applied sciences

MDPI

Article

Evolutionary-Based Deep Stacked Autoencoder for Intrusion Detection in a Cloud-Based Cyber-Physical System

Mesfer Al Duhayyim [1,*], Khalid A. Alissa [2], Fatma S. Alrayes [3], Saud S. Alotaibi [4], ElSayed M. Tag El Din [5], Amgad Atta Abdelmageed [6], Ishfaq Yaseen [6] and Abdelwahed Motwakel [6]

1 Department of Computer Science, College of Sciences and Humanities-Aflaj, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
2 SAUDI ARAMCO Cybersecurity Chair, Networks and Communications Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia; KaAlissa@iau.edu.sa
3 Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; Fsalrayes@pnu.edu.sa
4 Department of Information Systems, College of Computing and Information System, Umm Al-Qura University, Mecca 24382, Saudi Arabia; sotaibi@uqu.edu.sa
5 Department of Electrical Engineering, Faculty of Engineering and Technology, Future University in Egypt, New Cairo 11845, Egypt; ElSayed.TagElDin@fue.edu.eg
6 Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia; abzwesabi@gmail.com (A.A.A.); tsr2wesabi@gmail.com (I.Y.); mrbwesabi@gmail.com (A.M.)
* Correspondence: m.alduhayyim@psau.edu.sa

check for updates

Citation: Duhayyim, M.A.; Alissa, K.A.; Alrayes, F.S.; Alotaibi, S.S.; Tag El Din, E.M.; Abdelmageed, A.A.; Yaseen, I.; Motwakel, A. Evolutionary-Based Deep Stacked Autoencoder for Intrusion Detection in a Cloud-Based Cyber-Physical System. *Appl. Sci.* **2022**, *12*, 6875. https://doi.org/10.3390/app12146875

Academic Editor: João M.F. Rodrigues

Received: 5 June 2022
Accepted: 1 July 2022
Published: 7 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: As cyberattacks develop in volume and complexity, machine learning (ML) was extremely implemented for managing several cybersecurity attacks and malicious performance. The cyber-physical systems (CPSs) combined the calculation with physical procedures. An embedded computer and network monitor and control the physical procedure, commonly with feedback loops whereas physical procedures affect calculations and conversely, at the same time, ML approaches were vulnerable to data pollution attacks. Improving network security and attaining robustness of ML determined network schemes were the critical problems of the growth of CPS. This study develops a new Stochastic Fractal Search Algorithm with Deep Learning Driven Intrusion Detection system (SFSA-DLIDS) for a cloud-based CPS environment. The presented SFSA-DLIDS technique majorly focuses on the recognition and classification of intrusions for accomplishing security from the CPS environment. The presented SFSA-DLIDS approach primarily performs a min-max data normalization approach to convert the input data to a compatible format. In order to reduce a curse of dimensionality, the SFSA technique is applied to select a subset of features. Furthermore, chicken swarm optimization (CSO) with deep stacked auto encoder (DSAE) technique was utilized for the identification and classification of intrusions. The design of a CSO algorithm majorly focuses on the parameter optimization of the DSAE model and thereby enhances the classifier results. The experimental validation of the SFSA-DLIDS model is tested using a series of experiments. The experimental results depict the promising performance of the SFSA-DLIDS model over the recent models.

Keywords: Internet of Things; deep learning; cyber physical systems; cloud computing; intrusion detection; security

1. Introduction

With the emergence of disruptive technology, Industry 4.0 is experiencing huge transitions in terms of cost efficiency and performance [1]. In particular, this applies to smart computing on a big scale, namely, Cloud Computing, the Internet of Things (IoTs), and Cyber Physical System (CPS). CPS is a multi-dimensional, complex system that integrates a computer, network, and physical environment [2]. With the deep collaboration of 3C

(control, computation, and communication) techniques, the dynamic control, information servicing, and real-time perception of large engineering systems are realized [3]. CPS realizes the organic design of physical, computation, and transmission systems, making the system capable, reliable, and effective for simultaneous collaboration, resulting in extensive and important application prospects. CPS is utilized in different industries and fields [4,5]. In recent times, the information technology sector has expanded rapidly. Innovations and breakthroughs of several techniques have been established, resulting in earth-shaking changes in people's lives [6].

Particularly, in the process of rapid development, embedded technologies are often applied to human life [7,8]. CPS has become the most prominent in researches and development direction for scholars in different countries because of its extensible, scalable, and interactive features, and also it becomes a priority investment region for large enterprises. In contrast to embedded technologies, a CPS, as a combination of computer technology and physical equipment, transforms a computing object from distributed to unified, discrete to continuous, and digital to analog [9]. In contrast to the IoT system, the perceptibility of CPS after the connection of physical entities pay increased attention to dynamic or ongoing data control of the information services and major component of the device. In comparison to software system, CPS focuses on the feedback and control of the physical process, highlighting the dynamic response and interaction of data processing [10].

In relation to CPS security, a conventional pattern approaches the cyber and physical systems individually and cannot address vulnerability that is related to embedded controllers and networks that are intended for controlling and monitoring physical processes [11]. Hence, it is necessary for an organic security system to protect CPS from cyberattacks. In this study, there exists strong evidence of the necessity for security in this system and the havoc that can result if the security is disregarded. To identify unexpected errors and attacks in CPS, an anomaly detection method is suggested to mitigate the threat. For instance, state estimation (i.e., Kalman filter), rule, statistical models (histogram-based model and Gaussian model) based methods are applied for learning the regular status of CPS [12]. However, each method generally needs expert knowledge (for example, operator manually extracts some rules), or should know the fundamental distribution of data. The machine learning (ML) approach does not depend upon domain-specific knowledge [13]. However, it generally needs an abundance of labeled datasets (for example, classification-based method). As well, they could capture the unique attribute of CPS (for example, spatial-temporal relationship). The intrusion detection (ID) method is dedicated to ensuring network security.

This study develops a new Stochastic Fractal Search Algorithm with Deep Learning Driven Intrusion Detection system (SFSA-DLIDS) for a cloud-based CPS environment. The presented SFSA-DLIDS technique majorly focuses on the recognition and classification of intrusions for accomplishing security from the CPS environment. The presented SFSA-DLIDS approach primarily performs min-max data normalization approach to convert the input data to a compatible format. In order to reduce a curse of dimensionality, the SFSA technique is applied to select a subset of features. The SFSA uses the idea of fractals to satisfy the intensification (exploitation) property needed by optimization algorithms, and the stochasticity feature to guarantee the diversification (exploration) of the search space. Additionally, chicken swarm optimization (CSO) with deep stacked auto encoder (DSAE) technique was utilized for the identification and classification of intrusions. The experimental validation of the SFSA-DLIDS model is tested using a series of experiments.

2. Literature Review

Li et al. [14] present a novel federated DL approach termed DeepFed, for identifying cyber threats against industrial CPSs. Especially, the authors' primary design is a novel DL-based ID method for industrial CPSs, creating utilization of CNN and GRU. Secondary, the authors create a federated learning structure and permit several industrial CPSs to combine a detailed ID method from a privacy-preserving approach. de Araujo-Filho et al. [15]

present FID-GAN, a novel fog-based, unsupervised IDS for CPSs employing a generative adversarial network (GAN). The IDS was presented to a fog structure that takes computation resources as near as possible to the end node and so provides for a lesser latency requirement. For achieving superior detection rates, the presented structure estimates a reconstruction loss depending upon the reform of data instances mapped to latent spaces. Alohali et al. [16] project a novel AI-enabled multimodal fusion-based IDS (AIMMF-IDS) for CCPS from the Industry 4.0 environments. The presented method primarily executes the data pre-processed approach in two manners such as data conversion and data normalization. Moreover, an improved fish swarm optimization-based FS (IFSO-FS) system is utilized to suitable selective features. The IFSO approach was developed by utilizing a Levy Flight (LF) model as the search process of a typical FSO technique in order to avoid the local optimum problems.

Althobaiti et al. [17] examine a novel cognitive computing-based IDS approach for achieving security from industrial CPS. The presented method contains pre-processing for discarding the noise which exists from the data. Afterward, the proposed method utilizes a binary bacterial foraging optimization (BBFO) based FS approach for selecting the best subset of features. Additionally, the GRU method was executed for identifying the occurrence of intrusions from the industrial CPS environments. The authors in [18] primarily present a new self-learning spatial distribution technique called Euclidean distance-based between-class learning (EBC learning) that enhances between-class learning by computing the Euclidean distance (ED) amongst KNN of distinct classes. Moreover, a cognitive computing-based ID model termed order-line SMOTE and EBC learning dependent upon RF (BSBC-RF) is also presented as dependent upon EBC learning to industrial CPSs. Ibor et al. [19] present a new hybrid technique for intrusion forecast on a CPS's communication network. The authors utilize a bio-simulated hyperparameter searching approach for generating an enhanced DNN infrastructure dependent upon the basic hyperparameters of NNs. In addition, the authors develop a forecasting method dependent upon the enhanced NN infrastructure. Some other methods in the literature are available in [20–23].

