applied sciences

MDPI

Article

Optimal Machine Learning Based Privacy Preserving Blockchain Assisted Internet of Things with Smart Cities Environment

A. Al-Qarafi [1], Fadwa Alrowais [2], Saud S. Alotaibi [3], Nadhem Nemri [4], Fahd N. Al-Wesabi [4,*], Mesfer Al Duhayyim [5], Radwa Marzouk [6], Mahmoud Othman [7] and M. Al-Shabi [8]

[1] Department of Information Systems, College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia; aqarafi@taibahu.edu.sa

[2] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; falrowais@pnu.edu.sa

[3] Department of Information Systems, College of Computing and Information System, Umm Al-Qura University, Mecca 24382, Saudi Arabia; sotibi@uqu.edu.sa

[4] Department of Computer Science, College of Science & Arts at Mahayil, King Khalid University, Abha 62529, Saudi Arabia; nnemri@kku.edu.sa

[5] Department of Computer Science, College of Sciences and Humanities-Aflaj, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia; malduhayyim@psau.edu.sa

[6] Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; rmarzouk@pnu.edu.sa

[7] Department of Computer Science, Faculty of Computers and Information Technology, Future University in Egypt, New Cairo 11835, Egypt; abzwesabi@gmail.com

[8] Department of Management Information System, College of Business Administration, Taibah University, Medina 42353, Saudi Arabia; mrbwesabi@gmail.com

* Correspondence: falwesabi@kku.edu.sa

Abstract: Currently, the amount of Internet of Things (IoT) applications is enhanced for processing, analyzing, and managing the created big data from the smart city. Certain other applications of smart cities were location-based services, transportation management, and urban design, amongst others. There are several challenges under these applications containing privacy, data security, mining, and visualization. The blockchain-assisted IoT application (BIoT) is offering new urban computing to secure smart cities. The blockchain is a secure and transparent data-sharing decentralized platform, so BIoT is suggested as the optimum solution to the aforementioned challenges. In this view, this study develops an Optimal Machine Learning-based Intrusion Detection System for Privacy Preserving BIoT with Smart Cities Environment, called OMLIDS-PBIoT technique. The presented OMLIDS-PBIoT technique exploits BC and ML techniques to accomplish security in the smart city environment. For attaining this, the presented OMLIDS-PBIoT technique employs data pre-processing in the initial stage to transform the data into a compatible format. Moreover, a golden eagle optimization (GEO)-based feature selection (FS) model is designed to derive useful feature subsets. In addition, a heap-based optimizer (HBO) with random vector functional link network (RVFL) model was utilized for intrusion classification. Additionally, blockchain technology is exploited for secure data transmission in the IoT-enabled smart city environment. The performance validation of the OMLIDS-PBIoT technique is carried out using benchmark datasets, and the outcomes are inspected under numerous factors. The experimental results demonstrate the superiority of the OMLIDS-PBIoT technique over recent approaches.

Keywords: blockchain assisted IoT; smart city; security; privacy preserving; feature selection; intrusion detection

check for updates

Citation: Al-Qarafi, A.; Alrowais, F.; S. Alotaibi, 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. https://doi.org/10.3390/app12125893

Academic Editor: Faisal Jamil

Received: 4 May 2022
Accepted: 7 June 2022
Published: 9 June 2022

1. Introduction

The smart city is regarded as a technological structure utilized by the distinct city shareholders for achieving distinct objectives such as good governance, enhancing day-to-day living circumstances, maximizing resource usage, or building new commercial chances [1]. The supporters of the smart city, whether they are technology providers, researchers, or managers, have performed various designs, improvement studies, and standardization in response to the difficulties of smart city advancement referring to security, scalability, connectivity, or heterogeneity [2]. The Internet of Things (IoT) technology provides less cost and potential solutions for the advancement of smart cities owing to its greater suitability in an infinite number of cases. Specifically, the intelligent IoT gadgets could consistently combine multiple fields of a smart city phenomenon, interchanging a continual flow of data for granting quality services to the public as a supreme objective [3]. Thus, we refer to the IoT gadgets as brilliant, which means they are capable of communicating autonomously, demanding less or no human involvement.

Invaders will fix their places in smart cities for many reasons. Malevolent people might consider smart cities as play areas wherein they make a trail of their hacking skills through playing with existing technologies for individual fulfillment [4]. For cyber-criminals, the interconnectivity of systems and devices in a smart city can be operated for any feasible unauthorized accessibility of monetary properties, personal properties, delicate data, and causing damages or loss to the common public [5]. State-sponsored actors might exploit the ubiquity of smart city technologies for launching their hacktivist or espionage camps. In a few very dire situations, smart applications might be exploited for performances of terror [6]. Thus, engineers and researchers must grant legal methods and solutions to aid local administrations and metropolitan or urban developers in creating highly secured smart cities [7].

Security is not just a necessity for the further usage of blockchain (BC); it could participate in data dispersal since it operates in a quicker mode. However, according to the author's knowledge, autonomous mathematical evidence for quicker solutions is now not available [8]. Once after the existence of the mathematical evidence, an ideal ambiance is possible where speedy and trusted nodes are linked in the network and could avail the advantage from 5G with the help of particular fog layers of clouds. As a result, a BC-oriented decentralized system is one of the resolutions [9]. One such benefit of utilizing BC technologies is that it has capability of storing data in an absolute way that does not need a centralized database. Furthermore, it could offer a means for tracking and executing transactions amid several members in a trusted ambiance [10]. With the use of strong encryption with public private key sets, BC further grants higher levels of security to its members. Figure 1 depicts the process of BC in IoT with smart cities environment.

This study develops an Optimal Machine Learning-based Intrusion Detection System for Privacy Preserving BIoT with Smart Cities Environment, called OMLIDS-PBIoT technique. The proposed OMLIDS-PBIoT model can be applied to accomplish security in several areas of smart cities such as traffic management, emergency response, smart healthcare, air quality monitoring, disaster management, waste management, air quality monitoring, etc. The presented OMLIDS-PBIoT technique employs data pre-processing in the initial stage to transform the data into compatible format. Additionally, a golden eagle optimization (GEO)-based feature selection (FS) model is designed to derive useful feature subsets. In addition, heap-based optimizer (HBO) with random vector functional link network (RVFL) model was utilized for intrusion classification. In addition, BC technology is exploited for secure data transmission in the IoT-enabled smart city environment. The performance validation of the OMLIDS-PBIoT technique is performed utilizing benchmark datasets, and the outcomes are valued in many aspects.

