Computers, Materials & Continua
DOI: 10.32604/cmc.2022.031887
Article

Tech Science Press

Feature Selection with Stacked Autoencoder Based Intrusion Detection in Drones Environment

Heba G. Mohamed[1], Saud S. Alotaibi[2], Majdy M. Eltahir[3], Heba Mohsen[4], Manar Ahmed Hamza[5,*], Abu Sarwar Zamani[5], Ishfaq Yaseen[5] and Abdelwahed Motwakel[5]

[1]Department of Electrical Engineering, College of Engineering, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[2]Department of Information Systems, College of Computing and Information System, Umm Al-Qura University, Saudi Arabia
[3]Department of Information Systems, College of Science & Art at Mahayil, King Khalid University, Saudi Arabia
[4]Department of Computer Science, Faculty of Computers and Information Technology, Future University in Egypt, New Cairo, 11835, Egypt
[5]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Manar Ahmed Hamza. Email: Ma.hamza@psau.edu.sa

Received: 29 April 2022; Accepted: 02 June 2022

Abstract: The Internet of Drones (IoD) offers synchronized access to organized airspace for Unmanned Aerial Vehicles (known as drones). The availability of inexpensive sensors, processors, and wireless communication makes it possible in real time applications. As several applications comprise IoD in real time environment, significant interest has been received by research communications. Since IoD operates in wireless environment, it is needed to design effective intrusion detection system (IDS) to resolve security issues in the IoD environment. This article introduces a metaheuristics feature selection with optimal stacked autoencoder based intrusion detection (MFSOSAE-ID) in the IoD environment. The major intention of the MFSOSAE-ID technique is to identify the occurrence of intrusions in the IoD environment. To do so, the proposed MFSOSAE-ID technique firstly pre-processes the input data into a compatible format. In addition, the presented MFSOSAE-ID technique designs a moth flame optimization based feature selection (MFOFS) technique to elect appropriate features. Moreover, firefly algorithm (FFA) with stacked autoencoder (SAE) model is employed for the recognition and classification of intrusions in which the SAE parameters are optimally tuned with utilize of FFA. The performance validation of the MFSOSAE-ID model was tested utilizing benchmark dataset and the outcomes implied the promising performance of the MFSOSAE-ID model over other techniques with maximum accuracy of 99.72%.

Keywords: Internet of drones; unmanned aerial vehicles; security; intrusion detection; machine learning

1 Introduction

The Internet of Drones (IoD) is a layered network control structure devised for organizing the accessibility of several Unmanned Aerial Vehicles (UAVs) (called drones) for controlling airspace and offering navigation facilities [1]. The IoD renders several amenities for drone applications extending from military to civilian application areas, involving rescue and search operations, aerial surveys delivery services, environment, and traffic monitoring. The IoD is being familiar day after day due to its services and guidance to the community [2]. Every drone is fortified with computing unit, transmission module, recorder, sensors and actuators, and energy supply unit [3]. The default sensor nodes in the drone mean the physical criteria, namely concentration and temperature of harmful gases, and the camera which is built-into the UAV captures the images or records videos of the target and forwards every sensing monitored data to drone box through few wireless transmission technologies, like WiFi [4,5]. Every UAV box is linked to the server, server that is linked to control room. The UAVs are microcosm of the Internet of Things (IoT). For technology leaders and businesses to understand the capability of drones in their institutions, a quantity of general IoT-based product engineering difficulties occurs to beat these [6]. Fig. 1 depicts the applications of IoD.

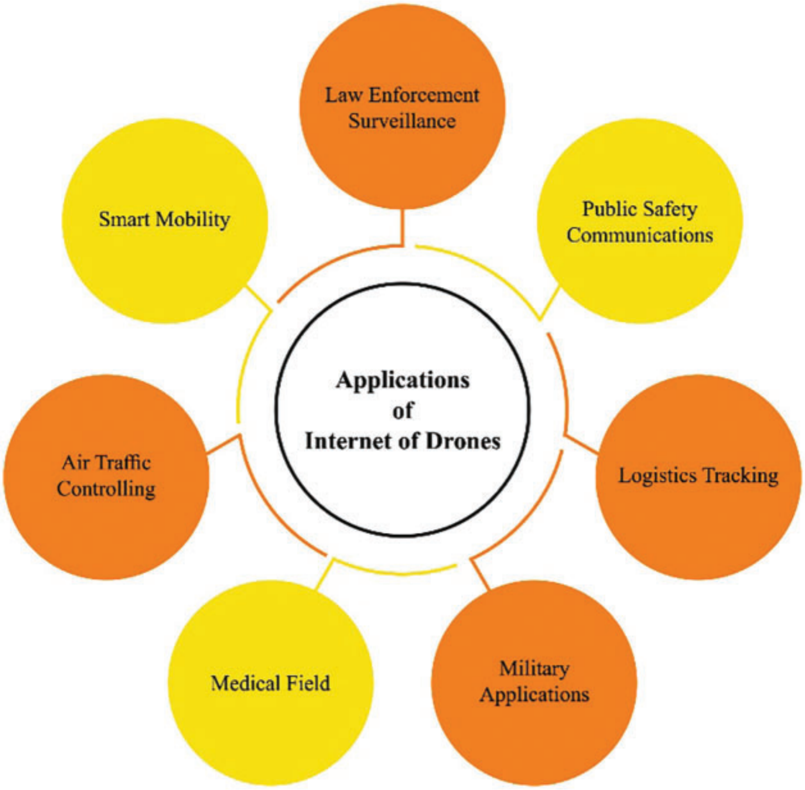

Law Enforcement Surveillance
Public Safety Communications
Logistics Tracking
Military Applications
Medical Field
Air Traffic Controlling
Smart Mobility
Applications of Internet of Drones

Figure 1: Applications of IoD

Currently, network security is becoming a crucial research field, specifically after advancing transmission and Internet methods. It makes use of various tools like Network Intrusion Detection Systems (NIDSs) and firewalls for security purposes of assets and networks [7]. For monitoring the network traffic for suspicious and worse conduct permanently NIDSs are commonly utilized. IDSs can be divided into three type's signature-related specification-related or anomaly-related, methods [8]. Anomaly-related approaches can be again categorized as statistical-related, knowledge-related, or machine learning-related [9]. The attack is the distraction among the 2 statistical paradigms they are the current captured and the normal memorized. In knowledge-related anomaly recognition, specialists present more rules in the standard of a fuzzy-related system or expert system for explaining the conduct of usual attacks and connections. The rule-related system is linked to input unit in fuzzy-related anomaly recognition [10]. On the basis of the input data, a subset of the rules is started. In a machine learning-related anomaly IDS, an implicit or explicit algorithm of the observed outlines is formed. Depending on earlier outcomes, these methods are upgraded on a day-to-day basis for enhancing intrusion detection efficiency.