3. The Proposed Model

In this article, a new SFSA-DLIDS technique has been projected for the classification and identification of intrusions from the CPS environment. The presented the SFSA-DLIDS model primarily performs a min-max data normalization approach to convert the input data to a compatible format, followed by the SFSA technique, which is applied to select a subset of features. Finally, the CSO-DSAE approach was utilized for the identification and classification of intrusions. Figure 1 depicts the block diagram of the SFSA-DLIDS approach.

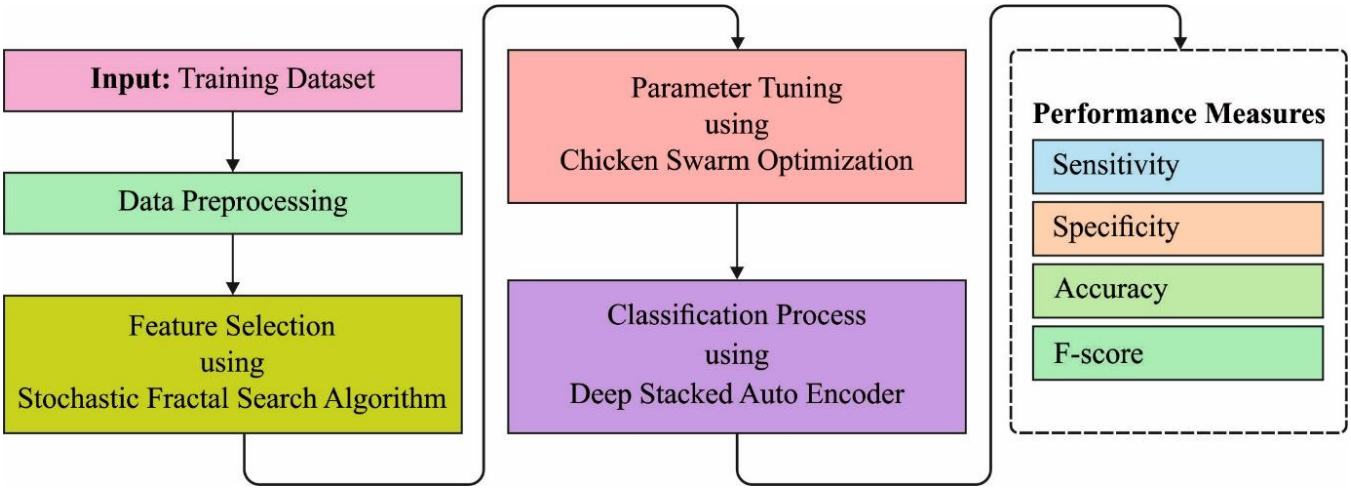

Figure 1. Block diagram of SFSA-DLIDS approach.

3.1. Data Pre-Processing

At the initial stage, the presented approach performs the min-max data normalization approach to convert the input data to a compatible format. It can be executed to scale the feature from the zero and one range with the execution in Equation (1):

$$v' = \frac{v - \min_A}{\max_A - \min_A} \tag{1}$$

At this point, $\min_A$ and $\max_A$ signifies the minimal and maximal values of features A. The original and normalized values of the elements, A have been demonstrated by v and v' correspondingly. It is apparent in the above formula that the maximal and minimal feature values were mapped to one and zero correspondingly.

3.2. Feature Selection Using SFSA Technique

In this work, the SFSA technique is applied to select a subset of features. SFSA is based on the specific development marvel of a random fractal and basically uses two processes afterward for population initialization: (1) diffusion and (2) update to enhance the searching [24]. In the arithmetical modeling of the SFSA, the finest solution is only preferred from the diffusion method to generate novel arrangements, while overlooking discrete arrangements. The procedure for making a new arrangement is characterized as Gaussian walks (Gw) that are determined in the following:

$$Gw_1 = Gaussian\ (\mu_G,\ \sigma) + (rand(0,1) \times P_B - rand\ (0,1) \times P_i) \tag{2}$$

$$Gw_2 = Gaussian(\mu_P,\ \delta) \tag{3}$$

The expression: $rand(0,\ 1)$ refers to an arbitrary value that lies between zero and one, P_i and P_B denotes the *i-th* solutions and every particle diffuses around its position and completes correspondingly; μ_G and μ_P indicates the Gaussian means walk that is equivalent to $|P_i|$ and $|P_B|$ correspondingly; δ denotes the standard deviation that is calculated by:

$$\delta \left| \frac{\log(g)}{g} (P_i - P_B) \right| \tag{4}$$

In Equation (4), g represents the iteration count. In the optimization technique, g is increased but lesser than ending criteria G_{max}, δ are attuned dynamically. All the particles are diffused around their current situation by using the Gaussian walk until a predetermined extreme dissemination number YD is obtained. Based on the following equation, many generated solutions are attained:

$$P_{ij} = LB_{ij} + rand(0,1) \times \left(UB_{ij} - LB_{ij}\right), j = 1,2,3,\ldots,Y_D \tag{5}$$

In Equation (5), UB_{ij} and LB_{ij} refers to the upper as well as lower limits of *j*-th values of solution i; y_D denotes the maximal diffusion count of solutions generated by the SFSA. Then, the quality of solution has been calculated and the optimal solution P_B is defined. In this step, two methods are used: initially update the solution of (p) probability according to the value of $Pa < rand\ (0,1)$ as follows:

$$P_a = \frac{rank(P_i)}{N_D} \tag{6}$$

$$P'_{ij} = P_{aj} - rand(0,1) \times (P_{a1} - P_{a2}) \tag{7}$$

In Equation (7), rank (p) refers to the rank of i^{th} solutions amongst different arrangements in the population; P'_i indicates the new solution of the *i-th* solution; P_{a1} and P_{a2} signifies the random solution of population. Next, improve the exploration accordingly and apply the variations to the solution based on discrete solutions from the population.

However, the process initiates by sorting each arrangement based on the following equation. When $P_a < rand(0,1)$ for the P_i', there is no update for the present solution P_i or else the solution was upgraded by the following expression:

$$P_i'' = \begin{cases} P_i' \bar{} rand(0,1) \times (P_t' - P_B) & rand(0,1) \leq 0.5 \\ P_i' + rand(0,1) \times (P_t' - P_r') & rand(0,1) > 0.5 \end{cases} \quad (8)$$

Let P_t' and P_t' be the solution randomly designated from the Gaussian distribution. The following phase is for comparing the quality of P_i'' with P_i' and $\left(P_i''\right)$ is superior to $\left(P_i'\right)$ then P_i'' is substituted P_i or else P_i' not upgraded.

The fitness function (FF) of the SFSA system utilized from the presented system was planned to contain a balance among the amount of chosen features from all the solutions (minimal) and the classifier accuracy (maximal) reached by utilizing these selective features. Equation (9) defines the FF for evaluating solutions:

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|C|} \quad (9)$$

whereas $\gamma_R(D)$ denotes the classifier error rate of provided classifier (the KNN technique was utilized). $|R|$ stands for the cardinality of chosen subset and $|C|$ signifies the entire amount of features from the dataset, α, and β signifies the 2 parameters equivalent to significance of classifier quality and subset length. $\in [1, 0]$ and $\beta = 1 - \alpha$.

