WILEY | Hindawi

Research Article

Cyber Security against Intrusion Detection Using Ensemble-Based Approaches

Mohammed Naif Alatawi,[1] **Najah Alsubaie,**[2] **Habib Ullah Khan** [iD],[3] **Tariq Sadad** [iD],[4] **Hathal Salamah Alwageed** [iD],[5] **Shaukat Ali** [iD],[6] **and Islam Zada** [iD][7]

[1]*Information Technology Department Faculty of Computing and Information Technology, University of Tabuk, Tabuk, Saudi Arabia*
[2]*Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[3]*Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar*
[4]*Department of Computer Science, University Engineering and Technology, Mardan, Khyber Pakhtunkhwa, Pakistan*
[5]*College of Computer and Information Sciences, Jouf University, Sakakah, Saudi Arabia*
[6]*Department of Computer Science, University of Peshawar, Peshawar, Khyber Pakhtunkhwa, Pakistan*
[7]*Department of Software Engineering, Faculty of Computing & Information Technology, International Islamic University, Islamabad, Pakistan*

Correspondence should be addressed to Islam Zada; islam.zada@iiu.edu.pk

The attacks of cyber are rapidly increasing due to advanced techniques applied by hackers. Furthermore, cyber security is demanding day by day, as cybercriminals are performing cyberattacks in this digital world. So, designing privacy and security measurements for IoT-based systems is necessary for secure network. Although various techniques of machine learning are applied to achieve the goal of cyber security, but still a lot of work is needed against intrusion detection. Recently, the concept of hybrid learning gives more attention to information security specialists for further improvement against cyber threats. In the proposed framework, a hybrid method of swarm intelligence and evolutionary for feature selection, namely, PSO-GA (PSO-based GA) is applied on dataset named CICIDS-2017 before training the model. The model is evaluated using ELM-BA based on bootstrap resampling to increase the reliability of ELM. This work achieved highest accuracy of 100% on PortScan, Sql injection, and brute force attack, which shows that the proposed model can be employed effectively in cybersecurity applications.

1. Introduction

The information technologies (IT) can be applied to fulfill the basics of smart cities. The idea of smart city is implementing in various countries to manage urbanization growth and employ the resources effectively. Moreover, the main aim of smart city is to connect various devices to promote Internet of Things (IoT) and to perform fast and accurate communication in the modern world [1]. IoT device used sensor to obtain real-time data from another object. Internet is the main source of communication for IoT devices, which makes them available all the time. IoT devices are contributed in the modern society and almost used in every field such as military, transport, education, agriculture, healthcare, and commerce as presented in Figure 1. IoT is working on approved protocols for communication exchange [2], but due to its diverse domains of appliances leads to the realization of several communication standards, devices, and protocols. IoT devices using the real-world data acquired from the sensors, which can further be employed to make an intelligent system. However, IoT devices can be protect against cyberattacks, and intelligent techniques of

intrusion detection system (IDS) must be applied before deployment in any organization.

Computing resources are protected from external threats by a computer security program to maintain their confidentiality, integrity, and availability. A network intrusion poses a risk to the resources of the victim server and the network as a whole [3]. System administrators can react to intrusions when they are identified by the intrusion detection system (IDS). People's distrust of the Internet has grown in tandem with the frequency of hacks. A well-executed security assault is a denial of service (DoS).

A company's computer network can be attacked from the inside or the outside using an IDS. It is important to realize that intrusion detection systems differ from burglar alarms despite their similarities. In this article, we describe how to detect and classify intrusions into agricultural Internet of Things networks. Not just in agriculture IoT networks, but throughout all Internet of Things applications, security and privacy are fundamental concerns.

1.1. Background of Intrusion Detection System. Detecting malicious activity on a network is a crucial element of intrusion detection systems (IDS) [4] . Software that detects harmful activities or actions might violate regulatory rules. A security information and event management (SIEM) system normally alert the administrator to any malicious activity or breach. To distinguish between true and false alerts, SIEM architectures combine data from various sources and use alert filtering algorithms. However, intrusion detection systems are susceptible to false alarms, as they monitor networks for suspicious activity. So, companies must fine-tune their IDS devices upon deployment. The system should distinguish legitimate network traffic from malicious activity by properly setting up intrusion prevention systems. Network packets entering the device are also monitored by intrusion detection systems to detect abnormal activity and send alerts.

There are four types of intrusion detection systems.

