Computer Systems
Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.037545
Article

Blockchain Assisted Optimal Machine Learning Based Cyberattack Detection and Classification Scheme

Manal Abdullah Alohali[1], Muna Elsadig[1], Fahd N. Al-Wesabi[2,*], Mesfer Al Duhayyim[3], Anwer Mustafa Hilal[4] and Abdelwahed Motwakel[4]

[1]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[2]Department of Computer Science, College of Science & Art at Mahayil, King Khalid University, Abha, 62529, Saudi Arabia
[3]Department of Computer Science, College of Sciences and Humanities-Aflaj, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[4]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University Al-Kharj, 16278, Saudi Arabia
*Corresponding Author: Fahd N. Al-Wesabi. Email: falwesabi@kku.edu.sa
Received: 08 November 2022; Accepted: 02 February 2023

Abstract: With recent advancements in information and communication technology, a huge volume of corporate and sensitive user data was shared consistently across the network, making it vulnerable to an attack that may be brought some factors under risk: data availability, confidentiality, and integrity. Intrusion Detection Systems (IDS) were mostly exploited in various networks to help promptly recognize intrusions. Nowadays, blockchain (BC) technology has received much more interest as a means to share data without needing a trusted third person. Therefore, this study designs a new Blockchain Assisted Optimal Machine Learning based Cyberattack Detection and Classification (BAOML-CADC) technique. In the BAOML-CADC technique, the major focus lies in identifying cyberattacks. To do so, the presented BAOML-CADC technique applies a thermal equilibrium algorithm-based feature selection (TEA-FS) method for the optimal choice of features. The BAOML-CADC technique uses an extreme learning machine (ELM) model for cyberattack recognition. In addition, a BC-based integrity verification technique is developed to defend against the misrouting attack, showing the innovation of the work. The experimental validation of BAOML-CADC algorithm is tested on a benchmark cyberattack dataset. The obtained values implied the improved performance of the BAOML-CADC algorithm over other techniques.

Keywords: Cyberattack; machine learning; blockchain; thermal equilibrium algorithm; feature selection

1 Introduction

Data analysis methods were used extensively in the cyber security field, and the current diffusion of advanced machine learning (ML) methods has permitted to precisely detect cyber-attacks and identify threats [1], both in post-incident analysis and real-time. Both unsupervised and supervised ML techniques were successfully used for supporting prevention systems and intrusion detection (ID), along with identifying security breaches and system misuses [2,3]. The scenarios of interest were generally characterized by a continuous data stream (like application-level or packet-level) that summarizes the conduct of the underlying system or network [4]. The ML methods' role is either in detecting familiar attacks, anomalous behaviour (unsupervised method) (supervised method), or Anomaly-related approaches that fit normal system functioning status, identifying and isolating anomalies as unexpected behavioural deviations. Therefore, anomaly detection methods are attractive for their capability to identify zero-day attacks, which are attacks using unknown vulnerabilities [5,6]. Fig. 1 represents the overview of the blockchain (BC)-assisted cyberattack detection.

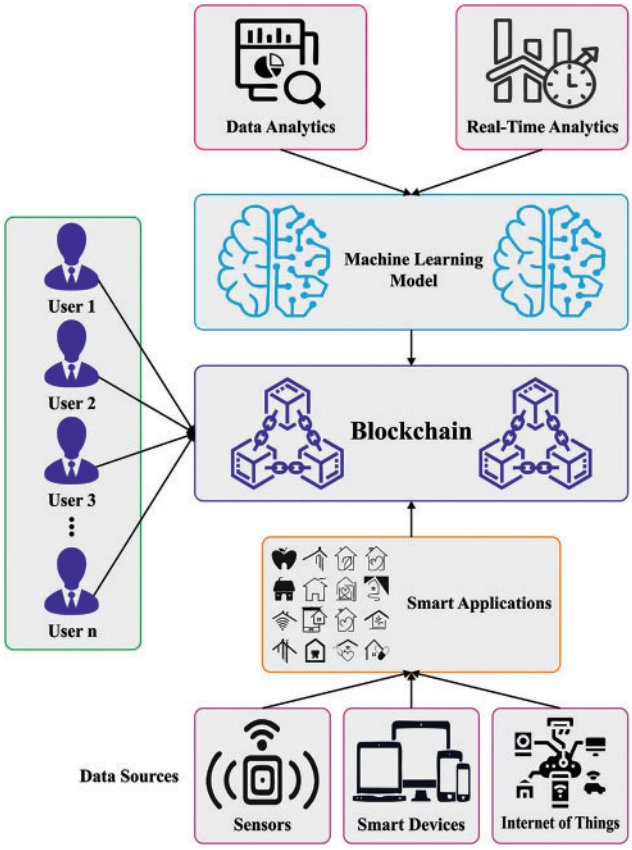

Figure 1: Blockchain-assisted cyberattack detection

In recent times, deep learning (DL) is becoming a hopeful analytic pattern because of its excellent function in examining huge volumes of data [7]. Dissimilar to the conventional ML method, DL supported effective feature engineering by handling automatic and reliable feature representation and extraction [8,9]. Since a strong analytic device, DL delivered existing latency and accuracy than the conventional ML method, and it is positioned for examining huge amounts of data in the 5G-based Internet of Things (IoT) [10]. This DL disposition supported forecasting future events, recognising

assaults, and offering important data for placement and content caching in dynamic cases of 5G-assisted IoT [11].