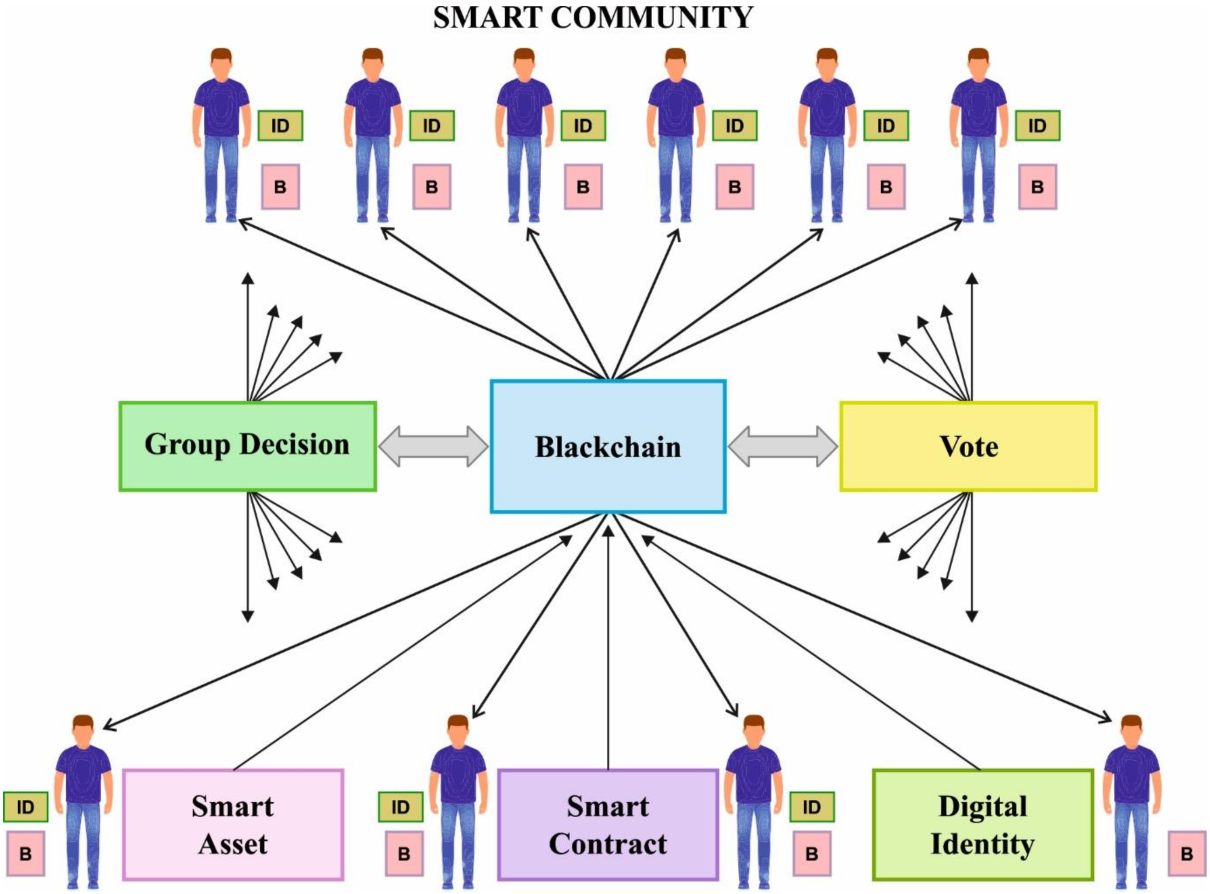

Figure 1. Blockchain in IoT with smart cities environment.

2. Literature Review

This section offers a brief review of security-based solutions for the IoT enabled smart city environment. Kumar et al. [11] introduce a Privacy Preserving and Secure Framework (PPSF) to IoT-driven smart cities. The suggested PPSF depends on two chief processes: one is a dual-level secrecy scheme, and another one is IDS. Firstly, in a dual-level secrecy scheme, a BC unit is devised for sending the IoT data securely, and a Principal Component Analysis (PCA) approach was implemented to convert original IoT data as to innovative outline. In the IDS, a Gradient Boosting Anomaly Detector (GBAD) was implied for training purposes. Meng et al. [12] recommend a BC-enabled single character frequency-related ESM, which could create a provable database of malicious payloads by means of BCs. In the valuation, we examine the act of the methods below flooding and character padding assaults in a real and simulated IoT network ambiance.

Bediya and Kumar [13] debated most feasible assaults on IoT network systems and distributed denial of service (DDoS) assault is such a risky one amongst them. BC technology could be used for developing a structure in order to protect or preserve IoT systems, and BC is a brand-new technology utilized in transacting processes of cryptocurrency. Rathore et al. [14] suggest a decentralized security substructure depends on Software Defined Networking (SDN) paired with BC technologies for IoT network from the smart cities, which hinges on the three main technologies of Fog, SDN, and BC, as well as mobile edge computing (MEC) for detecting assaults in the IoT network with higher effectiveness. Therefore, in the suggested substructure, SDN is accountable for continual observing and interpretation of traffic data in the whole IoT network for providing an optimum attack recognition method.

Botello et al. [15] suggest BlockSIEM, a BC-related and dispersed Security Information and Event Management (SIEM) solution structure for the safety of the above-mentioned smart city services. These security events were produced by IoT sentinels, which take responsibility for protecting groups of IoT gadgets. In [16], several various problems and distinct assault vectors are conferred, including the feasible results. To alleviate numerous assaults, BlockSDN, a BC as a service structure, for SDN is suggested. The substructure of permissioned BC is provided, pursued by two assault scenarios: one is a malware imperiled switch at data plane, and another one is DDoS assault at the control planes. Peneti et al. [17] inaugurates the BC-defined network system having a grey wolf enhanced modular neural network (NN) technique for managing the smart ambiance security. At the time of this process, translation, application, and construction, layers were formed, where user-authenticated related blocks are devised for handling the privacy and security property. After this, optimizing NN is implied for maintaining the computational resource utilization and latency in IoT-empowered smart applications.

The authors in [18] introduced a new AI-based multimodal fusion model to recognize intrusions in the industry 4.0 environment. This model involves improved fish swarm optimization-based feature selection (IFSO-FS) approach to choose optimal subset of features. In addition, the IFSO technique is derived by the use of the Levy Flight (LF) concept into the searching mechanism of the conventional FSO model. Additionally, the weighted voting-based ensemble technique is used for fusion procedure. The authors in [19] presented a hybrid metaheuristic-based energy efficiency resource allocation (HMEERA) technique for cloud platform. The HEERA technique carries out the feature extraction and principal component analysis (PCA)-based feature reduction. It also uses hybrid Group Teaching Optimization Algorithm (GTOA) with rat swarm optimizer (RSO) algorithm to allot the resources in an optimal way.

Though several models are available in the literature, it is still needed to design effective IDS models for smart city environment. In addition, most of the research works have chosen the parameter values based on the trial-and-error method, which is labor intensive and ineffective. Therefore, parameter optimization can be considered as the NP hard problem and can be solved by the use of metaheuristic algorithms. In this work, we have used the HBO algorithm for optimal parameter tuning process.

3. The Proposed Model

In this study, an effective OMLIDS-PBIoT technique was developed with the utilization of the BC and ML models for accomplishing security in the smart city environment. The presented OMLIDS-PBIoT technique employs different stages of subprocesses, namely preprocessing, GEO-FS, RVFL-based classification, and HBO-based parameter optimization. Moreover, BC technology is exploited for secure data transmission in the IoT enabled smart city environment.

3.1. Blockchain Technology

In this work, the BC technology is used to assure secure data transmission in the smart city environment. BC is a decentralized P2P network through which every transaction is authenticated by the registered node and recorded in an immutable and distributed ledger. In such case, the consensus approach is the center of the BC technique, since it ensures network reliability. Especially, no centralized authority exists to authenticate the produced event; all the transactions should be authenticated by the BC node via mutual agreement (viz., consensus) [20]. A few common types of consensuses are given in the following:

- Proof of Work (PoW): a transaction is authorized when the node accepts its P2P network.
- Proof of Stake (PoS): the node that has further wealth has great possibility to create a block and participate in the consensus.
- Proof of Importance (PoI): the node that can generate a block is the one with the maximum number of transactions.
- Proof of Authority (PoA): explicitly few nodes are permitted to generate new blocks and protect the BC. It should be apparent that the abovementioned algorithm features

potential drawbacks and possible advantages, depending mainly on the fundamental P2P network architecture.

PoS and PoW are explicitly portrayed as the more commonly used algorithms to achieve the consensus between the P2P nodes. Nonetheless, it has become evident that PoW needs high computation resources, whereas PoS is most demonstrated for attacking because the mining cost is almost zero. PoA and PoI are valid alternatives since they have better performance and are energy-friendly. Moreover, BC presents two techniques to construct a network, such as permissionless and permissioned BCs. Especially, permissionless BC (that is, public BC) allows a potential candidate to turn into a node and belong to the network. A node on this BC may execute other tasks as long as they pose the physical ability (for example, validate transactions, mine blocks, etc.). Consecutively, permissioned BC (that is, private BC) restricts the access of a node that belongs to the network and executing task. An appropriate characteristic of BC is that it is likely to select the levels of decentralization on the network, that is, partially decentralized or fully centralized. Specifically, open permissioned BC is partially decentralized, because all the entities have the capacity to read the saved information, whereas closed permissioned BC is completely centralized. After all, the saved information is perceptible to the participating node.

3.2. Data Pre-Processing

During this case, min-max normalized has been executed for scaling the data s to unit variance. It is commonly utilized to compute the similarity degree among points. Assume that A is data, which is mapped from the dataset range from A_{min} to A_{max}, employing in Equation (1):

$$A_{normalized} = \frac{A - A_{min}}{A_{max} \; - A_{min}} \tag{1}$$

The employ of min-max normalization ensures that the feature was exacted as to similar scales.