1.1.1. Network Intrusion Detection System (NIDS). Systematic analysis of multiple network devices is made possible by network intrusion detection systems (NIDS). A database of known attacks is used to track all subnet traffic. Any intrusion or suspicious behavior will be notified to the administrator. The goal of a NIDS is to detect attempts to breach firewalls on the subnet where they are installed.

1.1.2. Host Intrusion Detection System (HIDS). A host intrusion detection system (HIDS) detects and alerts the administrator when it detects suspicious or disruptive activity on a server. A HIDS measures only transmitted data and can detect threats over a network. Software compares the current state of the device's files with those on the most recent backup. Changes or losses of analytical system files are notified to the administrator so that he can inspect them. Devices that are unlikely to change their settings, such as mission-critical devices, can be equipped with HIDS.
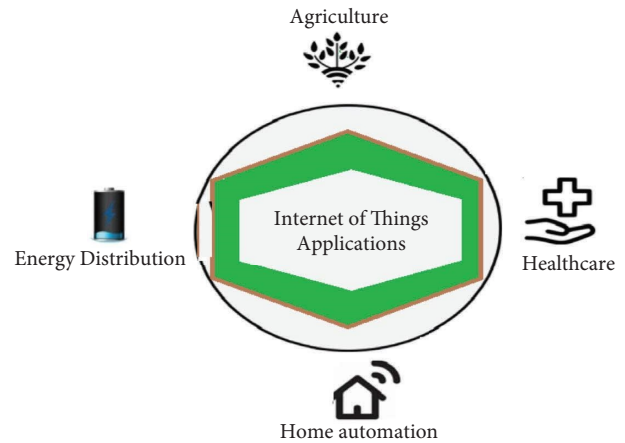

Figure 1: IoT deployment.

1.1.3. Protocol-Based Intrusion Detection System (PIDS). By accepting the corresponding HTTP protocol and managing the HTTPS stream regularly, the application seeks to keep the web server safe. Because HTTPS is not secure, the device must remain within this interface before it can proceed to the web presentation layer.

1.1.4. Application Protocol-Based Intrusion Detection System (APIDS). APIDS (application protocol-based intrusion detection systems) is a device or a set of agents that reside on a collection of servers. APIDS analyzes traffic between servers based on application-specific protocols to detect intrusions. By using this, for instance, the middleware can monitor the SQL communication from the webserver to the database.

1.2. Motivation. Our digital era is full of internet-connected objects. We rely significantly on these technologies to meet our daily demands. This will significantly increase the security and intrusion risks on these systems. The study on intrusion detection systems covers a wide range of machine learning approaches. It is still difficult for existing IDS to increase detection rates, reduce false positives, and identify unknown intrusions. Scholars have investigated how machine learning can be incorporated into IDSs to deal with existing issues. By using hybrid-based machine learning algorithms, the difference between normal and abnormal data can be automatically determined. A hotbed of research, hybrid learning has resulted in remarkable breakthroughs.

2. Literature Review

IoT devices are at high risk due to the increase ratio of cyberattacks, and recently, it required more attention. In literature, several solutions are proposed with the help of machine learning and deep learning to prevent and identify these attacks [3, 4]. Some well-known methods such as SVM, KNN, decision tree, ensemble methods, and CNN are used for classification [3]. For example, the authors employed autoencoders algorithms for online intrusion

detection [7]. NSL KDD data are used as input data, and it can be accessed online [8]. To preprocess the NSL-KDD data, all symbols are converted into numeric characteristics, and then, they are converted back into symbolic features. The principal component analysis method is used to extract characteristics. In this study, machine learning algorithms are compared on their accuracy, precision, and recall when used to classify preprocessed data. Support vector machines, linear regressions, and random forests are used as machine learning algorithms[9]. The authors used ANN for the detection of network intrusion [10]. In [11], the authors employed a hybrid method of feature selection before classification and decreased the false alarm rate. The authors applied an ensemble of ANN for multiclass intrusion detection and achieved 94.96% accuracy using KDD99 dataset [12]. The authors in [13] proposed productive IDS through deep learning for Internet of Medical Things (IoMT) networks. In [14], the authors used improved Seagull optimization algorithm (SOA) for feature selection followed by recurrent neural network (RNN) classifier to detect cyberattacks and obtained 94.12% accuracy using the KD-cup 99 dataset. Liu et al. [15] used CNN for feature extraction followed by MLP to detect the behavior of normal and abnormal user using KDD 99 dataset. The authors proposed DNN-based IDS system [16]. They claimed that DNN with antirectifier layer provide better results compared to others machine learning classifiers. The model was evaluated using various dataset such as UNSW_NB-15, NSL-KDD, and CIC-IDS-2017 dataset. In [17], the authors proposed network anomaly detection system using UNSW-NB15 dataset. The model was tested on various classifiers and achieved classification accuracy of 87.37% and 99.94% for worms class through reduced error pruning tree (REPTree). In [18], the authors proposed ensemble model using meta-classification technique for reliable predictions. The model was evaluated on two datasets called UNSW-NB15 and UGR'16 dataset and achieved 94.27% and 82.22% accuracy, respectively. Similarly, in [19], the authors applied several machine learning models using voting classifier and accomplished an accuracy of 99.7%.