3.3. DSAE-Based Data Classification

To recognize and classify intrusions, the DSAE model has been exploited in this study. In our study, the SAE used is developed by different LR and AE layers [25]. AE is the fundamental component of SAE classification. Figure 2 demonstrates the infrastructure of SAE. It comprises an encoding step (Layer 1 to Layer 2) and a reconstruction or decoding step (Layer 2 to Layer 3). This procedure is expressed in the following equation, where W and W^T (the transpose of W) refers to the weight matrix of mode b and b' are two distinct bias vectors, s indicates a nonlinearity function, namely the sigmoid function applied in the work, γ denotes a latent depiction of x input layer, and z is regarded as a prediction of x given γ and it must have the identical shape as x.

$$\gamma = s(Wx + b). \quad (10)$$

$$z = s\left(W^T\gamma + b'\right) \quad (11)$$

Numerous AE layers are collectively stacked to procedure in an unsupervised pre-training phase (Layer 1 to Layer 4). The latent depiction $\prime y$ calculated by an AE is utilized as the input to the following AE layers. All the layers are trained by an AE by reducing the reconstruction error that perform as a single layer at a time. The reconstructing error (loss function (x, z)) is evaluated in different methods. In our work, we apply cross-entropy for measuring the reconstructing error, as demonstrated in Equation (12), where x_k and z_k denotes the k^{th} component of x and z, correspondingly.

$$L(x, z) = -\sum_{k=1}^{d}[x_k lnz_k + (1 - x_k)ln(1 - z_k)] \quad (12)$$

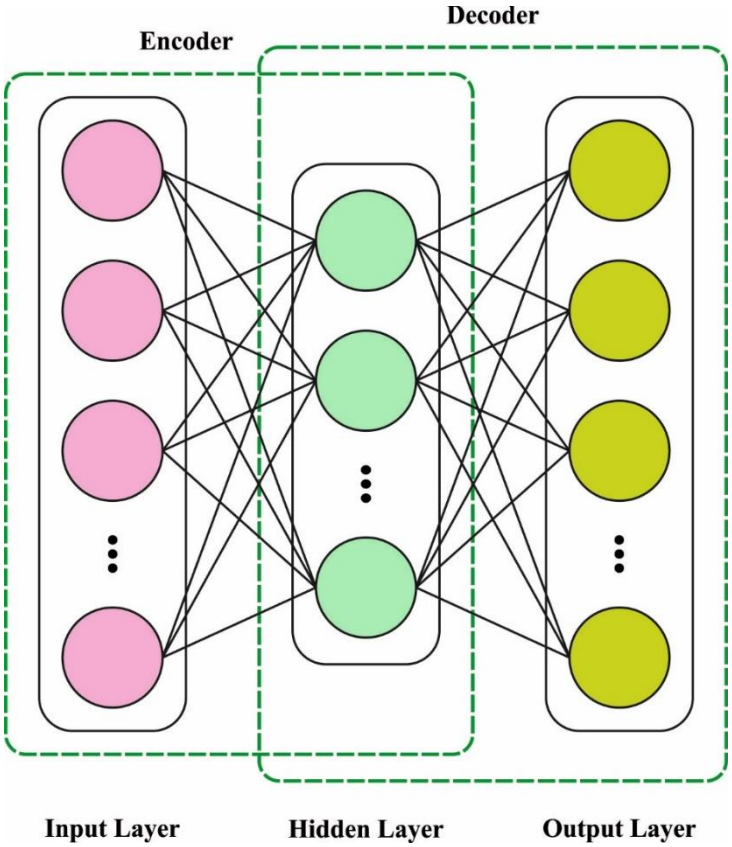

Figure 2. Structure of SAE.

The reconstructing error is minimalized by the Gradient Descent mechanism. The weight in Equations (10) and (11) need to be upgraded based on the following equations, where 0 indicates the learning rate:

$$W = W - \alpha \frac{\partial L(x,z)}{\partial W}. \tag{13}$$

$$b = b - \alpha \frac{\partial L(x,z)}{\partial b}. \tag{14}$$

$$b' = b' - \alpha \frac{\partial L(x,z)}{\partial b'}. \tag{15}$$

Once the layer is pre-trained, the network enters the supervised finetuning phase. In the supervised finetuning phase, add an LR layer to the resultant layer. In this work, probability that the x input vector (Layer 4) belonging to *i-th* class is determined in the above equation, where y represents the predicted class of input vector x. W and b denote the weight matrices and the bias vector, correspondingly, W_j and W_j denote the i^{rh} and j^{rh} row of matrixes W, correspondingly, b_i and b_j denote the i^{th} and j^{th} elements of vector b, correspondingly, and $softmax$ refers to the nonlinearity function. The class with the maximum probability was assumed as the prediction label y_{pred} of x input vector, as determined in Equation (17). The predictive error of sample dataset $D(Loss(D))$ is evaluated according to the true label, as demonstrated in Equation (18), where y_i indicates the true label of x_i. $Loss(D)$ is minimalized by the Gradient Descent model that is the same as the procedure of minimalizing the abovementioned reconstruction error:

$$P(Y = i|x,\ W,\ b) = softmax\ (Wx + b) = \frac{e^{W_i x + b_j}}{\sum_j e^{W_j x + b_j}}. \tag{16}$$

$$y_{pred} = \text{argmax}\ (P(Y = i|x,\ W,\ b)) \tag{17}$$

$$Loss\ (D) = -\sum_{i=0}^{D} \text{In}\ (P(V = y_i|x_i,\ W,\ b)) \tag{18}$$

3.4. Hyperparameter Tuning Using CSO Algorithm

Here, the CSO algorithm was executed for the parameter optimization of the DSAE approach and thereby enhances the classifier results. Meng et al. [26] suggest the CSO technique. A novel SI optimized technique was presented for simulating the hierarchy and foraging performance of chickens. The population was separated into many subgroups. All the subgroups contain cock, chick, and hen. The CSO technique follows the subsequent principles:

(1) The whole population contains many sub-populations, each of which comprises cock, amount of hens, and many chicks.
(2) The fitness value (FV) of all the particles from the population was computed. The particle is classified depending upon the FV. Some particles with optimum FVs were chosen as cocks, some particles with worse FVs were chosen as chickens, and remaining particles were chosen as hens.
(3) In specific hierarchy, the dominance connection and mother–child connection remained unaffected. However, as the chicks produced, the population connection was modified. The hierarchy control connection and maternal connection of chicken swarms were variations all the G time.
(4) The cock controls the flock, the hen follows the cock from its individual populations and the chick food was nearby the hens. The hen arbitrarily combines a subpopulation. The connection among mother as well as child from the flock was arbitrarily introduced. The cock with main searching range and an optimum searching capability was led from the flocks. The chick particle has the worse foraging capability and minimum foraging range. The foraging capability and searching range of hen particles were amongst cock as well as chick particles.

In CSO, there were N particles from the entire chick flock. The amount of roosters can be determined as N_r. The amount of hens was determined as N_h, and the amount of chickens was N_c. Distinct types of chickens are distinct place upgrade formulas if they can be determined as food [20]. The roosters are one of the adjustable individuals from chickens and, most apparently, for defining food from the entire population.

The formula for place upgrade of cock particles is depicted in Equation (19):

$$\begin{aligned} P_i^j(t+1) &= P_i^j(t) * \left(1 + Randn\left(0, \sigma^2\right)\right) \\ \sigma^2 &= \begin{cases} 1 & W_i < W_k \\ \exp\left(\frac{(W_k - W_i)}{|W_i| + \varepsilon}\right) & others \end{cases} \end{aligned} \tag{19}$$

In which, the $k \in [1, c_n]$, and $k \neq i$. $Randn\left(0, \sigma^2\right)$ means the Gaussian distribution with mean value of 0 and standard deviation of σ^2. An individual place of $P_i^j(t)$ is the value of j^{th} dimensional of i^{th} individual at t^{th} iterations. ε is some lesser constant; k refers to the random cock from every cock except i^{th} cock; W_i refers the FV equivalent to i^{th} cock; W_k denotes the FV equivalent to k^{th} cocks. The hens are maximal proportion of individuals from the entire chick population. Their place upgrade formulation is depicted in Equation (20):

$$\begin{aligned} P_i^j(t+1) = P_i^j(t) + K_1 * Random\ * \left(P_{r1}^i(t) - P_i^j(t)\right) + K_2 * Random\ * \left(P_{r2}^j(t) - P_i^j(t)\right) \\ K_1 = \frac{\exp\left(W_i - W_{r_1}\right)}{|W_i| + \varepsilon} \\ K_2 = \exp\ (W_{r_2} - W_i) \end{aligned} \tag{20}$$

whereas *Random* stands for the arbitrary number amongst zero and one, which follows the standard normal distribution. r_1 represents the cock from the group but i^{th} hen was placed. r_2 demonstrated that some cock excepting the cock from the set of i^{th} hen. Therefore r_1 is distinct in r_2. The chick follows the hen searching and chick place upgrade equation demonstrated in Equation (21):

$$P_i^j(t+1) = P_i^j(t) + FL * \left[P_m^j(t) - P_i^j(t)\right] \tag{21}$$

In which FL refers to the average amount equally distributed from zero and two. $P_m^j(t)$ implies the hen place equivalent to i^{th} chick. The CSO system determines an FF to accomplish superior classifier performances. It determines a positive integer for exemplifying the best efficiency of candidate results. In this scenario, the reduced classification error rate has been supposed that FF is providing in Equation (22). An optimum outcome is a decreased error rate and a worse solution accomplishes an improved error rate:

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \end{aligned} \tag{22}$$

4. Results and Discussion

The experimental validation of the SFSA-DLIDS model is tested using two benchmark datasets, namely, NSLKDD 2015 [27] and CICIDS 2017 [28] datasets. Table 1 illustrates the details on two benchmark datasets. The NSLKDD 2015 dataset holds samples under two classes. It includes 67,343 samples under normal class and 58,630 samples under anomaly class. In addition, the CICIDS 2017 dataset holds 50,000 samples under normal class and 50,000 samples under anomaly class.