3.3. Process Involved in GEO-FS Technique

In this study, the GEO-FS technique is exploited to choose an optimal subset of features from the preprocessed data. GEO is a novel meta-heuristic approach, which was established very recently for solving global optimized problems. The GEO technique was simulated and mathematically processed by the intelligence of golden eagle (GE) dependent upon adjusting the speed of its spiral tracking [21]. The GE is a specific type of swarm that has a superior propensity for cruising around and searching for prey initially to hunt. By controlling these two elements, i.e., cruise propensity and attack propensity, the GEO was rapidly capable of hunting an optimum accessible prey from the possible region.

The GE from cruise and hunt is a unique feature, i.e., it follows in a spiral trajectory, representing that prey was commonly on one side of the eagles. This allows them to control target prey wisely and use boulders for determining an appropriate angle of attack. Simultaneously, it can be checking other regions for optimum food. The hunting technique of GEs mostly dependent upon the subsequent feature.

The mathematical designs of GEs for mimicking the movement to search for the prey are mostly explained as:

- The spiral movement of GEs: In GEO, all the GE retains from their memory the optimum visited place previously. The eagle is an attraction near the cruising and nearing attack of the prey concurrently in order to search for optimum food. At all iterations, all the GEs j arbitrarily select prey, which is fixed by another GE l, and circles around optimum place stayed by GE l so far. The GE j also has the chosen features for circling their memory; therefore, it can be $l \in \{1, 2, \cdots, N_{GE}\}$, whereas N_{GE} stands for the number of GEs.
- Prey selection: At all the iterations, all the GEs can choose a prey for executing the cruising as well as attacking functions. Additionally, all the GEs select the chosen

prey in the memory of the entire flock. Thus, the cruising and attacking vectors are calculated based on the chosen prey. Next, it verifies their memory when the novel place was superior to the preceding place. Subsequently, the memory was upgraded with novel determining.

- Attack: The attack is defined by utilizing a vector deriving from the actual place of GE j and ending from the place of prey from the eagle's memory as:

$$\vec{A}_j = \vec{X}_l^* - \vec{X}_j, \tag{2}$$

where $\vec{A}_{j_*}$ implies the attack vector of GEs j, $\vec{X}_1^*$ denotes the optimum place visited by eagle l so far, and X_j signifies the present place of eagles j.

- Cruise: The cruise vector is a perpendicular vector to attack vectors and tangent to circle. It is also well-known as linear speed of GEs for attacking the prey. The target point on cruise vector was provided under:

$$\vec{C}_j = \frac{d - \sum_{f,f\neq j} a_f}{a_j{}'}, \tag{3}$$

where d signifies the hyperplane formula from n-dimension space, $a_j, a_f \in \vec{A}_j$, whereas $\vec{A}_j = [a_1, a_2, \cdots, a_n]$ signifies the attack vectors.

- Moving to novel places: Moving to novel places of GEs was mostly dependent upon the attacking and cruising vectors. Thus, the step vector of GEs j in iteration t was projected by the subsequent formula:

$$\triangle_{x_j} = \vec{r}_1 p_a \frac{\vec{A}_j}{\|\vec{A}_j\|} + \vec{r}_2 p_c \frac{\vec{C}_j}{\|\vec{c}_j\|} \tag{4}$$

In which p_a^t denotes the attack co-efficient at iteration t and p_c^t implies the cruise co-efficient at iteration t and controlling that the GE was affected by cruising and attacking. $\vec{r}_1$ and $\vec{r}_2$ imply the random vectors. A novel place of GEs is then provided as:

$$x_j^{t+1} = x_j^r + \triangle_{x_j}^t \tag{5}$$

When the fitness function j offers superior to the preceding places after their memory is upgraded with novel places.

- The transition from exploration to exploitation: The GEO technique utilizes the attacking co-efficient p_a and the cruising co-efficient p_c for switching from the state of exploration to exploitation. p_a and p_c is calculated utilizing the subsequent linear expression:

$$p_a = p_a^0 + \frac{t}{T}\left|p_a^T - p_a^0\right|, \tag{6}$$

$$p_c = p_c^0 + \frac{t}{T}\left|p_c^T - p_c^0\right|, \tag{7}$$

where p_a^0 and p_c^0 are correspondingly the primary values to propensity for attacking p_a and for propensity for cruising p_c, t signifies the present iteration, T denotes the maximal count of iterations, p_a^T and p_a^T are correspondingly the last values to propensity for attacking p_a and to propensity for cruising p_c.

The fitness function (FF) of the GEO-FS technique assumes the classifier accuracy and the number of chosen features. It maximizes the classifier accuracy and minimizing the set size of chosen features. Thus, the subsequent FF utilized for evaluating individual solutions is demonstrated in Equation (8).

$$Fitness = \alpha \times ErrorRate + (1-\alpha) \times \frac{\#SF}{\#All\_F} \tag{8}$$

where $ErrorRate$ implies the classifier error rate utilizing the chosen features. $ErrorRate$ is computed as the percentage of inappropriate classification (by 5-ANN classifier) to the number of classifiers developed, formulated as a value amongst zero and one [22]. ($ErrorRate$ implies the complement of classifier accuracy), $\#SF$ denotes the number of chosen features, and $\#All\_F$ denotes the entire number of elements from the original dataset. α denotes utilized for controlling the significance of classifier quality as well as subset length. During the experiments, α is fixed to 0.9.

3.4. Process Involved in HBO-RVFL Technique

At this stage, the chosen features are passed into the RVFL model to classify data. The basic framework of RVFL is generally the same as the artificial neural network that comprises hidden, output, and input layers [23]. However, the major difference between RVFL and ANN is that RVFL directly connects the output and input layers. This connection supports RVFL by an appropriate mechanism to prevent the over-fitting problems that take place in ANN. The training RVFL begins with the employment of the input data that comprise target y_i and sample x_i, to the input state. Next, the output of the hidden node is computed by the subsequent formula:

$$O_j(\alpha_j x_i + \beta_j) = \frac{1}{1 + e^{-(\alpha_j x_i + \beta_j)}},\ \beta_j \in [0,\ S],\ \alpha_j \in [-S,\ S] \tag{9}$$

From the above equation, α_j represents the value of weight, which connects the hidden and input nodes. S and β_j denote the scale bias and factor, correspondingly.

Then, last output is calculated by the Equation (11):

$$Z = Fw,\ w \in R^{n+P},\ F = [F_1,\ F_2] \tag{10}$$

$$F_1 = \begin{bmatrix} x_{11} & x_{1n} \\ \vdots & \vdots \\ x_{N1} & x_{Nn} \end{bmatrix} F_2 = \begin{bmatrix} G_1(\alpha_1 x_1 + \beta_1) & \cdots & G_P(\alpha_P x_1 + \beta_P) \\ \vdots & \ddots & \vdots \\ G_1(\alpha_1 x_N + \beta_1) & \cdots & G_P(\alpha_P x_N + \beta_P) \end{bmatrix} \tag{11}$$

Here, w is upgraded by the subsequent equation:

$$w = F^{\dagger} Z \tag{12}$$

In Equation (12), † indicates the Moore–Penrose pseudo-inverse. Figure 2 demonstrates the structure of RVFL technique.