It is clear from the literature that there is required some more effective models to cover the challenges of advance cyberattacks in the IoT domains. Moreover, ensemble methods of learning can increase the efficacy of ML-based IDS, because it provides better results of detection accuracy [20].

The main contribution of this article is as follows:

(i) A recent standard dataset is utilized and used

(ii) A novel feature selection strategy based on PSO-GA is proposed

(iii) The model is evaluated using various ELM models using bootstrap resampling

3. Proposed Method

Before implementing any hybrid-based ML technique, the feature selection methods are employed, namely, PSO-GA to select the optimum feature set. The flow diagram of the proposed IDS model is portrayed in Figure 2.

3.1. Dataset. The most defensive tools against ever-growing and sophisticated network attacks are IDS and intrusion prevention system (IPS). Anomaly based IDS suffers from the accurate performance development due to the lack of trustworthy/reliable test and validation datasets. Thus, we employed a benchmark dataset called CICIDS-2017 [21], which included denial-of-service (DoS), distributed denial-of-service (DDoS), brute force attack, web attack, botnet, infiltration, and PortScan [22, 23] presented, and the number of features are presented in Tables 1 and 2.

3.2. Features Selection. Features selection finds optimum range of features from the main data, which can effectively choose input data while reducing computational cost.

In this article, we proposed a hybrid based method for feature selection called PSO-GA. Particle swarm optimization (PSO) is a filtering processes and efficient method for feature subselection [24]. The local search competence of PSO is strong but that it cannot accomplish sufficient exploration. PSO is mostly stuck in local optima that stop the proficiency to explore further. PSO is unable to control the number of search features [25], and also, features' correlation knowledge is not using in the PSO-based method [24]. Genetic algorithm (GA) using the function of crossover, which can do an amazing exploration of the search space. However, it does not have capability to take advantage of that [25]. Thus, the benefit GA and PSO can be employed to become PSO-GA for effective and usable results.

In the proposed PSO-GA, exploring and exploiting is performed in a balance way [26]. PSO is thoroughly exploring the search space of the related particles with each other, while GA is effective for transmitting the valuable functions from production to production [20, 27].

3.3. Extreme Learning Machine Based on Bootstrap Aggregated (ELM-BA). ELM is a type of feed-forward neural network using single hidden layer mostly applied for classification and regression problems [28]. The training of ELM differs from conventional neural network, as it does not support back-propagation based on gradient. It eliminates all the restriction for biases and weights updates. ELM focuses on accomplishing the minimum ration of training error, and weight standards are also lowest to make this model more accurate. The ELM model produces the following output:

$$y_q = \sum_{p=1}^{n} \beta_k a\left(w_p x_k + b_p\right) k = 1, 2, 3, \ldots \ldots f, \qquad (1)$$

where n signifies the number of hidden neurons, a represents the activation function, b_p is used for bias value, w_p denotes vector of the input layer, β_k is used for output layer
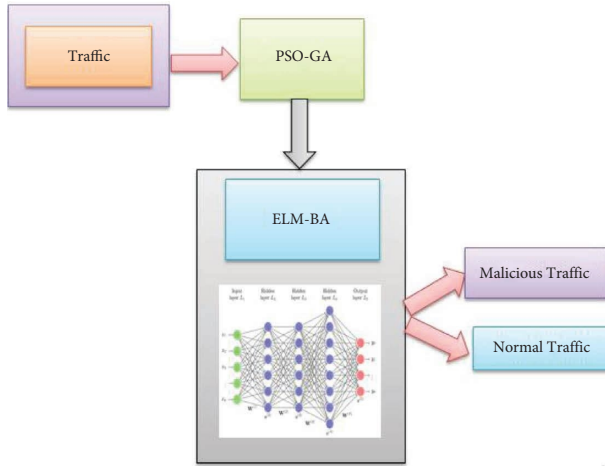
Figure 2: Proposed model.

