes 41 features and 148,517 samples. To evaluate the techniques, several estimation measurements are calculated. The results prove that DT is the most accurate and effective approach. Furthermore, the evaluation measurements indicate the efficacy of the presented approach compared to earlier studies.

Keywords—Software Defined Network (SDN), Deep Neural Network (DNN), Machine Learning (ML), NSL-KDD

1 Introduction

In wireless communication networks, sensor nodes in wireless sensor networks are considered as the fundamental backbone of WSN [1–2]. WSNs have hundreds to thousands of homogeneous or heterogeneous sensors. Most WSNs their sensor nodes handle essential functions like detecting, handling, communication and computation. Their neighboring nodes' communication is enabled through electromagnetic signals via radio frequency [3].

WSNs are utilized for both tracking and monitoring purposes. WSNs can be used to monitor patient health care, provide chemicals for the rubber industry, and monitor

toxic gas. WSN technologies are also found in tracking methods, such as tracking wild species, pets or people [4]. Recently, WSNs have been revised for efficient communication, innovativeness and cost-effectiveness.

Sensor nodes (SNs) are able to decipher, identify and transmit radio frequency information [5][6]. In addition, WSNs are helpful for checking and following purposes in unavailable and hostile conditions, in which human mediation does not or cannot happen. Checking purposes include observation of concoction vapors and gaseous tension observation; tracking purposes include human tracking and animal tracking [7]. Security of computer networks has become a primary concern; the cloud has exposed network technologies to different invasive activities, leading to intense failure. Hence, it is critical to combine security instruments to ensure that the system's security is not threatened. The purpose of a secure approach is to protect important information through determining anomalies.

It is essential to fuse security components so that the security of the framework is not threatened. The goal of a solid framework is to secure shrewd data through recognizing oddities. An Intrusion Detection System (IDS) is a powerful asset that identifies unwanted actions that can access, control, or incapacitate a PC framework, chiefly through the Internet. It screens approaching and active traffic to distinguish noxious activities that risk the security of the framework.

IDSs based on machine learning are the leading systems in the field of intrusion detection, and such an approach is critical for improving the quality of IDS performance [8–9]. Moreover, characteristics of ML techniques improve detection, reduce false alarm rates, and decrease computational and communication costs [10].

Machine learning enables systems to learn automatically from data and recognize hidden patterns, producing useful classification of unseen data [11]. Therefore, anomaly detection systems learn to identify attacks in normal traffic by generating patterns from several features of the traffic data set [12]. The most common algorithms used in this kind of IDS are support vector machines (SVMs) [13–14], modified SVMs [15–16], decision trees (DT), random forests (RF), naïve Bayes (NB), K-nearest neighbors (KNN) [17], artificial neural networks (ANN) and multi-layer perceptron [18].

The remaining parts of the paper consist of related works, methodology, feature selection, experimental results and conclusions.

2 Related works

Recently, a considerable number of research studies have been conducted in the Intrusion Detection System (IDS) field, which identifies various attacks on networks. In addition, a comprehensive inspection of the field has been conducted. ML techniques are employed to develop IDSs that handle a considerable amount of data. Feature selection methods are used to remove redundant or irrelevant features from the dataset to achieve efficiency of the model.

In [19], the authors proposed a method based on feature selection techniques. The proposed approach exploits pigeon-elucidated optimizers to determine a large number of attributes of the dataset. To analyses the presented system, NLS–KDD, KDDCUP99

and UNSW–NB15 were used. The results showed that the model had an efficient, accurate rate of detection.

In [20], the authors proposed the use of an RF classifier to eliminate unwanted features from the dataset and reduce computational complexity. Many ML classifiers, such as SVM, KNN, NB, DT and logistic regression (LR), were used for training and testing the proposed model. The results revealed 99% accuracy for the NSL–KDD dataset.

In [21], SVM was used to train and test the UNSW dataset for intrusion detection and performance improvement. This model was trained and tested with the UNSW–NB15 dataset. The results presented an efficient detection rate compared to other classification models (RepTree, ANN, RF and MLP).

In [22], a related research study, utilized different classifier techniques like RepTree, MLP and RF, achieving 94% accuracy. 99.85% accuracy resulted from utilizing SVM. To acquire data packets of the network, the Wireshark mechanism framework was adopted then various ML techniques were employed [23].