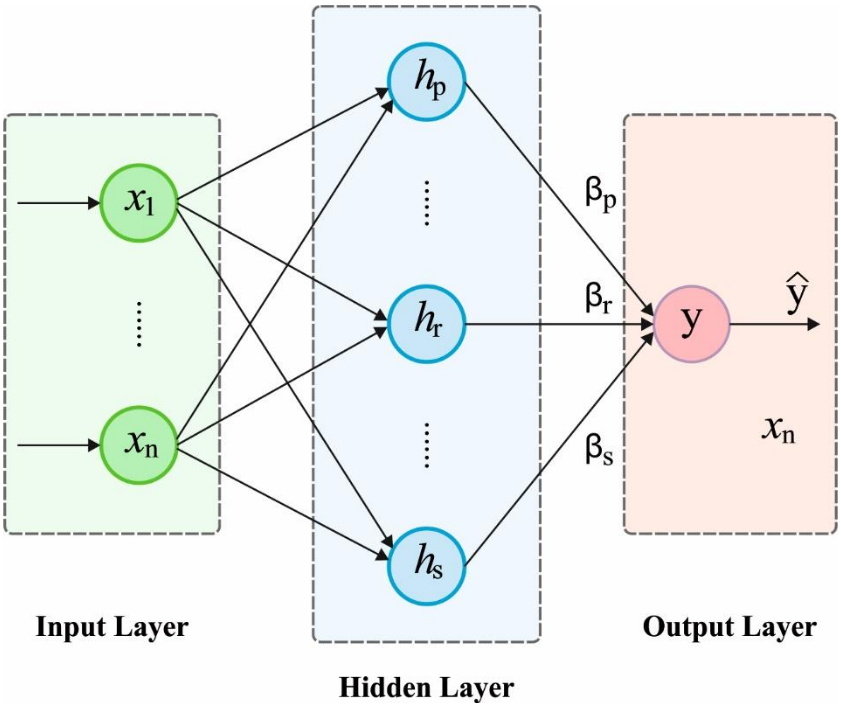

Figure 2. Structure of RVFL.

To optimally adjust the parameters related to the RVFL method, the HBO algorithm has been utilized. The presented algorithm is stimulated by the corporate rank hierarchy (CRH), which declares that a team could arrange in a hierarchal manner to fulfill organization goals [24]. The HBO algorithm is categorized as interaction between colleagues, interaction between subordinates, self-contribution of employees, and the immediate supervisor. In the CRH approach, the population is established as a CRH, while the heap node is established as a searching agent. The agent location of every search is upgraded by:

$$x_i^k(t+1) = B^k + \gamma(2r-1)\left|B^k - x_i^k(t)\right| \tag{13}$$

The k-th components of λ vector $\overrightarrow{\lambda}$ is signified as:

$$\lambda^k = 2r - 1 \tag{14}$$

γ is evaluated by using Equation (15):

$$\gamma = \left|2 - \frac{\left(tmod\frac{t}{c}\right)}{\frac{1}{4C}}\right| \tag{15}$$

The (C) parameter in Equation (16) controls the variation. However, this will complete in T iteration as in the following:

$$C = T^{max}/25 \tag{16}$$

Additionally, the interaction between colleagues is modeled. As expressed in Equation (17), the location of every agent $\left(\overrightarrow{x_i}\right)$ is upgraded by randomly designated colleague $\overrightarrow{S_r}$:

$$x_i^k(t+1) = \begin{cases} S_r^k + \gamma\lambda^k\left|S_r^k - x_i^k(t)\right|, & f(\overrightarrow{S_r} < f\left(\overrightarrow{\chi_i}(t)\right) \\ x_i^k + \gamma\lambda^k\left|S_r^k - x_i^k(t)\right|, & f(\overrightarrow{S_r} \geq f\left(\overrightarrow{\chi_i}(t)\right) \end{cases} \tag{17}$$

In Equation (17), the fitness of the searching agent is denoted as f.

Further modeled is the self-contribution of all the employees, whereby the location of every agent is upgraded based on the following equation:

$$x_i^k(t+1) = x_i^k(t) \tag{18}$$

At last, the updating position equation has been taking place. The roulette wheel probability, p_1, p_2, and p_3, are designated to balance the exploitation and exploration stages.

The searching agent upgrades the location by utilizing Equation (18). Selection of the proportion p_1 is implemented by Equation (19):

$$p_1 = 1 - \frac{t}{\tau^{\max}} \tag{19}$$

The searching agent upgrades their location by Equation (11). Selection of the proportion p_2 is implemented by Equation (20):

$$p_2 = p_1 + \frac{1-p_1}{2} \tag{20}$$

The searching agent upgrades their location by Equation (19). Selection of the proportion p_3 is implemented by Equation (21):

$$p_3 = p_2 + \frac{1-p_1}{2} = 1 \tag{21}$$

Therefore, the common position updating method of the HBO algorithm is expressed by the following equation:

$$x_i^k(t+1) = \begin{cases} x_i^k(t),\ p \le p_1 \\ B^k + \gamma\lambda^k \left| B^k - x_i^k(t) \right|,\ p_1 < p < p_2 \\ S_r^k + \gamma\lambda^k \left| S_r^k - x_i^k(t) \right|, p_2 < p \le p_3\ and\ f(\overrightarrow{S_r} < f\left(\overrightarrow{x}_{\mathrm{i}}(t)\right) \\ x_\gamma^k + \gamma\lambda^k \left| S_r^k - x_i^k(t) \right|, p_2 < p \le p_3\ and\ f\left(\overrightarrow{S_r} \ge f\left(\overrightarrow{x}_{\mathrm{i}}(t)\right)\right) \end{cases} \tag{22}$$

The HBO approach grows an FF for achieving superior classifier performances. It solves a positive integer for demonstrating the best performance of candidate results. During this study, the minimized classifier error rate has been regarded that FF is offered in Equation (23).

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} \times 100 \tag{23}$$

4. Performance Validation

This section inspects the experimental validation of the OMLIDS-PBIoT technique using two benchmark datasets such as NSL-KDD and UNSW-NB15 datasets.

The FS outcomes of the GEO-FS system on the test dataset are displayed in Table 1. The experimental values implied the GEO-FS method has accomplished enhanced performance over other models. For instance, on test NSL-KDD dataset, the OMLIDS-PBIoT technique has accomplished least good cost of 0.002218 with the selection of 13 features. Additionally, on test UNSW-NB15 dataset, the OMLIDS-PBIoT technique has established minimum best cost of 0.003095 with the selection of 20 features.

Table 1. Results analysis of OMLIDS-PBIoT based FS on applied dataset.

Dataset	Best Cost	Selected Features
NSL-KDD	0.002218	1, 2, 5, 6, 8, 21, 23, 26, 29, 31, 33, 34, 37
UNSW-NB15	0.003095	1, 3, 4, 7, 11, 17, 18, 20, 21, 25, 27, 28, 30, 31, 32, 34, 38, 39, 40, 42

A detailed intrusion detection of the outcomes of the OMLIDS-PBIoT technique on the NSL-KDD dataset are portrayed in Table 2 and Figure 3. The outcomes denoted by the OMLIDS-PBIoT technique have classified samples under all classes effectively. For example, the OMLIDS-PBIoT technique has recognized samples under DoS attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.30%, 99.50%, 99.64%, and 99.57% respectively. Additionally, the OMLIDS-PBIoT technique has recognized samples under R2L attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.53%, 98.52%, 99.18%, and 99.21% correspondingly. At the same time, the OMLIDS-PBIoT technique has recognized samples under U2R attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.89%, 99.12%, 98.45%, and 99.54% individually. Moreover, the OMLIDS-PBIoT technique has recognized samples under normal attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 98.65%, 98.84%, 98.74%, and 99.07%, correspondingly.

Table 2. Result analysis of OMLIDS-PBIoT technique on NSL-KDD dataset.