Table 1: Dataset labels.

Normal	DDoS	BOT
Brute force	XSS	Sql injection
DoS hulk	DoS hulk	DoS slowhttptest
FTP-patator	DoS slowloris	Heartbleed
SSH-patator	Infiltration	PortScan
DoS GoldenEye		

Table 2: Number of features.

No	Feature name
1	Bwd IAT std
2	Bwd IAT max
3	Bwd IAT min
4	Fwd PSH flags
5	Total length of fwd packets
6	Total length of bwd packets
7	Fwd packet length max
8	Fwd packet length min
9	Fwd packet length mean
10	Fwd packet length std
11	Bwd packet length max
12	Bwd packet length min
13	Bwd packet length mean
14	Bwd packet length std
15	Init_Win_bytes_forward
16	Init_Win_bytes_backward
17	act_data_pkt_fwd
18	min_seg_size_forward
19	Active mean
20	Active std
21	Active max
22	Active min
23	Idle mean
24	Idle std
25	Idle max
26	Idle min
27	Flow Bytes/s
28	Flow Packets/s
29	Flow IAT mean

Table 2: Continued.

No	Feature name
30	Flow IAT std
31	Bwd PSH flags
32	Fwd URG flags
33	Bwd URG flags
34	Fwd header len
35	Bwd header length
36	Fwd packets/s
37	Bwd packets/s
38	Min packet length
39	Max packet length
40	Packet length mean
41	Packet length std
42	Packet length variance
43	FIN flag count
44	SYN flag count
45	Destination port
46	Flow duration
47	Total fwd packets
48	Total backward packets
49	URG flag count
50	CWE flag count
51	Flow IAT max
52	Average packet size
53	AvgFwd segment size
54	AvgBwd segment size
55	Fwd header length
56	FwdAvg bytes/bulk
57	FwdAvg packets/bulk
58	FwdAvg bulk rate
59	BwdAvg bytes/bulk
60	BwdAvg packets/bulk
61	BwdAvg bulk rate
62	SubflowFwd packets
63	SubflowFwd bytes
64	SubflowBwd packets
65	SubflowBwd bytes
66	RST flag count
67	PSH flag count
68	ACK flag count
69	Flow IAT min
70	Fwd IAT total
71	Fwd IAT mean
72	Fwd IAT std
73	Fwd IAT max
74	Fwd IAT min
75	ECE flag count
76	Down/up ratio
77	Bwd IAT total
78	Bwd IAT mean

according to the kth hidden neuron, and f is utilized for the number of features

In this manuscript, ELM-BA is proposed to increase the accuracy and reliability of ELM where various ELM models are get trained using bootstrap resampling [28].

The ELM-BA is computed as

$$E(v) = \sum_{k=1}^{n} w_k p_k(v), \qquad (2)$$

TABLE 3: Accuracy against each attack.

Traffic	Accuracy
Normal	99.96
Bot	99.93
DDoS	92.60
FTP-patator	99.98
SSH-patator	90.12
PortScan	100
Heartbleed	99.80
Sql injection	100
Brute force	100
DoS hulk	93.25
DoS goldenEye	95.23
DoS slowhttptest	85.55
DoS slowloris	89.96
Infiltration	99.41
XSS	98.76
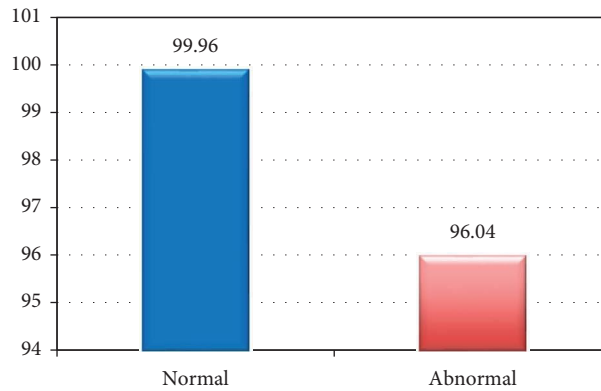
FIGURE 3: Normal vs. abnormal traffic.

TABLE 4: Comparison with state of the art.