In [24], a hybrid approach based on ML and knowledge was proposed to detect different kinds of attacks on KDD-99. Moreover, the knowledge-based method was used to navigate the target classes and choose the appropriate system based on prediction. Three classifier techniques, MLP, NB and SVM, were utilized to analyse various feature groupings in [25]. The results showed that MLP was better than the other classifier techniques in detecting composite, regular traffic and opposing windows. Moreover, the presented approach was integrated with existing IDSs and FARs, and accuracies of 0.27% and 92.09% were achieved, respectively. In [26], the authors suggested a framework established by employing neural networks and MLP to identify network intrusions; the system was first trained to recognize whether the data was malicious or normal. The proposed model was tested using the KDD-99 dataset and the back propagation algorithm in MLP. The results showed accuracies of 94% and 91% when using all features. In [27], RF was utilized for intrusion detection, depending on dimensional reduction. Comparative research was done using different ML techniques such as DT, SVM and NB, achieving an accuracy of 96.77%. The DT algorithm was utilized to design an IDS in which BOT–IoT and CICIDS datasets were used to analyse system performance. The results showed accuracies of 96.665% and 96.995% for CICIDS and BOT–IoT datasets, respectively. In [28], temporal and fuzzy rules with the DT classifier were utilized for system training and testing on the KDD-99 dataset. Accuracies of 92.67%, 99.99%, 57.39% and 95.23% were achieved for Probe, DoS, U2R and R2L, respectively.

This study applied ML techniques like SVM, DNN, NB and DT to an NLS–KDD dataset. This approach was undertaken to simultaneously increase the accuracy ratio and decrease the execution time.

3 Methodology

The results of the presented approach are examined in this section. DNN, NB, DT and SVM labelled as either abnormal or normal. The approach's performance was tested using various attribute subsets extracted from the NSL–KDD dataset. Figure 1

illustrates the structure of the presented method. This cycle started with the dataset's acquisition, which entailed extracting the attributes and pre-processing. Then, DNN, DT, NB and SVM techniques were applied to obtain final results.

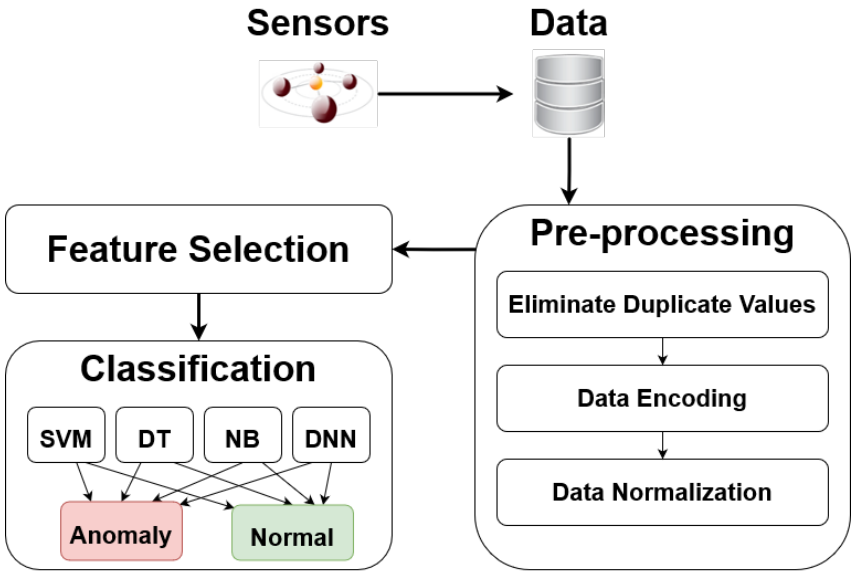

Fig. 1. System Architecture

The steps of the approach are summarized and directly examined in the following sections. The first step entailed pre-processing of the data. Then, feature selection was employed to determine the best subsets of features from the approach. Then, the DNN and three models of ML techniques were utilized to train and test the data. Finally, the experimental results of the parameters were investigated.

3.1 Dataset description

NSL–KDD is a dataset that was developed to avoid several of the intrinsic difficulties of the KDD-99 dataset. It can be used as an experimental model dataset to assist researchers in comparing various intrusion detection approaches. Also, the number of samples in the NSL–KDD set is not too large. This allows for reasonable investigation of the entire set; there is no need to randomly select a small portion. Therefore, the experimental evaluation results of various investigation works will be comparable and consistent.