Table 1. Dataset details.

Class	No. of Samples	
	NSLKDD 2015	CICIDS 2017
Normal	67,343	50,000
Anomaly	58,630	50,000
Total	125,973	100,000

Figure 3 indicates the confusion matrices produced by the SFSA-DLIDS approach on the test NSLKDD 2015 dataset. With 70% of training (TR) dataset, the SFSA-DLIDS model has recognized 46,762 samples into normal class and 40,054 samples into anomaly class. In addition, with 30% of the testing (TS) dataset, the SFSA-DLIDS method has recognized 20,147 samples into normal class and 17,062 samples into anomaly class. Additionally, with 20% of TS dataset, the SFSA-DLIDS approach has identified 13,390 samples into normal class and 11,665 samples into anomaly class.

Table 2 and Figure 4 showcase the overall classification output of the SFSA-DLIDS model on the test NSLKDD 2015 dataset. The results implied that the SFSA-DLIDS model has resulted to enhanced results under all aspects. For instance, with 70% of TR data, the SFSA-DLIDS model has offered average $accu_y$ of 97.74%, $prec_n$ of 97.76%, $reca_l$ of 97.74%, and F_{score} of 97.74%. Simultaneously, with 30% of TS data, the SFSA-DLIDS approach has rendered average $accu_y$ of 97.86%, $prec_n$ of 97.88%, $reca_l$ of 97.87%, and F_{score} of 97.86%. Concurrently, with 20% of TS data, the SFSA-DLIDS method has provided average $accu_y$ of 99.32%, $prec_n$ of 99.32%, $reca_l$ of 99.33%, and F_{score} of 99.32%.

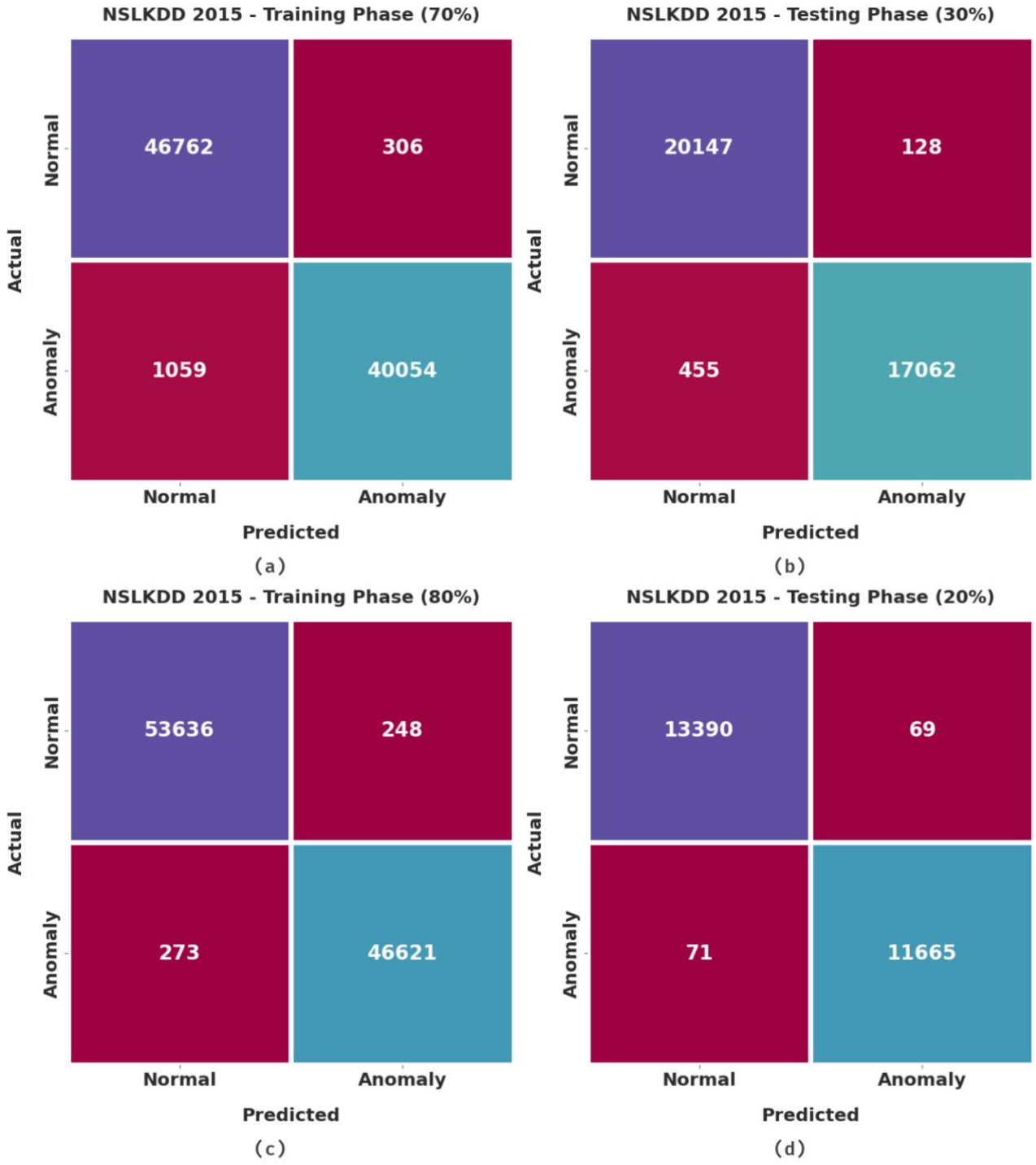

Figure 3. Confusion matrices of SFSA-DLIDS approach under NSLKDD 2015 dataset: (**a**) 70% of TR data, (**b**) 30% of TS data, (**c**) 80% of TR data, and (**d**) 20% of TS data.

The training accuracy (TA) and validation accuracy (VA) acquired by the SFSA-DLIDS approach on NSLKDD 2015 dataset is demonstrated in Figure 5. The experimental outcome denoted that the SFSA-DLIDS algorithm attained maximal values of TA and VA. In specific, the VA is higher than TA.

Table 2. Result analysis of SFSA-DLIDS approach with various measures on NSLKDD 2015 dataset.

Class Labels	Accuracy	Precision	Recall	F-Score
Training Phase (70%)				
Normal	97.74	98.76	96.67	97.71
Anomaly	97.74	96.76	98.80	97.77
Average	**97.74**	**97.76**	**97.74**	**97.74**
Testing Phase (30%)				
Normal	97.86	98.94	96.79	97.85
Anomaly	97.86	96.82	98.95	97.87
Average	**97.86**	**97.88**	**97.87**	**97.86**
Training Phase (80%)				
Normal	99.35	99.34	99.36	99.35
Anomaly	99.35	99.37	99.34	99.35
Average	**99.35**	**99.35**	**99.35**	**99.35**
Testing Phase (20%)				
Normal	99.32	99.39	99.28	99.33
Anomaly	99.32	99.26	99.37	99.32
Average	**99.32**	**99.32**	**99.33**	**99.32**