Ouiazzane et al. [11] presented a model dependent upon Muli-Agent System and on ML approaches for detecting Denial of Service (DoS) cyberattacks aiming at the network of drones. The presented method was autonomous, considered by their higher performance, and allows the recognition of known and unknown DoS attacks from UAV networks with higher accuracy and minimal false negative rate (FNR) and false positive rate (FPR). Condomines et al. [12] present a novel IDS, a hybrid technique dependent upon both spectral traffic analysis and robust controlling or observing to anomaly estimate inside UAV network. The presented hybrid technique assumes, as an initial step, a statistical signature of traffic replaced from the networks. With investigative the resulted signature, the variances were utilized for selecting the accurate method to accurate estimation abnormal traffics. Perumalla et al. [13] present a novel approach for secure broadcast from IoD by effectual blockchain (BC) supported access control and ID technique utilizing the recently devised Deep Neuro-Fuzzy Network model. Classically, BC based access control contains mostly 4 stages such as authentication, pre-deployment, access control, and registration stages to transfer the vital data from the IoD environments. Sharma et al. [14] execute the novel functional encryption (FE) approach from the presented UAV supported HetNet method for the dense urban area for securing data against such intrusion. During this network method, UAV performances as relay node to individuals UE that are in non-line-of-sight (NLoS) transmission with macro based station (MBS). Ouiazzane et al. [15–18] present novel methods of IDS to a fleet of UAVs used with ad hoc communication structures. The security of drone fleet was rarely discussed by the scientific community, and one of the research is concentrated on routing protocol and battery autonomy, but ignored the security feature.

This article introduces a metaheuristics feature selection with optimal stacked autoencoder based intrusion detection (MFSOSAE-ID) in the IoD environment. The proposed MFSOSAE-ID technique firstly pre-processes the input data into a compatible format. In addition, the presented MFSOSAE-ID technique designs a moth flame optimization based feature selection (MFOFS) technique to elect appropriate features. Moreover, firefly algorithm (FFA) with stacked autoencoder (SAE) model is employed for the recognition and classification of intrusions in which the SAE parameters are optimally tuned by the use of FFA. The performance validation of the MFSOSAE-ID model is tested using benchmark dataset.

2 The Proposed Model

In this study, a new MFSOSAE-ID technique has been developed to determine the existence of intrusions in the IoD environment. The proposed MFSOSAE-ID technique firstly pre-processes the input data into a compatible format. Followed this, the presented MFSOSAE-ID technique designs an MFOFS technique to elect appropriate features. In line with, the FFA-SAE model is employed for the recognition and classification of intrusions. Fig. 2 illustrates the overall process of MFSOSAE-ID technique.

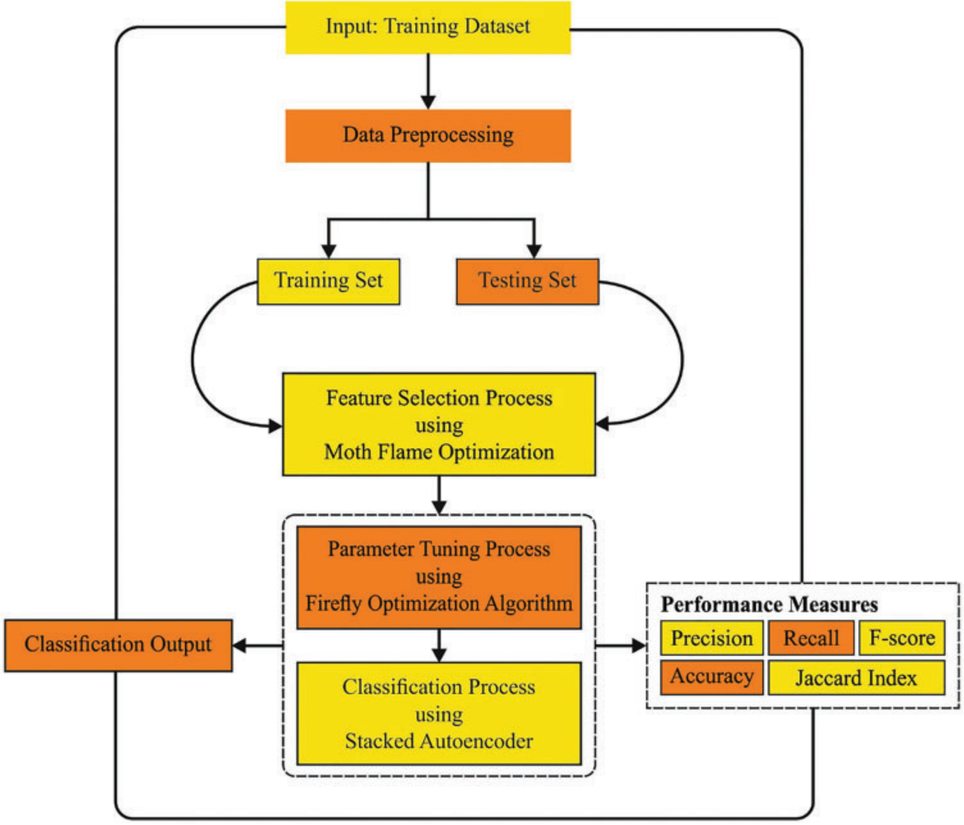

Input: Training Dataset
Data Preprocessing
Training Set
Testing Set
Feature Selection Process using Moth Flame Optimization
Parameter Tuning Process using Firefly Optimization Algorithm
Classification Process using Stacked Autoencoder
Classification Output
Performance Measures
Precision
Recall
F-score
Accuracy
Jaccard Index

Figure 2: Overall process of MFSOSAE-ID technique

2.1 Data Pre-Processing

The proposed MFSOSAE-ID technique firstly pre-processes the input data into a compatible format. In a primary phase, the z-score normalized system was developed that estimates the standard deviation (SD) and arithmetic mean of offered gene information. The normalized score reaches were offered in the subsequent:

$$s_k^{'} = \frac{s_k - \mu}{\sigma} \tag{1}$$

whereas σ signifies the SD and μ stands for the arithmetic mean of offered that data. During this case, the normalized of smoothed data has been executed using Z-score normalized.

2.2 Steps Involved in MFOFS Technique

Next to data pre-processing, the presented MFSO. The MSO procedure is based on the procedure of selecting optimum discriminatory characteristics of mitosis cells. Usually, Moth is a bug which belongs to the family of butterflies named Lepidoptera. About 160,000 moth classes appear worldwide, and it is noticeable during nighttime. With the comparison of moth and alternative species, Levy flight (LF) and phototaxis are found to be significant.

Phototaxis

The fundamental concept of the moth is that; it frequently moves toward the light called phototaxis. There is no precise technique for phototaxis and is comprised of the primary hypotheses to determine the phototaxis method. After that, it is intended to be a substantial hypothesis in the celestial orientation that is implemented in transverse direction during the flying period. Simultaneously, the angle is initiated by the source of light and the moth is directed, but, it is not relevant to obtain the transformation, since the celestial object is considered an external distance. Therefore, it travels to the light, because the moth creates flight direction for gaining the optimal position [19]. A spiral-path was established to approach the light.

Levy flight

Heavy-tailed, non-Gaussian statistics are inspected in traditional ways in variety of applications. On the other hand, Drosophila signifies a Levy flight, but, the flight is evaluated by energy source with features exponent with 3/2. In general, the process can be formulated by,

$$L\left(s\right) \sim |s|^{-\beta} \tag{2}$$

Now $1 < \beta \leq 3$ represents an index.