Attack Type	Precision	Recall	F1-Score	Accuracy
DoS	99.30	99.50	99.64	99.57
R2L	99.53	98.52	99.18	99.21
Probe	98.54	99.53	99.44	99.68
U2R	99.89	99.12	98.45	99.54
Normal	98.65	98.84	98.74	99.07
Average	99.18	99.10	99.09	99.41

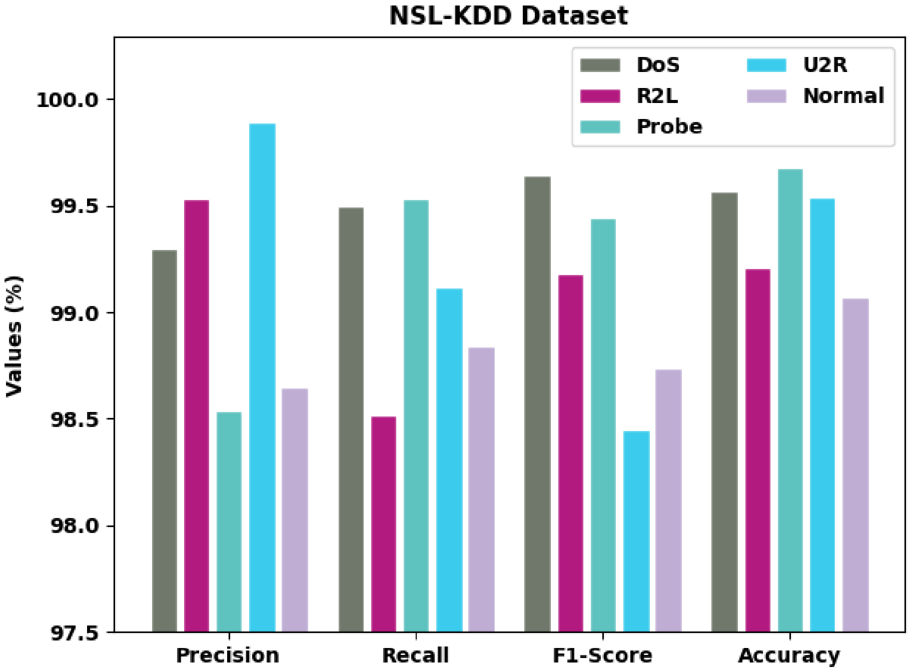

Figure 3. Result analysis of OMLIDS-PBIoT technique under NSL-KDD dataset.

The training accuracy (TA) and validation accuracy (VA) attained by the OMLIDS-PBIoT technique on NSL-KDD dataset is demonstrated in Figure 4. The experimental outcome implied that the OMLIDS-PBIoT technique has gained maximum values of TA and VA. In specific, the VA seemed to be higher than TA.

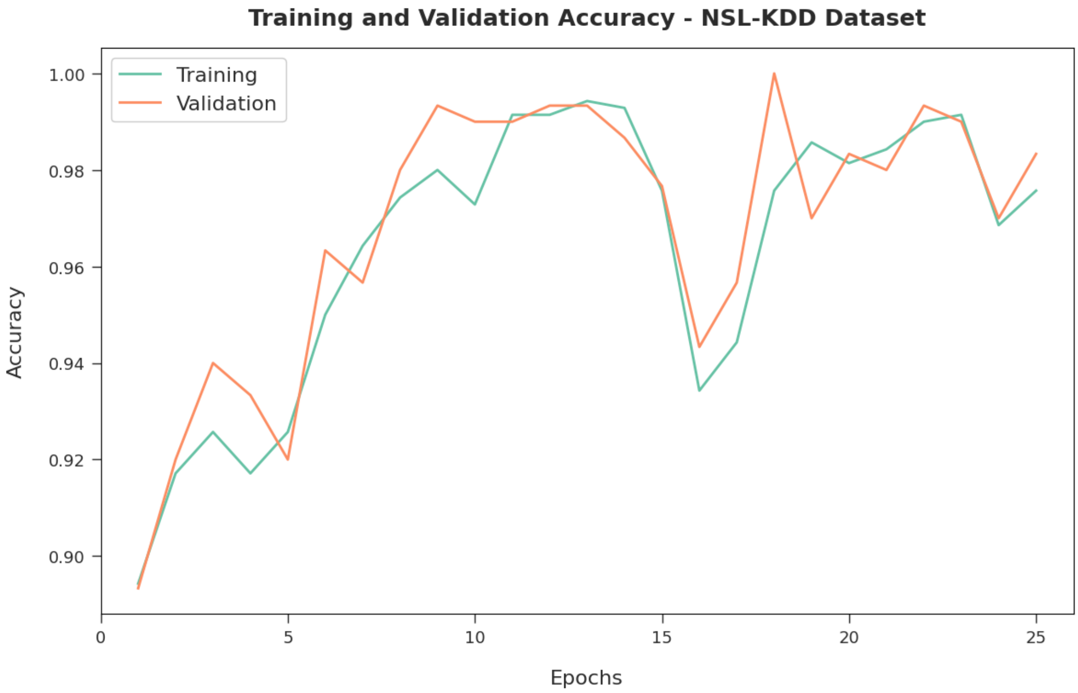

Figure 4. TA and VA analysis of OMLIDS-PBIoT technique under NSL-KDD dataset.

The training loss (TL) and validation loss (VL) achieved by the OMLIDS-PBIoT technique on NSL-KDD dataset are established in Figure 5. The experimental outcome inferred that the OMLIDS-PBIoT technique has accomplished least values of TL and VL. In specific, the VL seemed to be lower than TL.

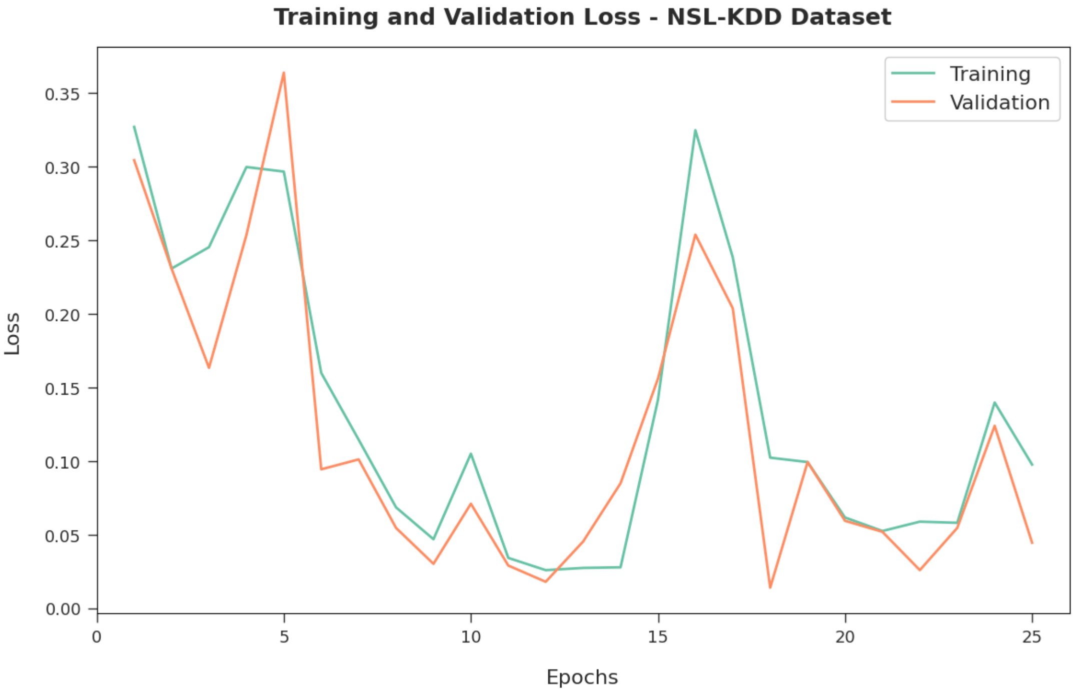

Figure 5. TL and VL analysis of OMLIDS-PBIoT technique under NSL-KDD dataset.

A detailed intrusion detection outcome of the OMLIDS-PBIoT technique on the UNSW-NB15 dataset is displayed in Table 3 and Figure 6. The outcomes exposed that the OMLIDS-PBIoT technique has classified samples under all classes effectively. For example, the OMLIDS-PBIoT model has recognized samples under DoS attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.69%, 99.33%, 99.54%, and 99.74%, correspondingly. Eventually, the OMLIDS-PBIoT technique has recognized samples under Generic attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.58%, 99.78%, 98.78%, and 99.67%, correspondingly. In the meantime, the OMLIDS-PBIoT technique has recognized samples under U2R attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.609%, 99.43%, 98.58%, and 99.54%, correspondingly. Furthermore, the OMLIDS-PBIoT technique has recognized samples under normal attack with $prec_n$, $reca_l$, $F1_{score}$, and $accu_y$ of 99.73%, 99.48%, 98.60%, and 99.65%, correspondingly.

Table 3. Result analysis of OMLIDS-PBIoT technique on UNSW-NB15 dataset.