References	Methods	Datasets	Accuracy
[29]	Deep learning	NSL-KDD	86.9
[20]	Deep random neural network	UNSW-NB15	99.5
Proposed	PSO-GA followed by ELM-BA	CICIDS-2017	99.96% (for normal) and 96.04% (for abnormal)

where $E(x)$ represents aggregated forecaster of the neural network, v represents vector of input neural network, n is the number of neural networks that are fused $p_k(v)$ used for k^{th} neural network, and w_k aggregated weight for combining k^{th} neural network

4. Performance Analysis

The proposed model is evaluated using different parameters such as true positive ($T+$), true negative ($T-$), false positive ($F+$), and false negative ($F-$) were calculated, and then, accuracy is calculated as follows:

$$\text{Accuracy} = \left\{ \frac{(T+) + (T-)}{(T+) + (F-) + (F+) + (T-)} \right\}. \quad (3)$$

5. Results and Discussion

The process of experimentation is carried out to detect normal and abnormal traffic. For this purpose, optimum features are chosen using PSO-GA, and then, ELM-BA model is used to train multiple ELM models using bootstrap aggregation to achieve better classification. We trained the ELM model using 100, 150, and 200 numbers of hidden neurons and then aggregated to achieve better results.

5.1. Analysis of ELM Models. The ELM model is trained using various ways and then aggregated the model. The number of hidden layer is chosen 100, 150, and 200, which are then finally aggregated. Table 2 provides the summarized result of accuracy. Table 3 reported the individual accuracy

of each label and proved that ELM-BA perform outstanding result. For example, PortScan, Sql injection, and brute force achieved 100% classification accuracy, while normal data is obtained 99.96% accuracy.

The efficacy of proposed model is further demonstrated in Figure 3, and the obtained results of abnormal attack are aggregated and obtained 96.04% accuracy. The chart clearly demonstrates that an obtained result of the proposed model is remarkable.

The proposed work is also compared with some existing works done for cyber security and is stated in Table 4. The proposed work achieved highest accuracy as illustrated in Table 4.

6. Conclusion

IoT-based systems facilitate users to retrieve their data smoothly, but on the contrary, it gives an insecure atmosphere so that security can be comprised. This research work provides intrusion detection model based on ensemble learning. Features are selected using evolutionary and swarm intelligence called PSO-GA followed by ELM-BA algorithm. The proposed method gives assurance to reveal all kinds of attacks. It presents noteworthy accuracy with ensemble model of feature selection and classification. Proposed model is evaluated on state of the art dataset called CICIDS-2017 and achieved 99.96% and 96.04% accuracy of normal and abnormal attack, respectively. The model will be evaluated on more datasets with advance techniques of deep learning in future.

Data Availability

The data used during the study for experiment is available online at http://www.unb.ca/cic/datasets/ids-2017.html.

Consent

Not applicable.

Disclosure

Research involves human participants and/or animals. No studies involving human participants or animals were performed by the authors for this article.

Conflicts of Interest

All the authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed equally to the work.

Acknowledgments

References

[1] T. Saba, "Intrusion detection in smart city hospitals using ensemble classifiers," in *Proceedings of the 2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 418–422, Liverpool, UK, December 2020.

[2] H. Shi, L. Zhai, H. Wu, M. Hwang, K. S. Hwang, and H. P. Hsu, "A Multitier reinforcement learning model for a cooperative multiagent system," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 636–644, 2020.

[3] T. Saba, T. Sadad, A. Rehman, Z. Mehmood, and Q. Javaid, "Intrusion detection system through advance machine learning for the internet of things networks," *IT Professional*, vol. 23, no. 2, pp. 58–64, 2021.

[4] T. Alhakami, A. ALharbi, T. Bourouis, H. Alroobaea, and S. A. Bouguila, "Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection," *IEEE Access*, vol. 7, pp. 52181–52190, 2019.

[5] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, "Anomaly-based intrusion detection system for IoT networks through deep learning model," *Computers and Electrical Engineering*, vol. 99, Article ID 107810, 2022.

[6] T. Sadad, A. Rehman, A. Hussain, A. A. Abbasi, and M. Q. Khan, "A review on multi-organ cancer detection using advanced machine learning techniques," *Current Medical Imaging*, vol. 17, no. 6, pp. 686–694, 2021.

[7] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," 2018, https://arxiv.org/abs/1802.09089.

[8] C. W. Tsai, T. P. Hong, and G. N. Shiu, "Metaheuristics for the lifetime of WSN: a review," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 2812–2831, 2016.