Fundamentally, moths have the property close to optimum one and fly in an LF method. Otherwise, it instructed the location and employs LF using the below formula, but, moth is maintained by the following equation:

$$X_j^{a+1} = X_j^a + \delta L\left(s\right) \tag{3}$$

Here, X_j^a and X_j^{a+1} indicates the original and refreshed position at generation a, but the present generation is referred to a. It is achieved by LF as denoted as $L\left(s\right)$. In problem of interest, scaling factor determines the parameter δ. Therefore, δ is represented as follows:

$$\delta = \frac{W_{\max}}{a^2} \tag{4}$$

Next, $W_{\max}$ represents maximal walk step and value for $W_{\max}$. $L\left(s\right)$ from abovementioned formula is rehabilitated by,

$$L\left(s\right) = \frac{\left(\alpha - 1\right)\Gamma\left(\alpha - 1\right)\sin\left(\frac{\pi\left(\alpha-1\right)}{2}\right)}{\pi s^{\alpha}} \tag{5}$$

In s is found to be high when compared to zero. $\Gamma\left(x\right)$, denotes a gamma operation. In previous to allocate from $L\left(s\right)$ with $\alpha = 1.5$ moth LF was initiated.

Learning feature presentation from each three dimensional is considerable for biomarker predictive process from volumetric medical dataset. Nevertheless, it is predictable to employ three dimensional convolutional kernels, from the point of encoded volumetric and spatial datasets. In the comparable volumetric image size of $X \times Y \times Z$, whereas utilizing three dimensional convolutional kernels to generate a three dimensional feature volume. At last, a three dimensional kernel was established and swept over the three dimensional topology. By decreasing the kernel distribution through three dimensional, the network employs the whole advantage of volumetric context dataset. The formula evaluates the applied three dimensional convolutional tasks from a component-wise method as follows:

$$u_{ki}^{l}\left(x, y, z\right) = \sum_{m,n,t} h_k^{l-1}\left(x - m, y - n, z - t\right) W_{ki}^{l}\left(m, n, t\right) \tag{6}$$

Here, W_{ki}^{l} indicates a three dimensional kernel from l-th layer which convolve through three dimensional feature volume h_k^{l-1} and $W_{ki}^{l}(m, n, t)$ represents the component-wise weights. In the application of Eq. (7), the three dimensional feature volume h_i^l can be accomplished by different three dimensional convolutional kernels as follows:

$$h_i^l = \sigma\left(\sum_k u_{ki}^l + b_i^l\right). \tag{7}$$

Fly straight

The existing moth is detached from light that flusters in straight line towards light. The function of moth j is developed by,

$$X_j^{a+1} = \lambda \times \left(X_j^a + \varphi\left(X_{best}^a - X_j^a\right)\right) \tag{8}$$

In the Eq. (8), X_{best}^a determines the optimum moth at a generation, λ designates a scaling factor. φ signifies accelerating factor.

Otherwise, moth moves out from light source to concluding location. Therefore, the last position of moth j is given as follows,

$$X_j^{a+1} = \lambda \times \left(X_j^a + \frac{1}{\varphi}\left(X_{best}^a - X_j^a\right)\right) \tag{9}$$

For reliability, the position of moth j is maintained by partial proportion. The best, original, and refreshed position of moth is evaluated by X_j, $X_{j,new}$ and, $X_{best}.\lambda$, attempted the technique with improved diversity of population.

The MFO purposes for discovering an optimum feature subset to offered dataset which is superior classifier accuracy and lesser features. All these 2 indicators are varying effects on classifier accuracy. At this point, it can be combined with a single weight indicator and use the same fitness function as:

$$fitness = \omega_1 \times acc\,(classifier) + \omega_2 \times \left(1 - \frac{s}{p}\right), \tag{10}$$

In which, p exemplifies the total amount of features, and s denotes the quantity of chosen features.

2.3 SAE Based Intrusion Detection

Next to FS process, the SAE model is employed for the recognition and classification of intrusions in the IoD environment [20]. The hidden layer h inside the autoencoder (AE) was capable of generating coding for expressing the input. The infrastructure has 2 parts such as encoding and decoding. The encoding is set for mapping the input data to hidden representations, but the decoding was mentioned that reconstructing an input data. The untagged input dataset $\{x_n\}_{n=1}^{N}$ was provided, whereas $x_n \in R^{m\times 1}$, h_n signifies the hidden encoding vector that is computed in x_n, and $\hat{x}_n$ represents the decoding vector of resultant layer. The encoder procedure was represented as:

$$h_n = f\left(Wx_n + b_1\right), \tag{11}$$

whereas $f(\cdot)$ implies the encoder function, W signifies the weighted matrix of encoding, and the bias vector was represented as b_1.

$$\hat{x}_n = g\left(W^T h_n + b_2\right), \tag{12}$$

In which $g(\cdot)$ represents the decoder function, W^T stands for the weighted matrix of decoding, and b_2 demonstrates the bias vector. For minimizing the error of reconstruction, the parameter of AEs was optimized, and the procedure was demonstrated as:

$$\psi\ (\Theta) = \operatorname{argmin}_{\theta,\theta} \frac{1}{n} \sum_{i=1}^{n} L\left(x^i, \hat{x}^i\right), \tag{13}$$

whereas L signifies the loss function: $L(x, \hat{x}) = \|x - \hat{x}\|^2$.

The AEs are superimposed layer-by-layer for procedure a deep neural network (DNN); i.e., all the hidden layers are obtained as the input of next layer that is progressive still the trained procedure was ended. The introduction of batch normalization (BN) is established as easy primary criterion to train SAE that creates the gradient upgrade in a very shallow path, speeding up the trained method. The labeled signal is bonded with softmax classification, and back propagation (BP) technique was executed for realizing the upgrade of network weighted and fine-tuning of parameters.

2.4 Steps Involved in FFA Based Parameter Optimization

Finally, the SAE parameters are optimally tuned [21–23] by the use of FFA with the fitness function of accomplishing maximum accuracy [24]. FFA was stimulated by the patterns of fireflies (FFs) and flashing behavior in nature. FFA is a meta-heuristic process based on nature and employed in most of the Np-hard and engineering optimization problems. It is a stochastic method. Specifically, random searching is utilized for finding a collection of solutions. At basic levels, the FFA focuses on making solutions within a searching region and chooses the optimal survival solutions. An arbitrary search prevents it from falling into local optimal. In metaheuristic algorithms, exploration refers to determining several solutions within the searching region, whereas exploitation refers to the searching procedure that aims at the optimal nearby solution. The mode and rate of brightness and the time intervals among the transmitted optical signal attract the two sexes to one another. The light intensity I, reduces by rising distance (R) from the light. In the following, the key characteristics of the FFA are given:

a. The FF becomes more attractive and brighter once it arbitrarily moves, and each FF is of the same sex.

b. The FF attractiveness is proportionate to the distance and the brightness of the light. The light absorption coefficient γ computes the decrease in light concentration. Also, the objective value defines the FF luminance.

c. The distance among FFs can be accomplished from Eq. (14), thus $X_{i,k}$ indicates the k-th part of spatial co-ordination and i-th FF.

$$r_{ij} = \sqrt{(xi - xj)^2 + (yi - yj)^2} \tag{14}$$

The FF movement and attractiveness towards a brighter one is defined as follows:

$$X_i = X_i + B_0 e^{rij^2} (X_j - X_j) + a(rand - 1/2) \tag{15}$$

In Eq. (15), a is a randomized parameter, a rand indicates an arbitrary value attained from [0,1], and B denotes attractiveness of the light. The attraction changes determine the variable γ. FFA has two inner loops while going through the n population and one outer loop for t iteration.