Attack Type	Precision	Recall	F1-Score	Accuracy
Normal	99.72	99.11	99.54	99.49
DoS	99.69	99.33	99.54	99.74
Shellcode	99.74	99.61	99.70	99.50
Backdoor	99.58	99.78	98.78	99.67
Generic	99.58	99.48	99.80	99.73
Exploits	99.41	99.39	99.51	99.63
Fuzzers	99.60	99.43	98.58	99.54
Reconnaissance	99.73	99.63	99.45	99.53
Analysis	99.41	99.80	99.55	99.56
Worms	99.73	99.48	98.60	99.65
Average	99.62	99.50	99.31	99.60

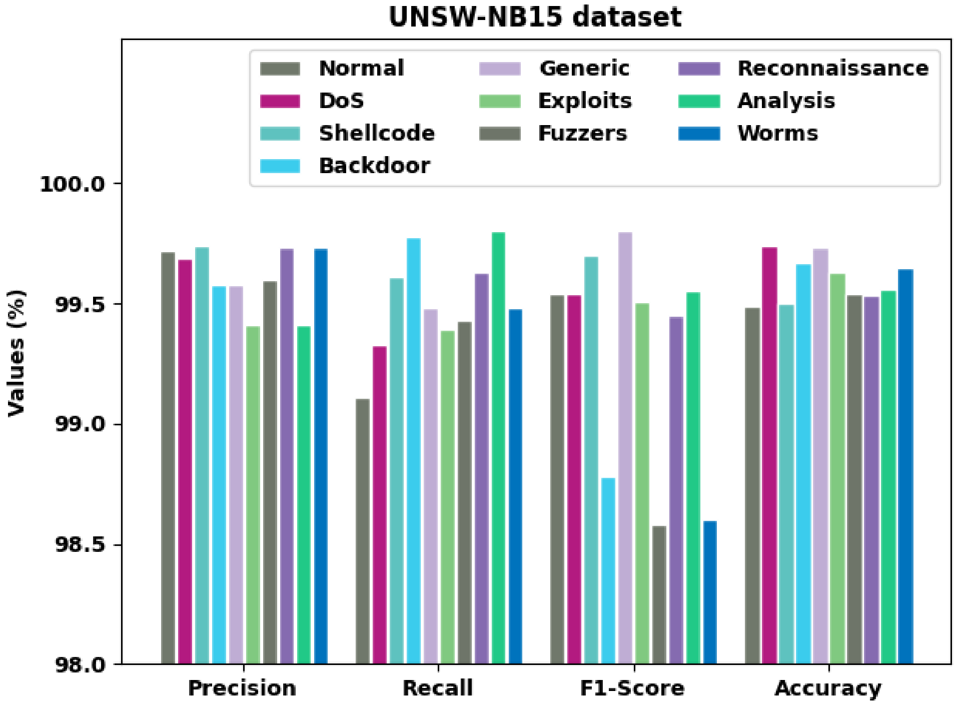

Figure 6. Result analysis of OMLIDS-PBIoT technique under UNSW-NB15 dataset.

The TA and VA gained by the OMLIDS-PBIoT technique on UNSW-NB15 dataset are shown in Figure 7. The experimental outcome implied that the OMLIDS-PBIoT technique has gained maximal values of TA and VA. Particularly, the VA seemed to be higher than TA.

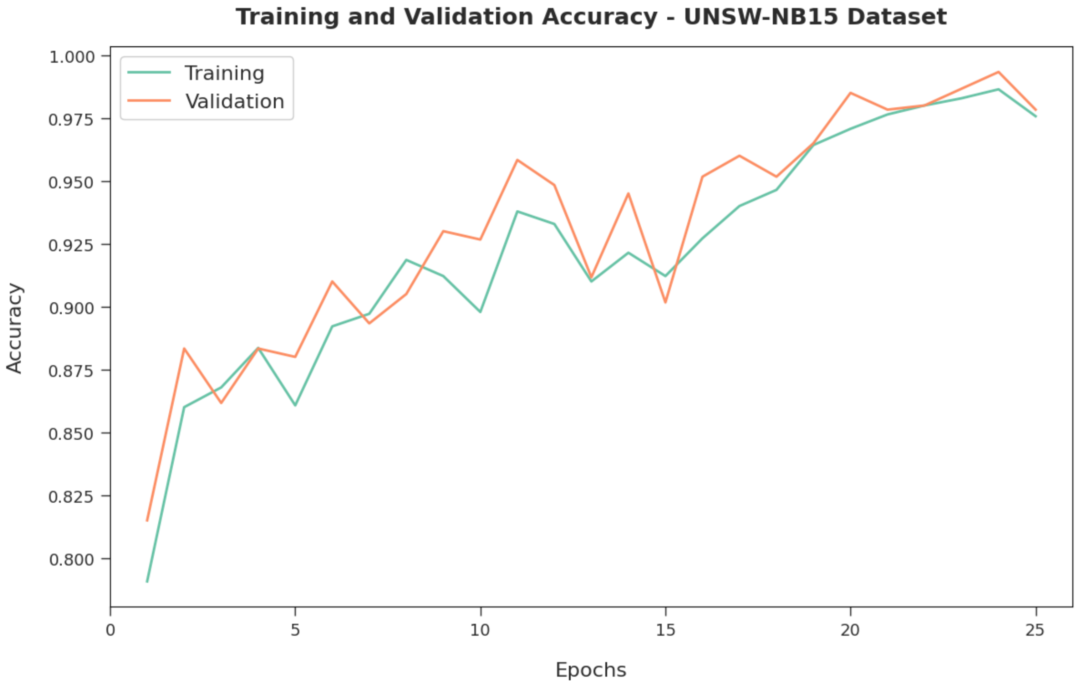

Figure 7. TA and VA analysis of OMLIDS-PBIoT technique under UNSW-NB15 dataset.

The TL and VL achieved by the OMLIDS-PBIoT techniques on UNSW-NB15 dataset are established in Figure 8. The experimental outcome inferred that the OMLIDS-PBIoT technique has accomplished the least values of TL and VL. Specifically, the VL seemed to be less than TL.

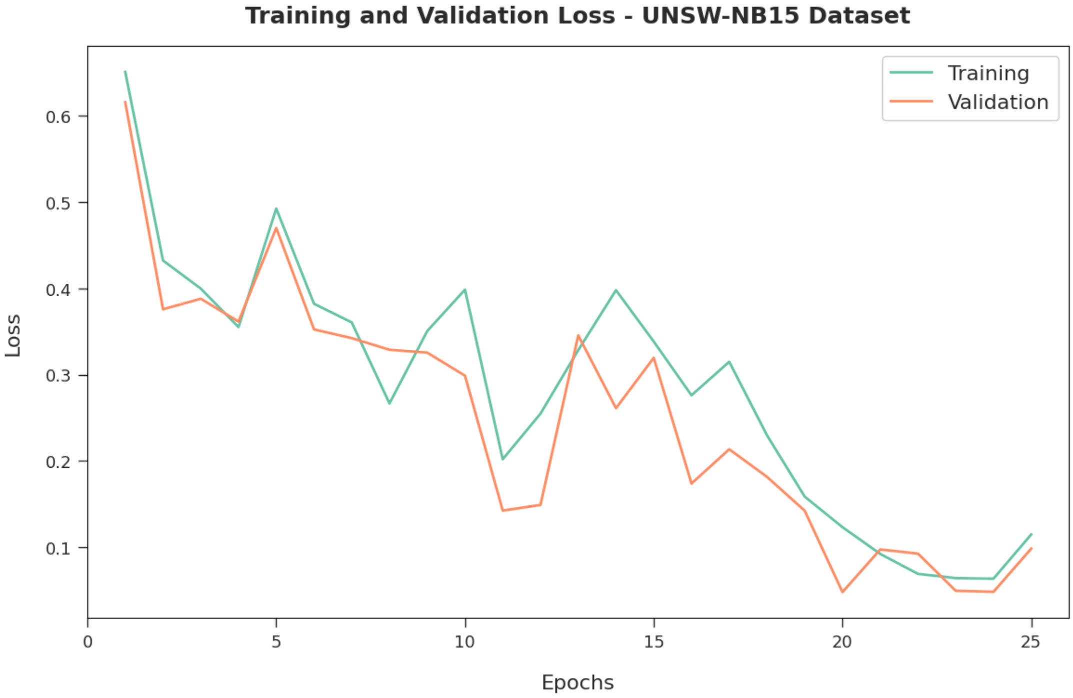

Figure 8. TL and VL analysis of OMLIDS-PBIoT technique under UNSW-NB15 dataset.