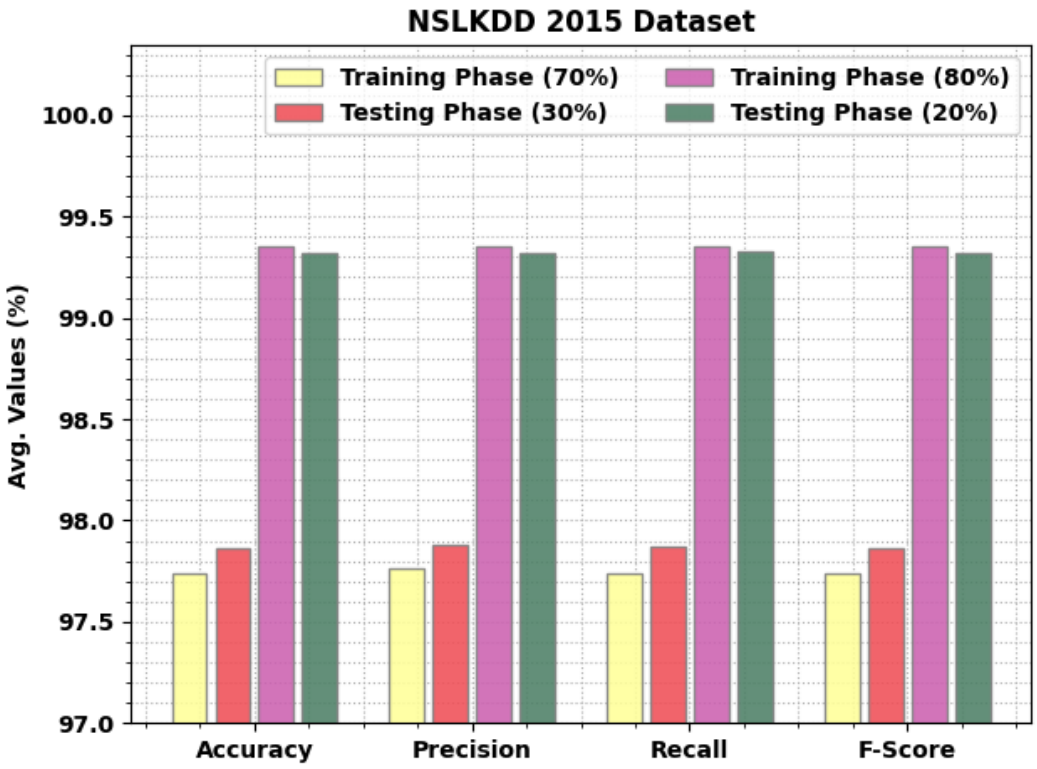

Figure 4. Average analysis of SFSA-DLIDS approach under NSLKDD 2015 dataset.

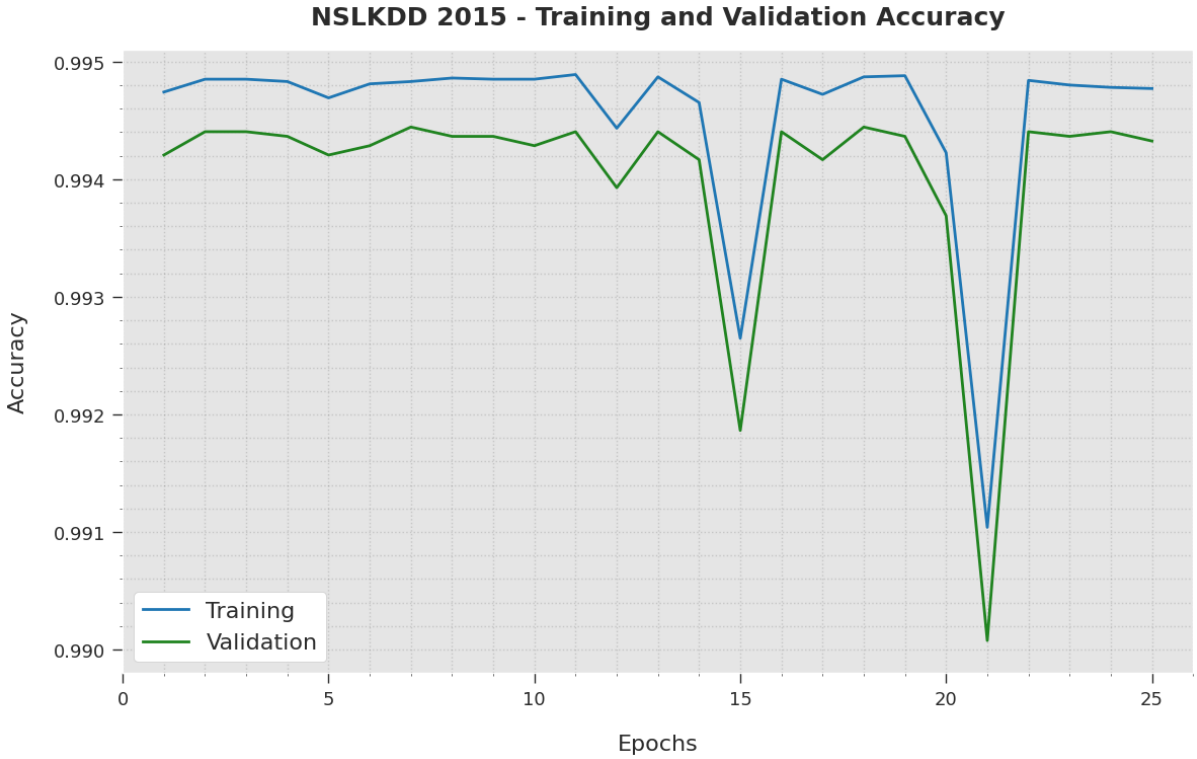

Figure 5. TA and VA analysis of SFSA-DLIDS approach under NSLKDD 2015 dataset.

The training loss (TL) and validation loss (VL) obtained by the SFSA-DLIDS methodology on NSLKDD 2015 dataset are accomplished in Figure 6. The experimental outcome represented that the SFSA-DLIDS technique exhibited minimal values of TL and VL. Particularly, the VL is less than TL.

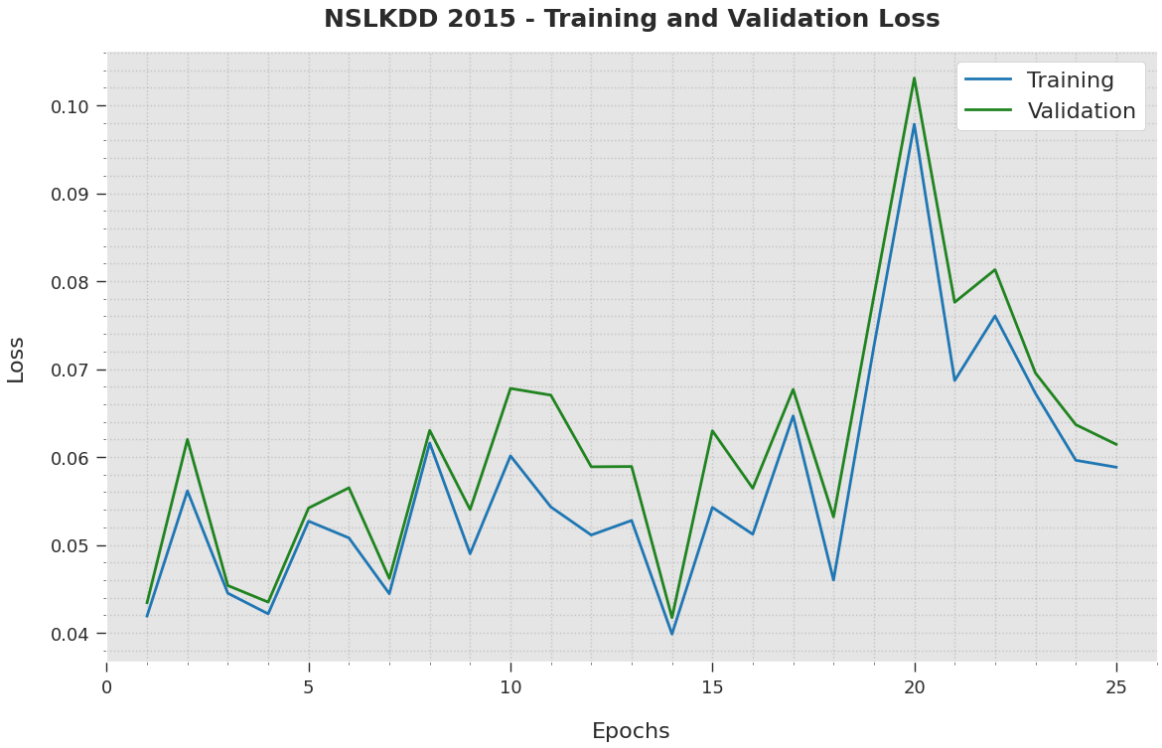

Figure 6. TL and VL analysis of SFSA-DLIDS approach under NSLKDD 2015 dataset.