The NSL–KDD dataset is available at no cost [29] and is mainly used in studies to validate ML techniques. It has 148,517 records with 41 attributes, including the class label.

3.2 Preprocessing

Pre-processing consists of drop-duplicated features, feature encoding and normalization [30]. This step is an essential phase in the performance development. This step aims to improve the extraction of selected features and reduce irrelevant data.

Eliminate duplicated features. Duplicated features negatively affect the performance of the approach. When these constraints are removed, the storage space will be freed up and the system's speed will be increased.

Feature encoding. Machine learning (ML) deals with numerical values [31]. The proposed approach uses the Python 'LabelEncoder' instruction to execute this process. Label encoding transforms labels into numerical form to make them machine-readable. ML algorithms can then more reliably determine how to achieve label requirements.

Normalization. Feature normalization is a strategy that standardizes data attributes in a specific range. Since the dataset utilized in this approach included varying values, normalization was a productive process that re-scaled the values of the feature. It can be calculated by using the following formula in Eq. (1):

$$X_{scale=} \frac{x-\mu}{\sigma} \quad (1)$$

Where X_{scale} is the standardization value, μ represents the mean value of the sample data and σ represents the standard deviation.

4 Feature selection

This approach used the python 'SelectKBest' instruction to perform this operation, where K represents the number of best features that we selected. More than one experiment was conducted. In the beginning, all the features were tested, and then we began to select the features by changing the value of K. In each experiment, we measured the performance of the system and the execution time until we reached perfect performance.

5 Experimental results

To evaluate and measure system performance, four types of alarms were needed: true positive (TP), false positive (FP), true negative (TN) and false negative (FN) [32].

- TP represents $normal\ connection\ record$ identified $as\ normal$.
- TN represents $attack\ connection\ record$ identified $as\ attack$.
- FP represents $normal\ connection\ record$ identified $as\ attack$.
- FN represents $attack\ connection\ record$ identified $as\ normal$.

The measures will be calculated as follows Eq. (2), (3), (4) and (5):

$$TP = \frac{TP}{TP+FN} \tag{2}$$

$$TN = \frac{TN}{TN+FP} \tag{3}$$

$$FN = \frac{FN}{FN+TP} \tag{4}$$

$$FP = \frac{FP}{FP+TN} \tag{5}$$

In this paper, the effectiveness of the proposed system was evaluated using performance metrics such as detection accuracy rate, shown in Eq. (6) [33–34].

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Precision and recall were determined as presented in Eq. (7) and (8) [33–34].

$$precision = \frac{TP}{TP+FP} \tag{7}$$

$$recall = \frac{TP}{TP+FN} \tag{8}$$

The results are presented in this section. DNN, NB, DT and SVM were exploited and labelled as abnormal or normal. The performance was tested on various subsets of features extracted from the NSL–KDD dataset, as shown in Table 1 and Figure 2.

Table 1. Performance metrics

No. of feature	SVM			DT			NB			DNN		
	accuracy	*precision*	*recall*	*accuracy*	*precision*	*Recall*	*accuracy*	*precision*	*recall*	*accuracy*	*precision*	*recall*
42	99.69	99.79	99.57	100	100	100	95.68	91.78	100	99.91	99.92	99.91
40	99.71	99.79	99.6	100	100	100	95.69	91.78	100	99.92	99.9	99.8
38	99.71	99.82	99.59	100	100	100	95.54	91.52	100	99.97	99.91	99.91
36	99.74	99.79	99.66	100	100	100	95.25	91.03	100	99.99	99.97	99.98
34	99.76	99.81	99.69	100	100	100	95.7	91.81	100	99.94	99.93	99.93
32	**99.77**	99.82	99.7	100	100	100	95.82	92.02	100	99.98	99.97	99.98
30	99.74	99.72	99.75	100	100	100	96.08	92.48	100	99.99	99.99	99.99
28	96.77	97.99	95.26	91.98	87.74	96.88	86.33	88.19	82.67	98.5	98.7	98.7
26	96.78	98.16	95.09	92.02	87.7	97.04	86.37	88.48	82.41	98.31	98.21	98.30
24	95.76	97.89	93.21	86.5	81.51	93.07	85.58	87.84	81.32	98.27	98.32	98.35
22	95.59	97.8	92.93	74.94	67.04	94.3	86.33	89.63	80.99	98.16	98.32	98.32
20	95.06	97.23	92.37	79.93	72.92	92.74	85.97	89.35	80.46	98	98.26	98.25
18	95.04	97.22	92.34	77.98	70.07	94.76	87.09	91.63	80.54	97.33	98.22	98.25
16	94.63	96.8	91.88	78.43	70.4	95.26	86.62	94.12	77.02	97.46	97.85	97.85
14	94.29	96.23	91.74	76.96	68.8	95.41	84.16	94.36	71.37	97.57	97.79	97.8
12	94.4	96.26	91.95	76.78	68.4	96.21	83.03	95.12	68.19	97.2	97.43	97.5
10	94.5	96.13	92.3	77.01	68.76	95.74	86.89	96.32	75.66	97.56	97.64	97.64