[9] S. Latif, Z. Idrees, Z. Zou, and J. A. Drann, "A deep random neural network model for intrusion detection in industrial iot," in *Proceedings of the 2020 International Conference on UK-China Emerging Technologies (UCET)*, pp. 1–4, IEEE, Glasgow, UK, August 2020.

[10] Y. R. Zhao, Y. Liu, D. Wang, W. R. Lv, and J. L. Zhou, "An ANN based sequential detection method for balancing performance indicators of IDS," in *Proceedings of the 2019 Seventh International Symposium on Computing and Networking (CANDAR)*, vol. 51, no. 2, pp. 239–244, Nagasaki, Japan, November 2019.

[11] N. Moustafa and J. Slay, "A hybrid feature selection for network intrusion detection systems: central points," 2017, https://arxiv.org/ftp/arxiv/papers/1707/1707.05505.pdf.

[12] M. M. Baig, M. M. Awais, and E. S. M. El-Alfy, "A multiclass cascade of artificial neural network for network intrusion detection," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 4, pp. 2875–2883, 2017.

[13] S. P. Rm, P. K. R. Maddikunta, M. Parimala et al., "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," *Computer Communications*, vol. 160, pp. 139–149, 2020.

[14] A. A. Ewees, R. R. Mostafa, R. M. Ghoniem, and M. A. Gaheen, "Improved seagull optimization algorithm using Lévy flight and mutation operator for feature selection,"

Neural Computing and Applications, vol. 34, no. 10, pp. 7437–7472, 2022.

[15] Y. Liu, S. Liu, and X. Zhao, "Intrusion detection algorithm based on convolutional neural network," in *Proceedings of the 2020 International Conference on Computer Engineering and Application (ICCEA)*, Guangzhou, China, March, 2018.

[16] R. Lohiya, A. Thakkar, and A. Thakkar E-Mail, "Intrusion detection using deep neural network with antirectifier layer," *Lecture Notes in Networks and Systems*, vol. 187, pp. 89–105, 2021.

[17] A. Roy and K. J. Singh, "Multi-classification of UNSW-NB15 dataset for network anomaly detection system BT," *Proceedings of International Conference on Communication and Computational Technologies*, pp. 429–451, Springer, Singapore, 2021.

[18] S. Rajagopal, P. P. Kundapur, and K. S. Hareesha, "A stacking ensemble for network intrusion detection using heterogeneous datasets," *Security and Communication Networks*, vol. 2020, Article ID 4586875, 9 pages, 2020.

[19] T. Saba, A. R. Khan, T. Sadad, and S. P. Hong, "Securing the IoT system of smart city against cyber threats using deep learning," *Discrete Dynamics in Nature and Society*, vol. 2022, Article ID 1241122, 9 pages, 2022.

[20] Y. Li, Y. Xu, Z. Liu et al., "Robust detection for network intrusion of industrial iot based on multi-cnn fusion," *Measurement*, vol. 154, Article ID 107450, 2020.

[21] S. Singh Panwar, Y. P. Raiwani, and L. S. Panwar, "Evaluation of network intrusion detection with features selection and machine learning algorithms on cicids-2017 dataset," *SSRN Electronic Journal*, vol. 2019, 2019.

[22] J. Krupp, M. Backes, and C. Rossow, "Identifying the scan and attack infrastructures behind amplification DDoS attacks," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, vol. 24-28, pp. 1426–1437, Vienna, Austria, October 2016.

[23] S. Bourouis, Y. Pawar, and N. Bouguila, "Entropy-based variational scheme with component splitting for the efficient learning of gamma mixtures," *Sensors*, vol. 22, no. 1, p. 186, 2021.

[24] Y. Xue, A. Aouari, R. F Mansour, and S. Su, "A hybrid algorithm based on PSO and GA for feature selection," *Journal of Cyber Security*, vol. 3, no. 2, pp. 117–124, 2021.

[25] M. Poongodi, S. Bourouis, A. N. Ahmed et al., "A novel secured multi-access edge computing based VANET with neuro fuzzy systems based blockchain framework," *Computer Communications*, vol. 192, pp. 48–56, 2022.

[26] S. A. Changazi, A. D. Bakhshi, M. Yousaf et al., "GA-based geometrically optimized topology robustness to improve ambient intelligence for future internet of things," *Computer Communications*, vol. 193, pp. 109–117, 2022.

[27] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Mathematical Problems in Engineering*, vol. 2015, Article ID 931256, 38 pages, 2015.

[28] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review*, vol. 44, 2015.

[29] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1982.