Figure 7 represents the confusion matrices generated by the SFSA-DLIDS algorithm on the test CICIDS 2017 dataset. With 70% of TR dataset, the SFSA-DLIDS methodology recognized 33,748 samples into normal class and 34,669 samples into anomaly class. Moreover, with 30% of TS dataset, the SFSA-DLIDS approach recognized 14,607 samples into normal class and 14,752 samples into anomaly class. Along with that, with 20% of TS dataset, the SFSA-DLIDS method recognized 10,026 samples into normal class and 9839 samples into anomaly class.

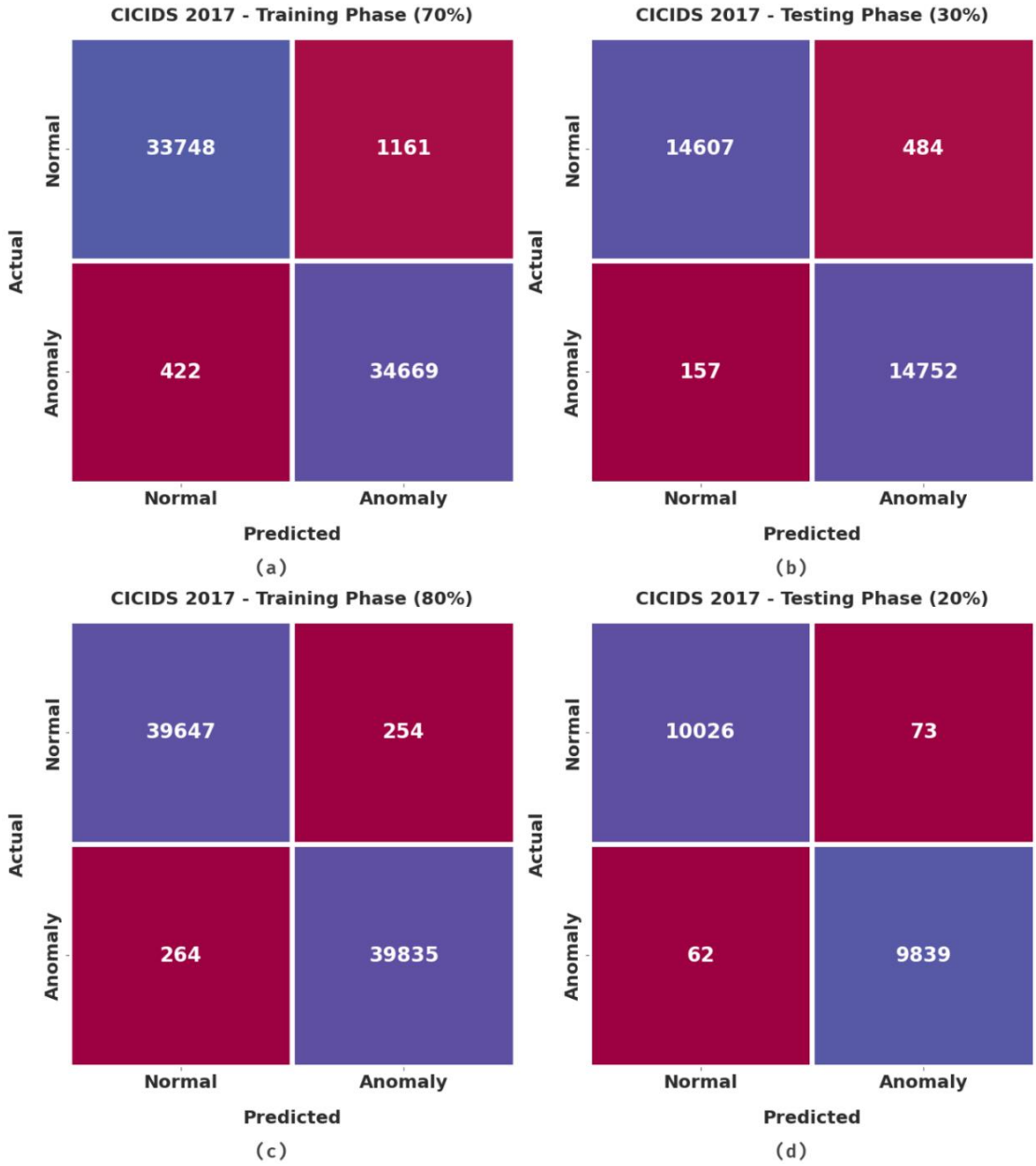

Figure 7. Confusion matrices of SFSA-DLIDS approach under CICIDS 2017 dataset: (**a**) 70% of TR data, (**b**) 30% of TS data, (**c**) 80% of TR data, and (**d**) 20% of TS data.

Table 3 and Figure 8 show the overall classification output of the SFSA-DLIDS technique on the test CICIDS 2017 dataset. The results portrayed that the SFSA-DLIDS approach resulted to improvised results under all aspects. For example, with 70% of TR data, the 3-DLIDS method rendered average $accu_y$ of 98.45%, $prec_n$ of 98.51%, $reca_l$ of 98.39%, and F_{score} of 98.44%. In the meantime, with 30% of TS data, the SFSA-DLIDS technique presented average $accu_y$ of 98.46%, $prec_n$ of 98.52%, $reca_l$ of 98.39%, and F_{score} of 98.45%.

Simultaneously, with 20% of TS data, the SFSA-DLIDS approach offered average $accu_y$ of 99.44%, $prec_n$ of 99.44%, $reca_l$ of 99.44%, and F_{score} of 99.44%.

Table 3. Result analysis of SFSA-DLIDS approach with various measures on CICIDS 2017 dataset.

Class Labels	Accuracy	Precision	Recall	F-Score
Training Phase (70%)				
Normal	98.45	97.79	99.35	98.56
Anomaly	98.45	99.24	97.42	98.32
Average	**98.45**	**98.51**	**98.39**	**98.44**
Testing Phase (30%)				
Normal	98.46	97.79	99.37	98.57
Anomaly	98.46	99.26	97.40	98.32
Average	**98.46**	**98.52**	**98.39**	**98.45**
Training Phase (80%)				
Normal	99.48	99.49	99.54	99.52
Anomaly	99.48	99.47	99.42	99.44
Average	**99.48**	**99.48**	**99.48**	**99.48**
Testing Phase (20%)				
Normal	99.44	99.47	99.49	99.48
Anomaly	99.44	99.41	99.40	99.40
Average	**99.44**	**99.44**	**99.44**	**99.44**

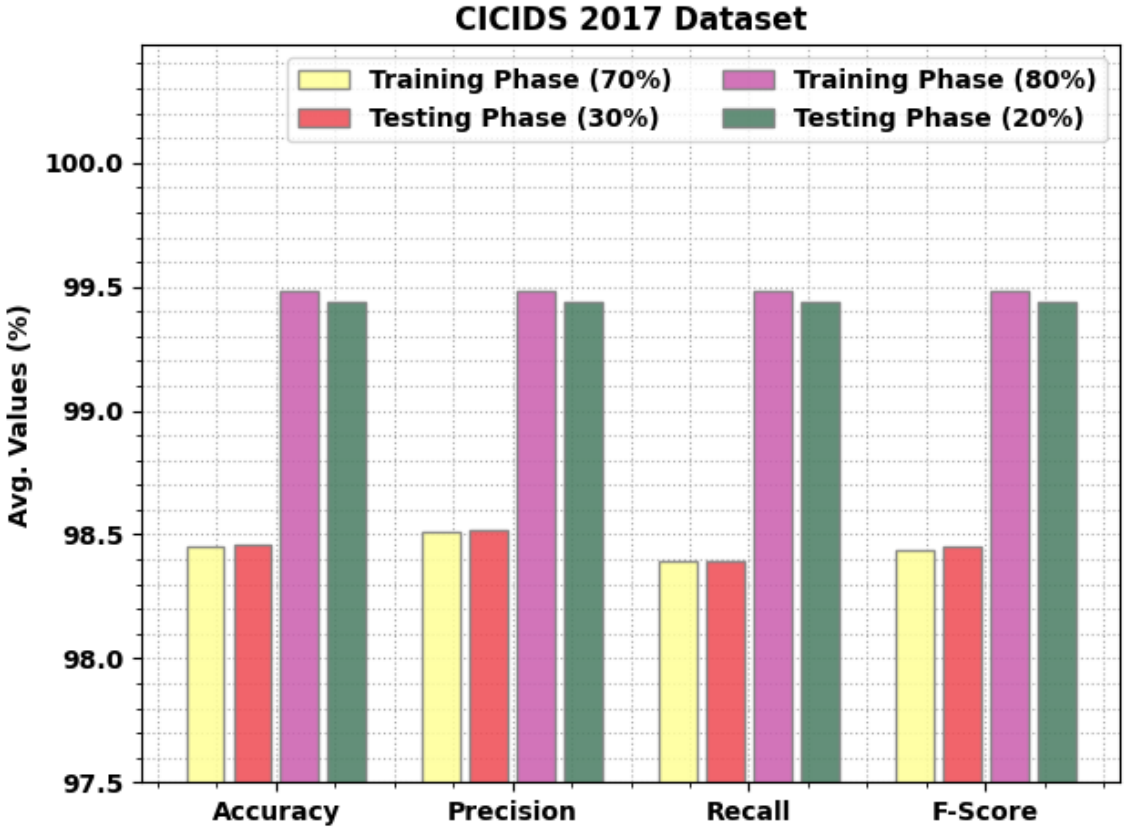

Figure 8. Average analysis of SFSA-DLIDS approach under CICIDS 2017 dataset.