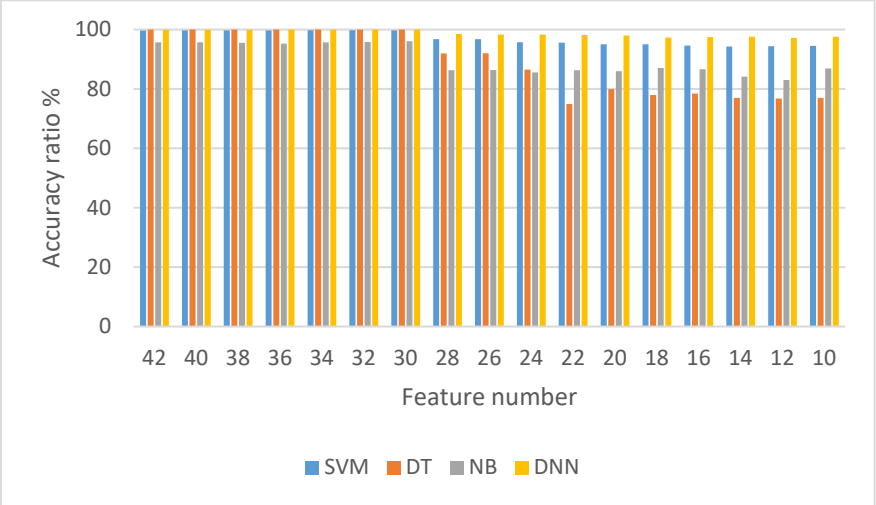

Fig. 2. Accuracy of the four classifiers

Table 1 and Figure 2 show the values of some features (accuracy, precision and recall) for each of the DNN and ML techniques. The optimum performance of all models is obtained by selecting 30 features. The best ML technique is DT.

6 Conclusion

WSNs are widely used in distinct areas to scrutinize required parametric values. The primary purpose of any WSN is to prolong the overall lifetime of the network as much as possible. In this paper, we have introduced an intelligent approach based on DNN and ML techniques. Experimental data show that optimal accuracy is obtained by applying the DT technique. The experimental results demonstrate that the proposed scheme achieves higher performance with the NSL–KDD data set than other methods.

7 References

[1] K. Haseeb, N. Abbas, M. Q. Saleem, O. E. Sheta, K. Awan et al., "RCER: Reliable Cluster-based Energy-aware Routing protocol for heterogeneous Wireless Sensor Networks," PLoS One, vol. 17, no. 9, pp. e0222009, 2019. https://doi.org/10.1371/journal.pone.0222009

[2] M. Ahmad, T. Li, Z. Khan, F. Khurshid and M. Ahmad, "A Novel Connectivity-Based LEACH-MEEC Routing Protocol for Mobile Wireless Sensor Network," Sensors, vol. 18, no. 12, pp. 4278, 2018. https://doi.org/10.3390/s18124278

[3] Q. Feng, D. He, S. Zeadally, M. K. Khan and N. Kumar, "A survey on privacy protection in blockchain system," Journal of Network & Computer Applications, vol. 126, pp. 45–58, 2019. https://doi.org/10.1016/j.jnca.2018.10.020