Algorithm 1: Pseudocode of FFA

```
Objective function of f(x), whereas $x = (x1, \ldots, xd)^T$
Create primary population of FFs or $xi(i = 1, 2, \ldots, n)$
Determine light intensity of Ii at xi using f(xi)
Determine light absorption co-efficient $\gamma$
While (t < MaxGeneration)
   For i = 1 to n (every n FFs)
     For j = 1 to n (every n FFs)
       If (li > li)
         Move FF i nearby j
         Attraction differs with distance r using exp[-$\gamma r$]
       end if
       Estimate a novel solution and upgrade light intensity
     End for j
   End for i
Raking the FFs and determining the present optimum
End while
```

3 Results and Discussion

The experimental validation of the MFSOSAE-ID model is tested using the benchmark UNSW-NB15 dataset [25]. It includes a set of 82332 samples with 10 class labels. The details related to the dataset are given in Tab. 1. The results are executed for four distinct runs. Fig. 3 highlights four confusion matrices offered by the MFSOSAE-ID model on the test UNSW-NB15 dataset. The figure pointed out that the UNSW-NB15 dataset has proficiently recognized all ten class labels under four distinct runs.

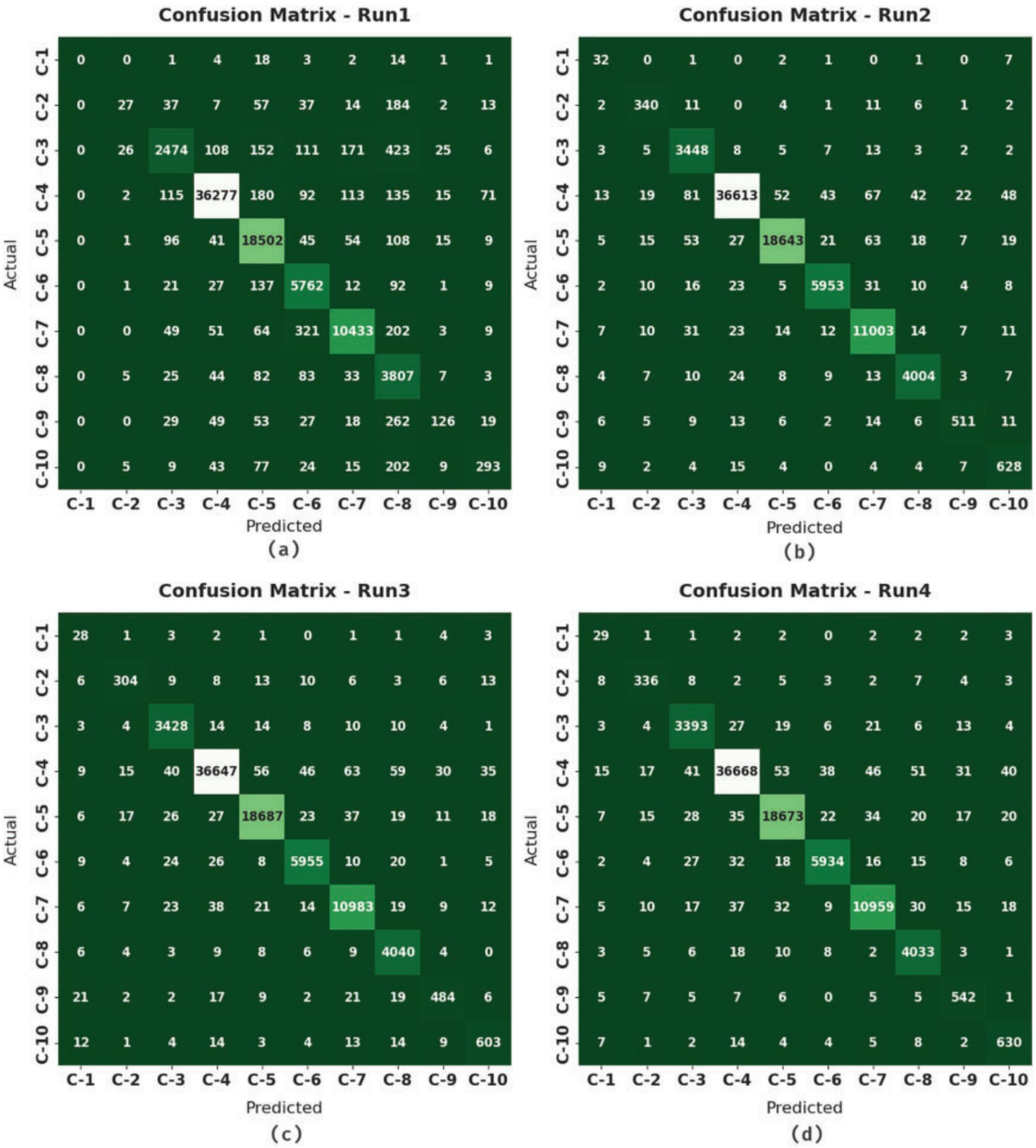
Table 1: Dataset details

Labels	Class names	No. of instances
C-1	Worms	44
C-2	Shell code	378
C-3	Reconnaissance	3496
C-4	Normal	37000
C-5	Generic	18871
C-6	Fuzzers	6062
C-7	Exploits	11132
C-8	DoS	4089
C-9	Backdoor	583

(Continued)

Table 1: Continued

Labels	Class names	No. of instances
C-10	Analysis	677
Total		82332



Actual \ Predicted	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
C-1	0	0	1	4	18	3	2	14	1	1
C-2	0	27	37	7	57	37	14	184	2	13
C-3	0	26	2474	108	152	111	171	423	25	6
C-4	0	2	115	36277	180	92	113	135	15	71
C-5	0	1	96	41	18502	45	54	108	15	9
C-6	0	1	21	27	137	5762	12	92	1	9
C-7	0	0	49	51	64	321	10433	202	3	9
C-8	0	5	25	44	82	83	33	3807	7	3
C-9	0	0	29	49	53	27	18	262	126	19
C-10	0	5	9	43	77	24	15	202	9	293

Actual \ Predicted	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
C-1	32	0	1	0	2	1	0	1	0	7
C-2	2	340	11	0	4	1	11	6	1	2
C-3	3	5	3448	8	5	7	13	3	2	2
C-4	13	19	81	36613	52	43	67	42	22	48
C-5	5	15	53	27	18643	21	63	18	7	19
C-6	2	10	16	23	5	5953	31	10	4	8
C-7	7	10	31	23	14	12	11003	14	7	11
C-8	4	7	10	24	8	9	13	4004	3	7
C-9	6	5	9	13	6	2	14	6	511	11
C-10	9	2	4	15	4	0	4	4	7	628