The TA and VA attained by the SFSA-DLIDS algorithm on CICIDS 2017 dataset are demonstrated in Figure 9. The experimental outcome shows the SFSA-DLIDS algorithm gained higher values of TA and VA. To be specific, the VA is higher than TA.

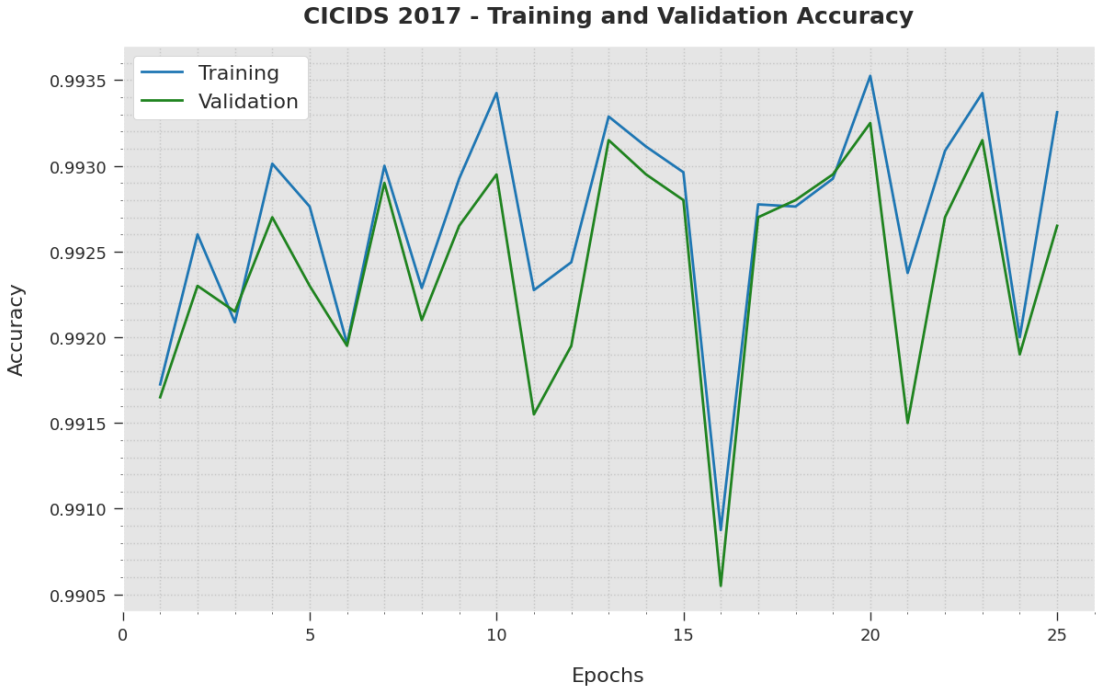

Figure 9. TA and VA analysis of SFSA-DLIDS approach under CICIDS 2017 dataset.

The TL and VL acquired by the SFSA-DLIDS technique on CICIDS 2017 dataset are exhibited in Figure 10. The experimental outcome denoted the SFSA-DLIDS approach accomplished minimal values of TL and VL. Particularly, the VL is lesser than TL.

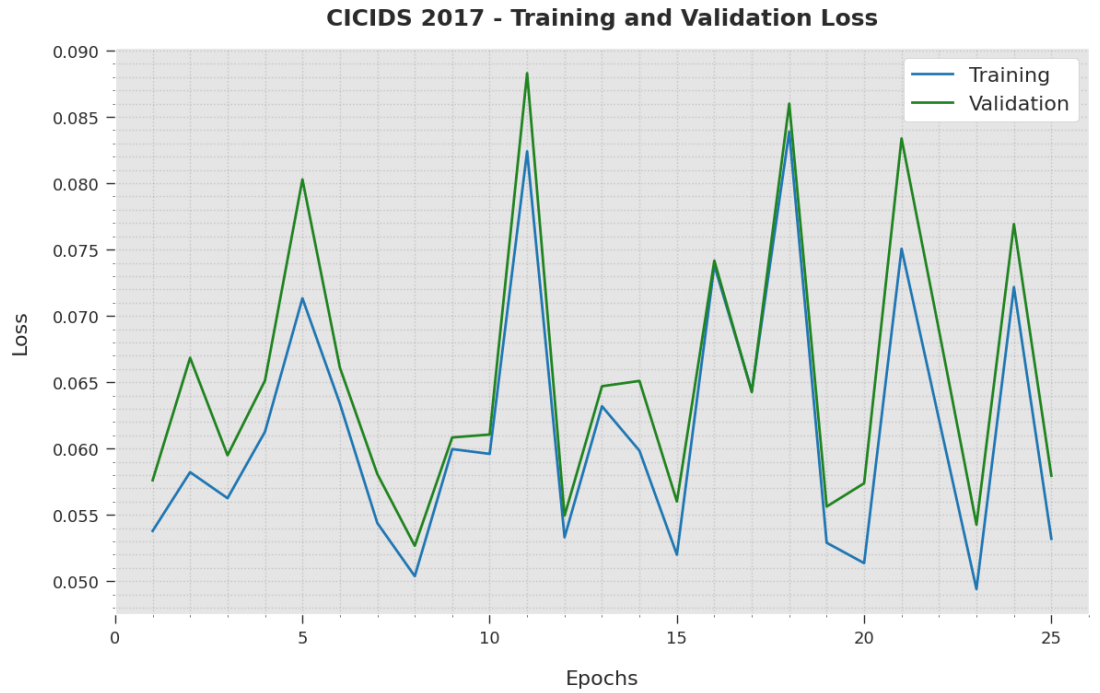

Figure 10. TL and VL analysis of SFSA-DLIDS approach under CICIDS 2017 dataset.

For ensuring the enhanced performance of the SFSA-DLIDS model, a comparative examination is made in Table 4 [3,13]. The results implied that the WISARD, Forest-PA, and LIB-SVM models have obtained lower $accu_y$ values of 96.22%, 96.53%, and 96.56% respectively. Followed by the GSAE and AE-RF models which attained slightly enhanced $accu_y$ values of 97.44% and 97.55%, respectively. Though the FURIA model resulted in reasonable $accu_y$ of 98.82%, the SFSA-DLIDS model accomplished maximum $accu_y$ of 99.44%. From the detailed results and discussion, it is obvious that the SFSA-DLIDS model has shown enhanced security in the CPS environment.

Table 4. Comparative analysis of SFSA-DLIDS approach with existing algorithms.

Methods	Accuracy	Precision	Recall	F1-Score
SFSA-DLIDS	99.44	99.44	99.44	99.44
GSAE	97.44	96.44	98.79	97.74
AE-RF	97.55	97.08	98.15	97.66
WISARD	96.22	97.27	96.85	98.75
Forest-PA	96.53	96.99	96.85	97.88
LIB-SVM	96.56	97.38	97.32	97.75
FURIA	98.82	97.83	96.94	98.55

5. Conclusions

In this article, an innovative SFSA-DLIDS method was devised for the classification and identification of intrusions from the CPS environment. The presented SFSA-DLIDS model primarily performed a min-max data normalization approach to convert the input data to a compatible format, followed by the SFSA technique which was applied to select a subset of features. Finally, the CSO-DSAE approach was utilized for the identification and classification of intrusions. The design of the CSO algorithm majorly focuses on the parameter optimization of the DSAE model and thereby enhances the classifier results. The experimental validation of the SFSA-DLIDS model was tested using a series of experiments. The experimental results established the enhanced performance of the SFSA-DLIDS method over the existing ones with maximum accuracy of 99.35% and 99.48% on the test NSLKDD 2015 and CICIDS 2017 datasets, respectively. Therefore, the presented SFSA-DLIDS model was implemented as an effectual tool to recognize intrusions in the CPS environment. In future, outlier detection approaches should be integrated for improving the overall detection efficiency of the SFSA-DLIDS technique. In addition, the proposed model can be realized on a big data environment in our future work.