Table 4 portrays an extensive comparison study of the OMLIDS-PBIoT technique with recent models in terms of different measures [18]. Figure 9 offers a comparative $prec_n$ and $reca_l$ examination of the OMLIDS-PBIoT technique with existing models on test data. The results implied that the AKNN-IDS model has obtained lower $prec_n$ and $reca_l$ values of 0.9219 and 0.9376, respectively. Meanwhile, the DPCDBN-IDS and DT-IDS models have shown slightly improved values of $prec_n$ and $reca_l$. Moreover, the PTDSAE-IDS, AB-IDS, and RF-IDS models have obtained moderately closer values of $prec_n$ and $reca_l$. In line with this, the DBN-IDS, LSTM-IDS, and RNN-IDS models have accomplished reasonable $prec_n$ and $reca_l$ values. Though the AIMMFIDS model has shown near-optimal performance with $prec_n$ and $reca_l$ of 0.9946 and 0.9907, the presented OMLIDS-PBIoT technique has accomplished higher $prec_n$ and $reca_l$ of 0.9962 and 0.9950, respectively.

Table 4. Comparative analysis of OMLIDS-PBIoT method with existing approaches.

Methods	Precision	Recall	F1-Score	Accuracy
OMLIDS-PBIoT	0.9962	0.9950	0.9931	0.9960
AIMMFIDS	0.9946	0.9907	0.9898	0.9936
DBN-IDS	0.9867	0.9883	0.9897	0.9874
LSTM-IDS	0.9845	0.9857	0.9891	0.9863
RNN-IDS	0.9823	0.9868	0.9875	0.9852
PTDSAE-IDS	0.9791	0.9865	0.9860	0.9849
DPCDBN-IDS	0.9512	0.9499	0.9508	0.9498
AKNN-IDS	0.9219	0.9376	0.9292	0.9199
DT-IDS	0.9659	0.9284	0.9542	0.9365
AB-IDS	0.9742	0.9321	0.9568	0.9587
RF-IDS	0.9756	0.9384	0.9592	0.9598

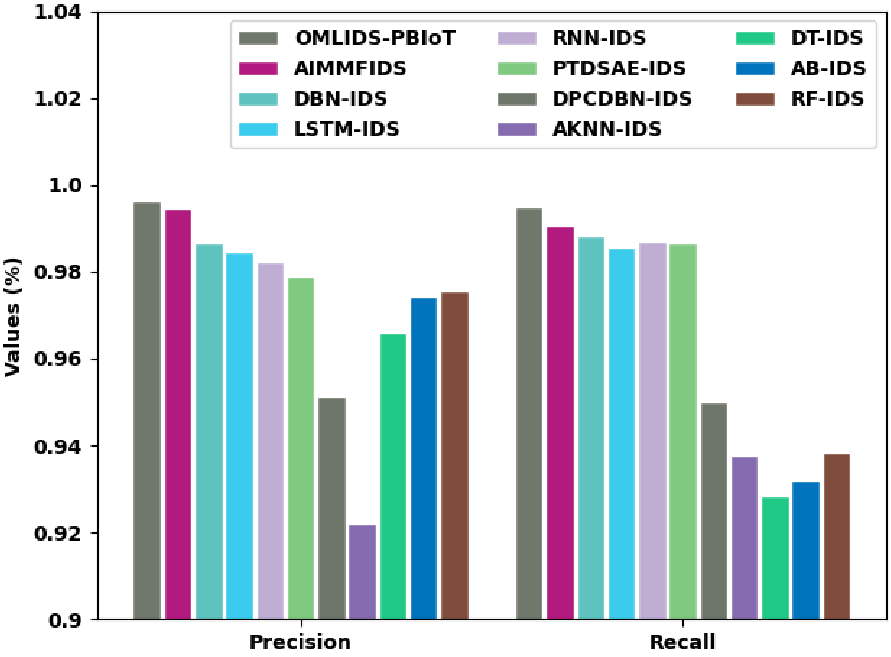

Figure 9. $Prec_n$ and $reca_l$ analysis of OMLIDS-PBIoT technique with existing approaches.

Figure 10 provides a comparative $F1_{score}$ and $accu_y$ inspection of the OMLIDS-PBIoT technique with existing techniques on test data. The results implied that the AKNN-IDS methodology has obtained lower $F1_{score}$ and $accu_y$ values of 0.9292 and 0.9199, respectively. In the meantime, the DPCDBN-IDS and DT-IDS methodologies have shown slightly improved values of $F1_{score}$ and $accu_y$. Next, the PTDSAE-IDS, AB-IDS, and RF-IDS algorithms have gained moderately closer values of $F1_{score}$ and $accu_y$. In accordance with this, the DBN-IDS, LSTM-IDS, and RNN-IDS methods have accomplished reasonable $F1_{score}$ and $accu_y$ values. Even though the AIMMFIDS system has displayed near-optimal performance with $F1_{score}$ and $accu_y$ of 0.9898 and 0.9936, the presented OMLIDS-PBIoT technique has accomplished higher $F1_{score}$ and $accu_y$ of 0.9931 and 0.9960, correspondingly.

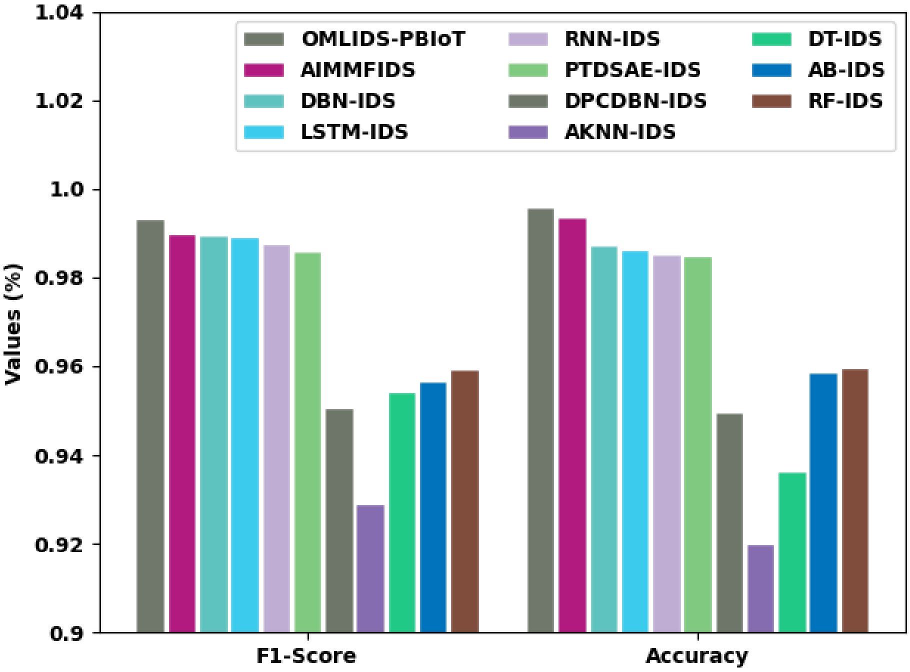

Figure 10. $F1_{score}$ and $accu_y$ analysis of OMLIDS-PBIoT technique with existing approaches.

Finally, a detailed computation time inspection of the OMLIDS-PBIoT technique with recent models takes place in Table 5. The results implied that the AKNN-IDS, DT-IDS, AB-IDS, and RF-IDS, models have shown the least performance with maximum values of TRT. Additionally, the LSTM-IDS, RNN-IDS, PTDSAE-IDS, and DPCDBN-IDS models have demonstrated moderately closer values of TRT and TST. In line with this, the DBN-IDS model has shown a reasonable TRT of 61 s. At the same time, the AIMMFIDS model has demonstrated considerable TRT of 55 s. However, the OMLIDS-PBIoT technique has accomplished superior performance with minimal TRT of 35 s.