Actual \ Predicted	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
C-1	28	1	3	2	1	0	1	1	4	3
C-2	6	304	9	8	13	10	6	3	6	13
C-3	3	4	3428	14	14	8	10	10	4	1
C-4	9	15	40	36647	56	46	63	59	30	35
C-5	6	17	26	27	18687	23	37	19	11	18
C-6	9	4	24	26	8	5955	10	20	1	5
C-7	6	7	23	38	21	14	10983	19	9	12
C-8	6	4	3	9	8	6	9	4040	4	0
C-9	21	2	2	17	9	2	21	19	484	6
C-10	12	1	4	14	3	4	13	14	9	603

Actual \ Predicted	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
C-1	29	1	1	2	2	0	2	2	2	3
C-2	8	336	8	2	5	3	2	7	4	3
C-3	3	4	3393	27	19	6	21	6	13	4
C-4	15	17	41	36668	53	38	46	51	31	40
C-5	7	15	28	35	18673	22	34	20	17	20
C-6	2	4	27	32	18	5934	16	15	8	6
C-7	5	10	17	37	32	9	10959	30	15	18
C-8	3	5	6	18	10	8	2	4033	3	1
C-9	5	7	5	7	6	0	5	5	542	1
C-10	7	1	2	14	4	4	5	8	2	630

Figure 3: Confusion matrices of MFSOSAE-ID technique (a) run-1, (b) run-2, (c) run-3, and (d) run-4

Tab. 2 demonstrates the classification results offered by the MFSOSAE-ID model on UNSW-NB15 dataset under run-1. The results implied that the MFSOSAE-ID model has resulted in enhanced classifier results under all ten classes. For instance, with class C-3, the MFSOSAE-ID model has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 98.29%, 86.62%, 70.77%, 77.90%, and 63.80% respectively. In line with, with class C-5, the MFSOSAE-ID methodology has obtainable $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 98.56%, 95.76%, 98.04%, 96.89%, and 93.96% correspondingly. Eventually, with class C-10, the MFSOSAE-ID system has offered $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.36%, 67.67%, 43.28%, 52.79%, and 35.86% correspondingly.

Table 2: Result analysis of MFSOSAE-ID technique with various measures under run-1

Run-1					
Labels	Accuracy	Precision	Recall	F-Score	Jaccard Index
C-1	99.95	0.00	0.00	0.00	0.00
C-2	99.53	40.30	7.14	12.13	6.46
C-3	98.29	86.62	70.77	77.90	63.80
C-4	98.67	98.98	98.05	98.51	97.06
C-5	98.56	95.76	98.04	96.89	93.96
C-6	98.73	88.58	95.05	91.70	84.67
C-7	98.63	96.02	93.72	94.86	90.22
C-8	97.69	70.12	93.10	80.00	66.66
C-9	99.35	61.76	21.61	32.02	19.06
C-10	99.36	67.67	43.28	52.79	35.86
Average	98.88	70.58	62.08	63.68	55.78

Tab. 3 depicts the classification results presented by the MFSOSAE-ID technique on UNSW-NB15 dataset under run-2. The outcomes implied that the MFSOSAE-ID model has resulted in improved classifier results under all ten classes. For instance, with class C-3, the MFSOSAE-ID system has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.68%, 94.10%, 98.63%, 96.31%, and 92.89% correspondingly. Also, with class C-5, the MFSOSAE-ID algorithm has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.60%, 99.47%, 98.79%, 99.13%, and 98.27% correspondingly. Finally, with class C-10, the MFSOSAE-ID algorithm has accessible $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.80%, 84.52%, 92.76%, 88.45%, and 79.29% correspondingly.

Table 3: Result analysis of MFSOSAE-ID technique with various measures under run-2

Run-2					
Labels	Accuracy	Precision	Recall	F-Score	Jaccard Index
C-1	99.92	38.55	72.73	50.39	33.68
C-2	99.87	82.32	89.95	85.97	75.39
C-3	99.68	94.10	98.63	96.31	92.89

(Continued)

Table 3: Continued

Run-2					
Labels	Accuracy	Precision	Recall	F-Score	Jaccard Index
C-4	99.37	99.64	98.95	99.29	98.60
C-5	99.60	99.47	98.79	99.13	98.27
C-6	99.75	98.41	98.20	98.31	96.67
C-7	99.58	98.07	98.84	98.46	96.96
C-8	99.77	97.47	97.92	97.69	95.49
C-9	99.85	90.60	87.65	89.10	80.35
C-10	99.80	84.52	92.76	88.45	79.29
Average	99.72	88.32	93.44	90.31	84.76

Tab. 4 depicts the classification results obtainable by the MFSOSAE-ID approach on UNSW-NB15 dataset under run-3. The outcomes exposed that the MFSOSAE-ID methodology has resulted in higher classifier results under all ten classes. For instance, with class C-3, the MFSOSAE-ID system has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.75%, 96.24%, 98.05%, 97.14%, and 94.44% correspondingly. Besides, with class C-5, the MFSOSAE-ID approach has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.61%, 99.29%, 99.02%, 99.16%, and 98.33% correspondingly. Finally, with class C-10, the MFSOSAE-ID model has offered $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.80%, 86.64%, 89.07%, 87.84%, and 78.31% respectively.

Table 4: Result analysis of MFSOSAE-ID technique with various measures under run-3

Run-3					
Labels	Accuracy	Precision	Recall	F-Score	Jaccard Index
C-1	99.89	26.42	63.64	37.33	22.95
C-2	99.84	84.68	80.42	82.50	70.21
C-3	99.75	96.24	98.05	97.14	94.44
C-4	99.38	99.58	99.05	99.31	98.63
C-5	99.61	99.29	99.02	99.16	98.33
C-6	99.73	98.14	98.23	98.19	96.44
C-7	99.61	98.48	98.66	98.57	97.18
C-8	99.74	96.10	98.80	97.43	94.99
C-9	99.79	86.12	83.02	84.54	73.22
C-10	99.80	86.64	89.07	87.84	78.31
Average	99.72	87.17	90.80	88.20	82.47

Tab. 5 illustrates the classification results offered by the MFSOSAE-ID system on UNSW-NB15 dataset under run-4. The outcomes revealed that the MFSOSAE-ID algorithm has resulted in maximal classifier results under all ten classes. For sample, with class C-3, the MFSOSAE-ID model has accessible $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.71%, 96.17%, 97.05%, 96.61%, and 93.45% respectively. Followed by, with class C-5, the MFSOSAE-ID methodology has offered $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.58%, 99.21%, 98.95%, 99.08%, and 98.18% correspondingly. At last, with class C-10, the MFSOSAE-ID model has provided $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.83%, 86.78%, 93.06%, 89.81%, and 81.50% correspondingly.