[4] M. Mittal and C. Iwendi, "A Survey on Energy-Aware Wireless Sensor Routing Protocols," EAI Endorsed Transactions on Energy Web, vol. 6, no. 24, pp. e5, 2019. https://doi.org/10.4108/eai.11-6-2019.160835
[5] A. Awad, R. German and F. Dressler, "Exploiting Virtual Coordinates for Improved Routing Performance in Sensor Networks," IEEE Transactions on Mobile Computing, vol. 10, no. 9, pp. 1214–1226, 2011. https://doi.org/10.1109/TMC.2010.218
[6] H. Chen, F. Gao, M. Martins, P. Huang and J. Liang, "Accurate and Efficient Node Localization for Mobile Sensor Networks," Mobile Networks & Applications, vol. 18, no. 1, pp. 141–147, 2013. https://doi.org/10.1007/s11036-012-0361-7
[7] C. Intanagonwiwat, R. Govindan and D. Estrin, "Direct Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," in Proc. ACM (Mobi-Com), Boston, MA, USA, pp. 56–67, 2007.
[8] S. M. Kasongo and Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," Journal of Big Data, vol. 7, no. 1, pp.105, 2020. https://doi.org/10.1186/s40537-020-00379-6
[9] W.A.H. Ghanem, A. Jantan, S.A.A. Ghaleb and A.B. Nasser, "An efficient intrusion detection model based on hybridization of artificial bee colony and dragonfly algorithms for training multilayer perceptrons," IEEE Access, vol. 8, pp.130452-130475, 2020. https://doi.org/10.1109/ACCESS.2020.3009533
[10] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint, arXiv:1312.2177, 2015.
[11] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis and R. Atkinson, "Shallow and deep networks intrusion detection system: a taxonomy and survey," arXiv preprint, arXiv:1701.02145, 2017.
[12] A. Ahmad, E. Harjula, M. Ylianttila and I. Ahmad, "Evaluation of machine learning techniques for security in SDN," in Proc. 2020 IEEE Globecom Workshops (GC Wkshps), pp. 1-6, 2020. https://doi.org/10.1109/GCWkshps50303.2020.9367477
[13] K. Li and G. Teng, "Unsupervised SVM based on p-kernels for anomaly detection," in Proc. First International Conference on Innovative Computing, Information and Control (ICICIC'06), Beijing, China, pp. 59–62, 2006.
[14] C. A. Catania, F. Bromberg and C. G. Garino, "An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection," Expert Systems with Applications, vol. 39, no. 2, pp. 1822–1829, 2012. https://doi.org/10.1016/j.eswa.2011.08.068
[15] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," Information Sciences, vol. 177, no. 18, pp. 3799–3821, 2007. https://doi.org/10.1016/j.ins.2007.03.025
[16] Z. Zhang and H. Shen, "Online training of SVMs for real-time intrusion detection," in Proc. 18th International Conference on Advanced Information Networking and Applications, Fukuoka, Japan, pp. 568–573, 2004.
[17] P. R. Kanna and P. Santhi, "Hybrid Intrusion Detection using MapReduce based Black Widow Optimized Convolutional Long Short-Term Memory Neural Networks," Expert Systems with Applications, vol. 194, pp.116545, 2022. https://doi.org/10.1016/j.eswa.2022.116545
[18] M. Alauthaman, N. Aslam, L. Zhang, R. Alasem and M. A. Hossain, "A P2P botnet detection scheme based on decision tree and adaptive multilayer neural networks," Neural Computing and Applications, vol. 29, no. 11, pp. 991–1004, 2018. https://doi.org/10.1007/s00521-016-2564-5

[19] H. Alazzam, A. Sharieh and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer," Expert Systems with Applications, vol. 148, pp. 113249, 2020. https://doi.org/10.1016/j.eswa.2020.113249

[20] N. Kunhare, R. Tiwari and J. Dhar, "Particle swarm optimization and feature selection for intrusion detection system," Sādhanā, vol. 45, no. 1, pp. 109, 2020. https://doi.org/10.1007/s12046-020-1308-5

[21] A. Bachar, N. Makhfi and O. Bannay, "Towards a behavioral network intrusion detection system based on the SVM model," in Proc. 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, pp. 1–7, 2020. https://doi.org/10.1109/IRASET48871.2020.9092094

[22] D. Wang and G. Xu, "Research on the Detection of Network Intrusion Prevention with SVM Based Optimization Algorithm," Informatica, vol. 44, no. 2, pp. 269–274, 2020. https://doi.org/10.31449/inf.v44i2.3195

[23] F. Fenzl, R. Rieke, Y. Chevalier, A. Dominik and I. Kotenko, "Continuous fields: Enhanced in-vehicle anomaly detection using machine learning models," Simulation Modelling Practice and Theory, vol. 105, pp. 102143, 2020. https://doi.org/10.1016/j.simpat.2020.102143