At the same time, the AKNN-IDS, DT-IDS, AB-IDS, and RF-IDS, models have displayed the least performance with maximal values of TST. Following this, the LSTM-IDS, RNN-IDS, PTDSAE-IDS, and DPCDBN-IDS methodologies have illustrated moderately closer values of TST and TST. In compliance with this, the DBN-IDS method has shown reasonable TST of 34 s. Meanwhile, the AIMMFIDS model has demonstrated considerable TST of 25 s. However, the OMLIDS-PBIoT technique has accomplished superior performance with a minimum TST of 20 s. From the detailed outcomes and discussion, it is clear that the OMLIDS-PBIoT technique has resulted in more enhanced outcomes than other models.

Table 5. TRT and TST analysis of OMLIDS-PBIoT technique with existing algorithms.

Methods	Training Time (s)	Testing Time (s)
OMLIDS-PBIoT	35	20
AIMMFIDS	55	25
DBN-IDS	61	34
LSTM-IDS	68	34
RNN-IDS	69	35
PTDSAE-IDS	71	36
DPCDBN-IDS	73	37
AKNN-IDS	79	42
DT-IDS	80	45
AB-IDS	81	43
RF-IDS	83	45

5. Conclusions

In this study, an effective OMLIDS-PBIoT technique was advanced with the utilization of the BC and ML models for accomplishing security in the smart city environment. The presented OMLIDS-PBIoT technique employs different stages of subprocesses, namely pre-processing, GEO-FS, RVFL-based classification, and HBO-based parameter optimization. Moreover, BC technology is exploited for secure data transmission in the IoT-enabled smart city environment. The performance validation of the OMLIDS-PBIoT technique is performed using benchmark datasets, and the outcomes are inspected under numerous aspects. The experimental outcomes demonstrated the superiority of the OMLIDS-PBIoT technique over recent approaches. Thus, the OMLIDS-PBIoT technique could be exploited as a proficient tool to accomplish security in the smart city environment in sectors such as healthcare, smart buildings, transportation, etc. In upcoming days, the performance of the OMLIDS-PBIoT technique could be enhanced using hybrid DL classification methods.

Author Contributions: Conceptualization, A.A.-Q. and S.S.A.; methodology, F.A.; software, R.M.; validation, A.A.-Q. and M.O.; investigation, R.M.; resources, F.N.A.-W.; writing—original draft preparation, F.N.A.-W.; M.A.D. and M.A.-S.; writing—review and editing, N.N. and M.O.; visualization, M.O.; supervision, F.N.A.-W.; project administration, S.S.A.; funding acquisition, F.A. All authors have read and agreed to the published version of the manuscript.

Funding: 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 (71/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R77), 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: 22UQU4210118DSR19.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing 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.

References

1. Li, D.; Cai, Z.; Deng, L.; Yao, X.; Wang, H.H. Information security model of block chain based on intrusion sensing in the IoT environment. *Clust. Comput.* **2019**, *22*, 451–468. [CrossRef]
2. Rathee, G.; Iqbal, R.; Waqar, O.; Bashir, A.K. On the design and implementation of a blockchain enabled e-voting application within iot-oriented smart cities. *IEEE Access* **2021**, *9*, 34165–34176. [CrossRef]
3. Mukherjee, P.; Barik, R.K.; Pradhan, C. A comprehensive proposal for blockchain-oriented smart city. In *Security and Privacy Applications for Smart City Development*; Springer: Cham, Switzerland, 2021; pp. 55–87.
4. Banerjee, M.; Lee, J.; Choo, K.K.R. A blockchain future for internet of things security: A position paper. *Digit. Commun. Netw.* **2018**, *4*, 149–160. [CrossRef]
5. Singh, S.K.; Azzaoui, A.E.; Kim, T.W.; Pan, Y.; Park, J.H. DeepBlockScheme: A deep learning-based blockchain driven scheme for secure smart city. *Hum.-Cent. Comput. Inf. Sci* **2021**, *11*, 12.
6. Liang, C.; Shanmugam, B.; Azam, S.; Karim, A.; Islam, A.; Zamani, M.; Kavianpour, S.; Idris, N.B. Intrusion detection system for the internet of things based on blockchain and multi-agent systems. *Electronics* **2020**, *9*, 1120. [CrossRef]
7. Khan, M.A.; Abbas, S.; Rehman, A.; Saeed, Y.; Zeb, A.; Uddin, M.I.; Nasser, N.; Ali, A. A machine learning approach for blockchain-based smart home networks security. *IEEE Netw.* **2020**, *35*, 223–229. [CrossRef]
8. Shen, M.; Tang, X.; Zhu, L.; Du, X.; Guizani, M. Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities. *IEEE Internet Things J.* **2019**, *6*, 7702–7712. [CrossRef]
9. Singh, S.; Sharma, P.K.; Yoon, B.; Shojafar, M.; Cho, G.H.; Ra, I.H. Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustain. Cities Soc.* **2020**, *63*, 102364. [CrossRef]
10. Kumar, P.; Gupta, G.P.; Tripathi, R. TP2SF: A Trustworthy Privacy-Preserving Secured Framework for sustainable smart cities by leveraging blockchain and machine learning. *J. Syst. Archit.* **2021**, *115*, 101954. [CrossRef]
11. Kumar, P.; Kumar, R.; Srivastava, G.; Gupta, G.P.; Tripathi, R.; Gadekallu, T.R.; Xiong, N.N. PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2326–2341. [CrossRef]
12. Meng, W.; Li, W.; Tug, S.; Tan, J. Towards blockchain-enabled single character frequency-based exclusive signature matching in IoT-assisted smart cities. *J. Parallel Distrib. Comput.* **2020**, *144*, 268–277. [CrossRef]
13. Bediya, A.K.; Kumar, R. A novel intrusion detection system for internet of things network security. *J. Inf. Technol. Res.* **2021**, *14*, 20–37. [CrossRef]
14. Rathore, S.; Kwon, B.W.; Park, J.H. BlockSecIoTNet: Blockchain-based decentralized security architecture for IoT network. *J. Netw. Comput. Appl.* **2019**, *143*, 167–177. [CrossRef]
15. Botello, J.V.; Mesa, A.P.; Rodríguez, F.A.; Díaz-López, D.; Nespoli, P.; Mármol, F.G. BlockSIEM: Protecting smart city services through a blockchain-based and distributed SIEM. *Sensors* **2020**, *20*, 4636. [CrossRef] [PubMed]
16. Aujla, G.S.; Singh, M.; Bose, A.; Kumar, N.; Han, G.; Buyya, R. Blocksdn: Blockchain-as-a-service for software defined networking in smart city applications. *IEEE Netw.* **2020**, *34*, 83–91. [CrossRef]
17. Peneti, S.; Sunil Kumar, M.; Kallam, S.; Patan, R.; Bhaskar, V.; Ramachandran, M. BDN-GWMNN: Internet of things (IoT) enabled secure smart city applications. *Wirel. Pers. Commun.* **2021**, *119*, 2469–2485. [CrossRef]
18. 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]
19. Al-Wesabi, F.N.; Obayya, M.; Hamza, M.A.; Alzahrani, J.S.; Gupta, D.; Kumar, S. Energy Aware Resource Optimization using Unified Metaheuristic Optimization Algorithm Allocation for Cloud Computing Environment. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100686. [CrossRef]
20. Bach, L.M.; Mihaljevic, B.; Zagar, M. Comparative analysis of blockchain consensus algorithms. In Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 1545–1550.
21. Amor, N.; Noman, M.T.; Petru, M.; Sebastian, N. Comfort evaluation of ZnO coated fabrics by artificial neural network assisted with golden eagle optimizer model. *Sci. Rep.* **2022**, *12*, 6350. [CrossRef]
22. Hussien, A.G.; Houssein, E.H.; Hassanien, A.E. A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection. In Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 5–7 December 2017; pp. 166–172.
23. Shi, Q.; Katuwal, R.; Suganthan, P.N.; Tanveer, M. Random vector functional link neural network based ensemble deep learning. *Pattern Recognit.* **2021**, *117*, 107978. [CrossRef]
24. Askari, Q.; Saeed, M.; Younas, I. Heap-based optimizer inspired by corporate rank hierarchy for global optimization. *Expert Syst. Appl.* **2020**, *161*, 113702. [CrossRef]