Table 5: Result analysis of MFSOSAE-ID technique with various measures under run-4

Run-4					
Labels	Accuracy	Precision	Recall	F-Score	Jaccard Index
C-1	99.91	34.52	65.91	45.31	29.29
C-2	99.87	84.00	88.89	86.38	76.02
C-3	99.71	96.17	97.05	96.61	93.45
C-4	99.39	99.53	99.10	99.31	98.64
C-5	99.58	99.21	98.95	99.08	98.18
C-6	99.74	98.51	97.89	98.20	96.46
C-7	99.63	98.80	98.45	98.62	97.28
C-8	99.76	96.55	98.63	97.58	95.28
C-9	99.83	85.09	92.97	88.85	79.94
C-10	99.83	86.78	93.06	89.81	81.50
Average	99.72	87.92	93.09	89.98	84.60

Fig. 4 exhibits a detailed average result obtained by the MFSOSAE-ID model under four runs on UNSW-NB15 dataset. The figure notified that the MFSOSAE-ID model has resulted in maximum classifier outcomes under all runs. For instance, with run-1, the MFSOSAE-ID model has gained average $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 98.88%, 70.58%, 62.08%, 63.68%, and 55.78% correspondingly. Besides, with run-3, the MFSOSAE-ID methodology has reached average $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.72%, 87.17%, 90.80%, 88.20%, and 82.47% correspondingly. Meanwhile, with run-4, the MFSOSAE-ID system has attained average $accu_y$, $prec_n$, $reca_l$, F_{score}, and J_{index} of 99.72%, 87.92%, 93.09%, 89.98%, and 84.60% correspondingly.

The training accuracy (TA) and validation accuracy (VA) attained by the MFSOSAE-ID system on test dataset is demonstrated in Fig. 5. The experimental outcome implied that the MFSOSAE-ID method has gained maximum values of TA and VA. In specific, the VA seemed to be superior to TA.

The training loss (TL) and validation loss (VL) achieved by the MFSOSAE-ID approach on test dataset are established in Fig. 6. The experimental outcome inferred that the MFSOSAE-ID methodology has been able least values of TL and VL. In specific, the VL seemed to be lower than TL.

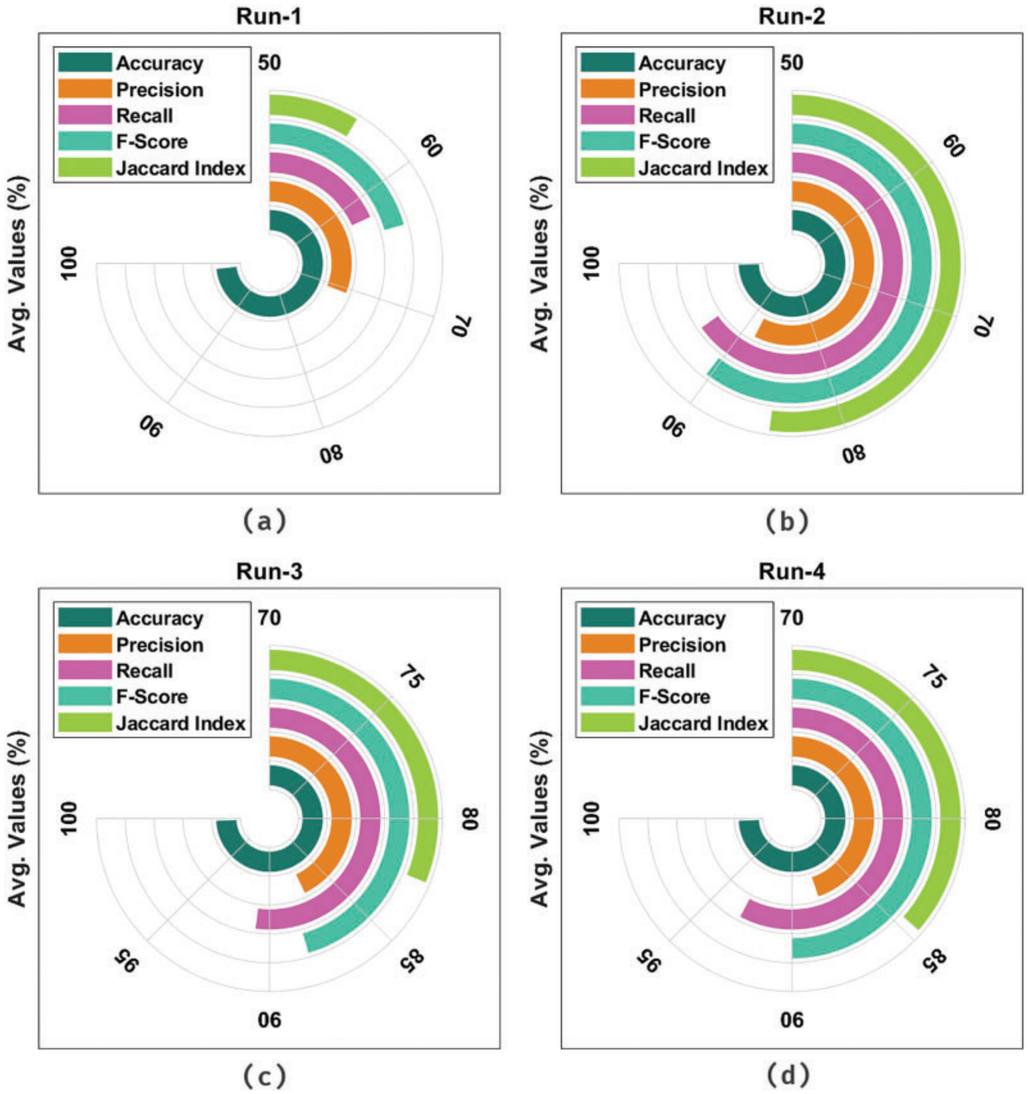


Figure 4: Average analysis of MFSOSAE-ID technique (a) run-1, (b) run-2, (c) run-3, and (d) run-4

At last, a comprehensive comparative examination of the MFSOSAE-ID technique with recent methodologies is offered in Tab. 6 [26]. Fig. 7 illustrates a comparison study of the MFSOSAE-ID model with other models interms of $accu_y$. The figure portrayed that the logistic regression (LR) model has shown ineffectual outcome with least $accu_y$ value of 89.62%. Followed by, the k-nearest neighbor (KNN) and kernel extreme learning machine (KELM) models have shown slightly enhanced $accu_y$ values of 91.84% and 92.31% correspondingly. Besides, the extreme learning machine (ELM) and multilayer perceptron (MLP) models have obtained moderately closer $accu_y$ values of 94.17% and 93.52% respectively. Though the long short term memory with recurrent neural network (LSTM-RNN) model has resulted in reasonable $accu_y$ of 97.51%, the MFSOSAE-ID model has shown maximum $accu_y$ of 99.72%.