Author Contributions: Conceptualization, M.A.D. and K.A.A.; methodology, F.S.A.; software, I.Y.; validation, S.S.A., K.A.A.; formal analysis, E.M.T.E.D.; investigation, K.A.A.; resources, A.A.A.; data curation, A.A.A.; writing—original draft preparation, S.S.A.; writing—review and editing, K.A.A.; visualization, A.M.; supervision, F.S.A.; project administration, A.M.; funding acquisition, F.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project number PNURSP2022R319, 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, grant number 22UQU4210118DSR34.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated during the current study.

Conflicts of Interest: The authors declare that they have no conflict of interest. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

References

1. Schneble, W.; Thamilarasu, G. Optimal feature selection for intrusion detection in medical cyber-physical systems. In Proceedings of the 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 18–20 December 2019; pp. 238–243.
2. Wickramasinghe, C.S.; Marino, D.L.; Amarasinghe, K.; Manic, M. Generalization of deep learning for cyber-physical system security: A survey. In Proceedings of the IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 21–23 October 2018; pp. 745–751.
3. Thakur, S.; Chakraborty, A.; De, R.; Kumar, N.; Sarkar, R. Intrusion detection in cyber-physical systems using a generic and domain specific deep autoencoder model. *Comput. Electr. Eng.* **2021**, *91*, 107044. [CrossRef]
4. Teyou, D.; Kamdem, G.; Ziazet, J. Convolutional neural network for intrusion detection system in cyber physical systems. *arXiv* **2019**, arXiv:1905.03168.
5. Al-Qarafi, A.; Alrowais, F.; Alotaibi, S.S.; Nemri, N.; Al-Wesabi, F.N.; Al Duhayyim, M.; Marzouk, R.; Othman, M.; Al-Shabi, M. Optimal Machine Learning Based Privacy Preserving Blockchain Assisted Internet of Things with Smart Cities Environment. *Appl. Sci.* **2022**, *12*, 5893. [CrossRef]
6. Panigrahi, R.; Borah, S.; Pramanik, M.; Bhoi, A.K.; Barsocchi, P.; Nayak, S.R.; Alnumay, W. Intrusion detection in cyber–physical environment using hybrid Naïve Bayes—Decision table and multi-objective evolutionary feature selection. *Comput. Commun.* **2022**, *188*, 133–144. [CrossRef]
7. Albraikan, A.A.; Hassine, S.B.H.; Fati, S.M.; Al-Wesabi, F.N.; Hilal, A.M.; Motwakel, A.; Hamza, M.A.; Al Duhayyim, M. Optimal Deep Learning-based Cyberattack Detection and Classification Technique on Social Networks. *Comput. Mater. Contin.* **2022**, *72*, 907–923.
8. Yadav, S.; Kalpana, R. A Survey on Network Intrusion Detection Using Deep Generative Networks for Cyber-Physical Systems. In *Artificial Intelligence Paradigms for Smart Cyber-Physical Systems*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 137–159.
9. Alohali, M.A.; Al-Wesabi, F.N.; Hilal, A.M.; Goel, S.; Gupta, D.; Khanna, A. Artificial intelligence enabled intrusion detection systems for cognitive cyber-physical systems in industry 4.0 environment. *Cogn. Neurodyn.* **2022**. [CrossRef]
10. Maleh, Y. Machine learning techniques for IoT intrusions detection in aerospace cyber-physical systems. In *Machine Learning and Data Mining in Aerospace Technology*; Springer: Cham, Switzerland, 2020; pp. 205–232.
11. Jamal, A.A.; Majid, A.-A.M.; Konev, A.; Kosachenko, T.; Shelupanov, A. A review on security analysis of cyber physical systems using Machine learning. *Mater. Today Proc.* **2021**. [CrossRef]
12. Sharma, M.; Elmiligi, H.; Gebali, F. A Novel Intrusion Detection System for RPL-Based Cyber–Physical Systems. *IEEE Can. J. Electr. Comput. Eng.* **2021**, *44*, 246–252. [CrossRef]
13. Alkayem, N.F.; Shen, L.; Asteris, P.G.; Sokol, M.; Xin, Z.; Cao, M. A new self-adaptive quasi-oppositional stochastic fractal search for the inverse problem of structural damage assessment. *Alex. Eng. J.* **2021**, *61*, 1922–1936. [CrossRef]
14. Li, B.; Wu, Y.; Song, J.; Lu, R.; Li, T.; Zhao, L. DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber–Physical Systems. *IEEE Trans. Ind. Inform.* **2020**, *17*, 5615–5624. [CrossRef]
15. de Araujo-Filho, P.F.; Kaddoum, G.; Campelo, D.R.; Santos, A.G.; Macedo, D.; Zanchettin, C. Intrusion Detection for Cyber–Physical Systems Using Generative Adversarial Networks in Fog Environment. *IEEE Internet Things J.* **2020**, *8*, 6247–6256. [CrossRef]
16. Althobaiti, M.M.; Kumar, K.P.M.; Gupta, D.; Kumar, S.; Mansour, R.F. An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems. *Measurement* **2021**, *186*, 110145. [CrossRef]
17. Gao, Y.; Chen, J.; Miao, H.; Song, B.; Lu, Y.; Pan, W. Self-Learning Spatial Distribution-Based Intrusion Detection for Industrial Cyber-Physical Systems. *IEEE Trans. Comput. Soc. Syst.* **2022**, 1–10. [CrossRef]
18. Ibor, A.E.; Okunoye, O.B.; Oladeji, F.A.; Abdulsalam, K.A. Novel Hybrid Model for Intrusion Prediction on Cyber Physical Systems' Communication Networks based on Bio-inspired Deep Neural Network Structure. *J. Inf. Secur. Appl.* **2022**, *65*, 103107. [CrossRef]
19. Kaddoura, S.; Arid, A.E.; Moukhtar, M. Evaluation of Supervised Machine Learning Algorithms for Multi-class Intrusion Detection Systems. In Proceedings of the Future Technologies Conference, Vancouver, BC, Canada, 28–29 October 2021; Springer: Cham, Switzerland, 2021; pp. 1–16.
20. Quincozes, S.E.; Passos, D.; Albuquerque, C.; Mossé, D.; Ochi, L.S. An extended assessment of metaheuristics-based feature selection for intrusion detection in CPS perception layer. *Ann. Telecommun.* **2022**, 1–15. [CrossRef]
21. Nagarajan, S.M.; Deverajan, G.G.; Bashir, A.K.; Mahapatra, R.P.; Al-Numay, M.S. IADF-CPS: Intelligent Anomaly Detection Framework towards Cyber Physical Systems. *Comput. Commun.* **2022**, *188*, 81–89. [CrossRef]
22. Wang, Z.; Li, Z.; He, D.; Chan, S. A lightweight approach for network intrusion detection in industrial cyber-physical systems based on knowledge distillation and deep metric learning. *Expert Syst. Appl.* **2022**, *206*, 117671. [CrossRef]
23. Çelik, E. Improved stochastic fractal search algorithm and modified cost function for automatic generation control of interconnected electric power systems. *Eng. Appl. Artif. Intell.* **2020**, *88*, 103407. [CrossRef]

24. Adem, K. Diagnosis of breast cancer with Stacked autoencoder and Subspace kNN. *Phys. A Stat. Mech. Appl.* **2020**, *551*, 124591. [CrossRef]
25. Meng, X.; Liu, Y.; Gao, X.; Zhang, H. A new bio-inspired algorithm: Chicken swarm optimization. *Adv. Swarm Intell.* **2014**, *5*, 86–94.
26. Fu, C.; Li, G.-Q.; Lin, K.-P.; Zhang, H.-J. Short-Term Wind Power Prediction Based on Improved Chicken Algorithm Optimization Support Vector Machine. *Sustainability* **2019**, *11*, 512. [CrossRef]
27. NSL-KDD Dataset. Available online: https://www.unb.ca/cic/datasets/nsl.html (accessed on 4 June 2022).
28. CICIDS 2017 Dataset. Available online: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 4 June 2022).