[24] M. Sarnovsky and J. Paralic, "Hierarchical Intrusion Detection Using Machine Learning and Knowledge Model," Symmetry, vol. 12, no. 2, pp. 203, 2020. https://doi.org/10.3390/sym12020203

[25] L. Sun, A. Ho, Z. Xia, J. Chen and W. Meng, "Development of an Early Warning System for Network Intrusion Detection using Benford's Law Features," in Proc. Security and Privacy in Social Networks and Big Data, Singapore, pp. 57–73, 2019. https://doi.org/10.1007/978-981-15-0758-8_5

[26] M. Darkaie and R. Tavoli, "Providing a method to reduce the false alarm rate in network intrusion detection systems using the multilayer Perceptron technique and backpropagation algorithm," in Proc. 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Tehran, Iran, pp. 1–6, 2019. https://doi.org/10.1109/KBEI.2019.8735024

[27] S. Waskle, L. Parashar and U. Singh, "Intrusion Detection System Using PCA with Random Forest Approach," in Proc. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 803–808, 2020. https://doi.org/10.1109/ICESC48915.2020.9155656

[28] P. Nancy, S. Muthurajkumar, S. Ganapathy, S.V.N. Santhosh Kumar and K. Arputharaj, "Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks," IET Commununications, vol. 14, no. 5, pp. 888–895, 2020. https://doi.org/10.1049/iet-com.2019.0172

[29] K. M. A. Alheeti, A. Gruebler and K. D. Mcdonald-maier, "An Intrusion Detection System Against Malicious Attacks on the Communication Network of Driverless Cars," in Proc. 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, pp. 916–921, 2015. https://doi.org/10.1109/CCNC.2015.7158098

[30] Alzahrani, Abdulkareem; Alheeti, Khattab M. Ali; Thabit, Samer Salah; Al-Dosary, Duaa; Al-Ani, Muzhir Shaban "Intelligent Mobile Coronavirus Recognition Centre Based on IEEE 802.15.4" August 2021International Journal of Interactive Mobile Technologies (iJIM) 15(16):4. https://doi.org/10.3991/ijim.v15i16.24193

[31] Al-Rami, Bandar; Alheeti, Khattab M. Ali; Aldosari, Waleed M.; Alshahrani, Saeed Matar; Al-Abrez, Shahad Mahdi, " A New Classification Method for Drone-Based Crops in Smart Farming" in International Journal of Interactive Mobile Technologies . 2022, Vol. 16 Issue 9, p164-174. 11p. https://doi.org/10.3991/ijim.v16i09.30037

[32] Alheeti, Khattab Alsukayti, Ibrahim Alreshoodi, Mohammed, "Intelligent Botnet Detection Approach in Modern Applications", VOL. 15 NO. 16 (2021). https://doi.org/10.3991/ijim.v15i16.24199

[33] A. A. Bahashwan, M. Anbar, I. H. Hasbullah, Z. R. Alashhab and A. Bin-Salem, "Flow-Based Approach to Detect Abnormal Behavior in Neighbor Discovery Protocol (NDP)," *IEEE Access*, vol. 9, pp. 45512–45526, 2021. https://doi.org/10.1109/ACCESS.2021.3066630

[34] K. M. A. Alheeti, V. W. Venus and M. S. Al Rababaa, "The affect of fuzzification on neural networks intrusion detection system," in *Proc. 4th IEEE Conference on Industrial Electronics and Applications*, Xi'an, China, pp. 1236–1241, 2009.

8 Authors

Khattab M. Ali Alheeti is with College of Computer and Information Technology, Computer Networking Systems Department, University of Anbar, Anbar, Iraq (email: co.khattab.alheeti@uoanbar.edu.iq).

Abdulkareem Alzahrani is with Faculty of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia.

Maha Alamri is with Faculty of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia.

Aythem Khairi Kareem is with Department of Heet Education General Directorate of Education in Anbar, Ministry of Education, Heet, Anbar, Iraq.

Duaa Al_Dosary is with College of Computer and Information Technology, Computer Networking Systems Department, University of Anbar, Anbar, Iraq.

Article submitted 2023-02-22. Resubmitted 2023-04-29. Final acceptance 2023-05-01. Final version published as submitted by the authors.