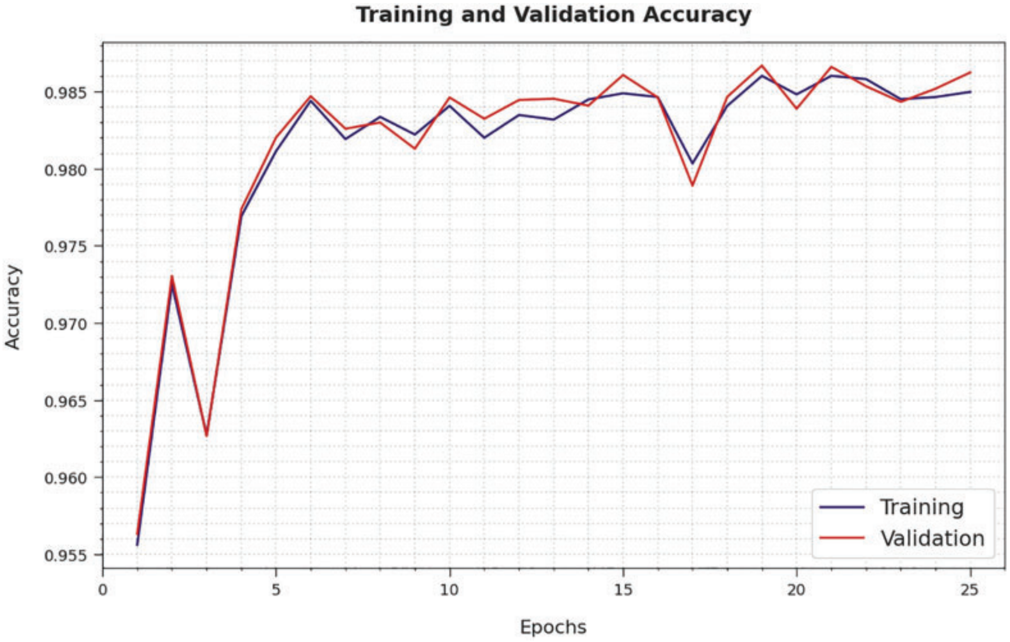


Figure 5: TA and VA analysis of MFSOSAE-ID technique

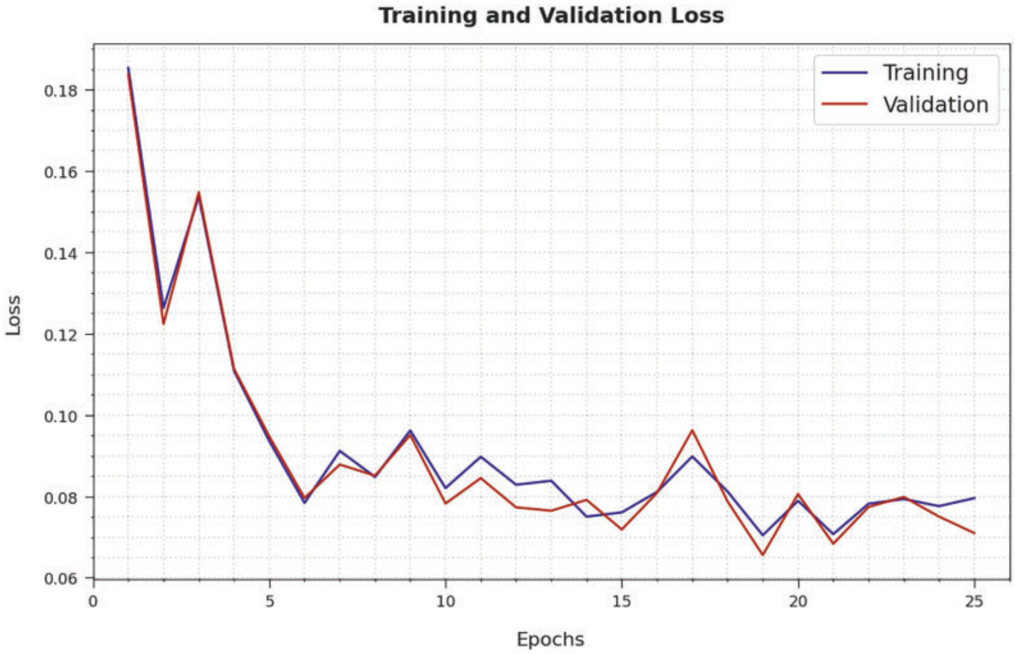


Figure 6: TL and VL analysis of MFSOSAE-ID technique

Table 6: Comparative analysis of MFSOSAE-ID algorithm with existing approaches

Methods	Accuracy	Precision	Recall	F-score
LSTM-RNN	97.51	84.56	91.12	86.58
Logistic regression	89.62	87.85	89.12	89.26
K Nearest neighbor	91.84	81.97	90.56	85.87
ELM	94.17	81.57	87.61	89.54
KELM	92.31	86.37	92.11	90.05
MLP	93.52	84.57	91.54	87.11
MFSOSAE-ID	99.72	88.32	93.44	90.31

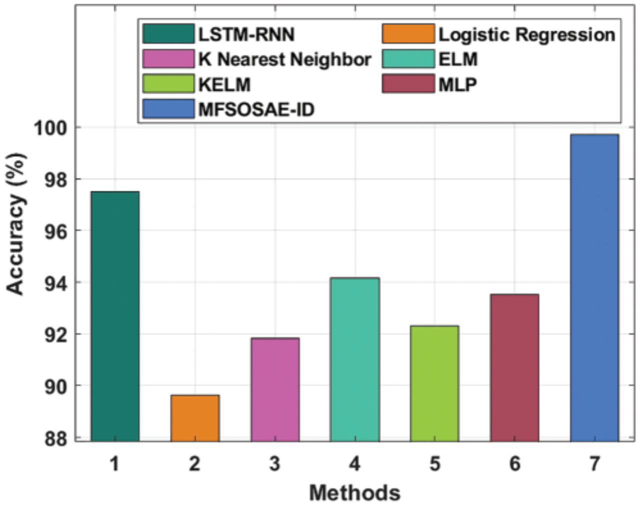


Figure 7: $Accu_y$ analysis of MFSOSAE-ID technique with existing approaches

Fig. 8 showcases a comparison study of the MFSOSAE-ID model with other models with respect to $prec_n$. The figure exposed that the LR method has presented ineffectual outcome with least $prec_n$ value of 87.85%. In addition, the KNN and KELM methodologies have shown somewhat higher $prec_n$ values of 81.97% and 86.37% respectively. Moreover, the ELM and MLP techniques have obtained moderately closer $prec_n$ values of 81.57% and 84.57% correspondingly. However, the LSTM-RNN approach has resulted in reasonable $prec_n$ of 84.56%, the MFSOSAE-ID approach has shown maximal $prec_n$ of 88.32%. These results and discussion reported that the MFSOSAE-ID model has reached effectual outcomes over other models.

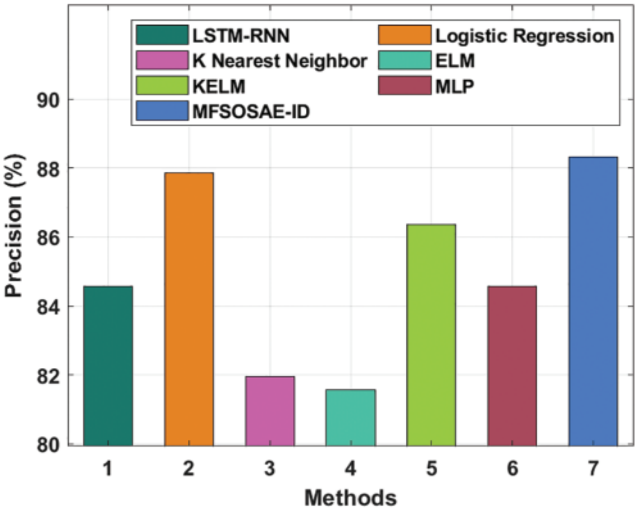


Figure 8: *Prec$_n$* analysis of MFSOSAE-ID algorithm with existing approaches

4 Conclusion

In this study, a novel MFSOSAE-ID approach was established to determine the existence of intrusions in the IoD environment. The presented MFSOSAE-ID technique firstly pre-processes the input data into a compatible format. Followed this, the presented MFSOSAE-ID technique designs an MFOFS technique to elect appropriate features. In line with, the SAE model is employed for the recognition and classification of intrusions in which the SAE parameters are optimally tuned with utilize of FFA. The performance validation of the MFSOSAE-ID model is tested utilizing benchmark dataset and the outcomes implied the promising performance of the MFSOSAE-ID model over other techniques. Thus, the presented MFSOSAE-ID technique can be employed to accomplish security in the IoD environment. In future, advanced deep learning (DL) based classification models can be employed to boost detection efficiency of the MFSOSAE-ID model.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (158/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R140), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: 22UQU4210118DSR18.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Gharibi, R. Boutaba and S. L. Waslander, "Internet of drones," *IEEE Access*, vol. 4, pp. 1148–1162, 2016.

[2] Z. Lv, "The security of internet of drones," *Computer Communications*, vol. 148, no. 2, pp. 208–214, 2019.

[3] M. Yahuza, K. Dimyati, K. A. Noordin, N. S. M. Shah, S. H. Alsamhi *et al.,* "Internet of drones security and privacy issues: Taxonomy and open challenges," *IEEE Access*, vol. 9, pp. 57243–57270, 2021.

[4] T. Long, M. Ozger, O. Cetinkaya and O. B. Akan, “Energy neutral internet of drones,” *IEEE Communications Magazine*, vol. 56, no. 1, pp. 22–28, 2018.

[5] A. Koubâa, B. Qureshi, M. F. Sriti, A. Allouch, Y. Javed *et al.,* “Dronemap Planner: A service-oriented cloud-based management system for the Internet-of-Drones,” *Ad Hoc Networks*, vol. 86, no. 2, pp. 46–62, 2019.

[6] A. Abdelmaboud, “The internet of drones: Requirements, taxonomy, recent advances, and challenges of research trends,” *Sensors*, vol. 21, no. 17, pp. 5718, 2021.

[7] J. V. Anchitaalagammai, T. Jayasankar, P. Selvaraj, M. Y. Sikkandar, M. Zakarya *et al.,* “Energy efficient cluster-based optimal resource management in iot environment,” *Computers, Materials & Continua*, vol. 70, no. 1, pp. 1247–1261, 2022.

[8] M. Wazid, A. K. Das and J. H. Lee, “Authentication protocols for the internet of drones: Taxonomy, analysis and future directions,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 101, no. 3, pp. 42, 2018.

[9] M. A. Ferrag, L. Shu, O. Friha and X. Yang, “Cyber security intrusion detection for agriculture 4.0: Machine learning-based solutions, datasets, and future directions,” *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 407–436, 2022.

[10] I. V. Pustokhina, D. A. Pustokhin, E. L. Lydia, P. Garg, A. Kadian *et al.,* “Hyperparameter search based convolution neural network with Bi-LSTM model for intrusion detection system in multimedia big data environment,” *Multimedia Tools and Applications*, vol. 13, no. 5, pp. 111, 2021.

[11] S. Ouiazzane, M. Addou and F. Barramou, “A multiagent and machine learning based denial of service intrusion detection system for drone networks,” in *Geospatial Intelligence, Advances in Science, Technology & Innovation Book Series (ASTI)*. Cham: Springer, pp. 51–65, 2022.

[12] J. P. Condomines, R. Zhang and N. Larrieu, “Network intrusion detection system for UAV ad-hoc communication: From methodology design to real test validation,” *Ad Hoc Networks*, vol. 90, no. 2, pp. 101759, 2019.

[13] S. Perumalla, S. Chatterjee and A. P. S. Kumar, “Block chain-based access control and intrusion detection system in IoD,” in *2021 6th Int. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatre, India, pp. 511–518, 2021.

[14] D. Sharma, S. K. Gupta, A. Rashid, S. Gupta, M. Rashid *et al.,* “A novel approach for securing data against intrusion attacks in unmanned aerial vehicles integrated heterogeneous network using functional encryption technique,” *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 7, pp. e3667, 2021.

[15] S. Ouiazzane, F. Barramou and M. Addou, “Towards a multi-agent based network intrusion detection system for a fleet of drones,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 10, 2020. DOI 10.14569/issn.2156-5570.

[16] M. A. Hamza, S. B. Haj Hassine, I. Abunadi, F. N. Al-Wesabi, H. Alsolai *et al.,* “Feature selection with optimal stacked sparse autoencoder for data mining,” *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2581–2596, 2022.

[17] A. A. Albraikan, S. B. Haj Hassine, S. M. Fati, F. N. Al-Wesabi, A. M. Hilal *et al.,* “Optimal deep learning-based cyberattack detection and classification technique on social networks,” *Computers, Materials & Continua*, vol. 72, no. 1, pp. 907–923, 2022.

[18] A. M. Hilal, J. S. Alzahrani, I. Abunadi, N. Nemri, F. N. Al-Wesabi *et al.,* “Intelligent deep learning model for privacy preserving IIoT on 6g environment,” *Computers, Materials & Continua*, vol. 72, no. 1, pp. 333–348, 2022.

[19] S. Mirjalili, “Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm,” *Knowledge-Based Systems*, vol. 89, pp. 228–249, 2015.

[20] P. Zhou, J. Han, G. Cheng and B. Zhang, “Learning compact and discriminative stacked autoencoder for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4823–4833, 2019.

[21] I. V. Pustokhina, D. A. Pustokhin, E. L. Lydia, P. Garg, A. Kadian *et al.,* "Hyperparameter search based convolution neural network with Bi-LSTM model for intrusion detection system in multimedia big data environment," *Multimedia Tools and Applications*, vol. 13, no. 5, pp. 111, 2021.

[22] K. Shankar, E. Perumal, P. Tiwari, M. Shorfuzzaman and D. Gupta, "Deep learning and evolutionary intelligence with fusion-based feature extraction for detection of COVID-19 from chest X-ray images," *Multimedia Systems*, vol. 66, no. 2, pp. 1921, 2021.

[23] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C. H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020.

[24] X. Yang and X. He, "Firefly algorithm: recent advances and applications," *International Journal of Swarm Intelligence*, vol. 1, no. 1, pp. 36, 2013.

[25] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Military Communications and Information Systems Conf. (MilCIS)*, Poland, pp. 1–6, 2015.

[26] R. A. Ramadan, A. H. Emara, M. A. Sarem and M. Elhamahmy, "Internet of drones intrusion detection using deep learning," *Electronics*, vol. 10, no. 21, pp. 2633, 2021.