Since an evolving technology, BC has become a hopeful choice for managing privacy and security in next-generation transmission structures [12,13]. It constitutes a peer-to-peer (P2P) transactions platform where the data is exchanged, recorded, and authenticated in a decentralized way for delivering verification data and security independent from centralized authorities [14]. The significant features of BC involving anonymity, decentralization, and security can apply secure data transactions and overcome centralized server dependence to support security in 5G-based IoT. Additionally, the inimitable property of dispersed data storage, smart contracts, and asset tracking make BC technology required for 5G-based IoT [15].

This study designs a new Blockchain Assisted Optimal Machine Learning based Cyberattack Detection and Classification (BAOML-CADC) technique. In the BAOML-CADC technique, the major focus lies in identifying cyberattacks. To do so, the presented BAOML-CADC technique applies a thermal equilibrium algorithm-based feature selection (TEA-FS) method for the optimal choice of features. The BAOML-CADC technique uses the extreme learning machine (ELM) model for cyberattack recognition. In addition, a BC-based integrity verification technique is developed to protect against the misrouting attack. The experimental validation of BAOML-CADC algorithm is tested on a benchmark cyberattack dataset.

2 Related Works

In [16], the authors identify false data injection attack (FDIA) in the microgrid mechanism, Hilbert-Huang transforms technique, along with BC-related ledger technology, was employed to scale up the security in smart direct current (DC)-microgrids by evaluating the voltage and current signals in controller and smart sensor nodes by deriving signal information. Ajayi et al. [17] devised a BC-related solution that guarantees the consistency and integrity of attack features shared in a cooperative ID mechanism. The modelled structure attains by preventing and identifying compromised intrusion detection system (IDS) nodes and fake feature injection. It even enables scalable attack features to exchange amongst IDS nodes, assures heterogeneous IDS node contribution, and is powerful to public IDS node leaving and joining the networks.

Kumar et al. [18] presented a new BC system for secured clinical data management that minimizes the computational and communicational overhead cost than the lightweight BC architecture and the prevailing bitcoin network. Kumar et al. [19] present a Privacy-Preserving and Secure Framework (PPSF) for IoT-driven smart city. This devised method is dependent upon 2 main systems: an ID system and a two-level privacy system. Firstly, in a 2-level privacy method, a BC was devised to transfer information of IoT securely, and the PCA method can be adapted to convert raw IoT data into a new structure. A Gradient Boosting Anomaly Detector (GBAD) was implemented in the ID method to evaluate and train the devised two-level privacy method related to IoT network datasets, such as BoT-IoT and ToN-IoT.

In [20], modelled a new process based on DL and Hilbert-Huang Transform for cyberattack recognition in DC-MGs along with identifying the assaults in distributed generation (DG) sensors and units. Then, an innovative elective group DL approach and Krill Herd Optimization (KHO) was devised. Then, Hilbert-Huang Transform was employed to extract the signals feature. Then such attributes were implemented as multiple deep input basis methods were constituted for capturing sentient traits automatically from raw fluctuation signals. Liang et al. [21] modelled a novel, distributed BC-related security system to improve modern power systems' self-defensive ability against cyberattacks. The

authors discussed how BC technology improves the power grid's security and robustness by leveraging meters as nodes in dispersed networks and encapsulating meter measurements as blocks.

3 The Proposed Model

In this study, we have developed a new BAOML-CADC technique for Cyberattack Detection and Classification process. In the BAOML-CADC technique, the major focus lies in identifying cyberattacks. Initially, the presented BAOML-CADC technique applied the TEA-FS approach for the optimal choice of features. For cyberattack recognition, the BAOML-CADC technique used the ELM model. In addition, a BC-based integrity verification technique is developed to defend against the misrouting attack.

3.1 Algorithmic Steps of TEA-FS Technique

Primarily, the presented BAOML-CADC technique applied the TEA-FS approach for the optimal choice of features. TEA was simulated by thermodynamic phenomena and is classified as an evolutionary technique [22]. During this approach, the coordinates of all the systems can be transformed into volume or temperature, which are thermodynamic parameters. The study aims to search for an optimum solution based on the problem variable. In genetic algorithm (GA), the group of variables should be enhanced, often known as the chromosome. However, the term can be replaced by the thermodynamic system. In the N_{var}-dimension problems, every system is a dimensional array determined using the following expression:

$$x = [T, V_1, V_2, \ldots, V_{N_{var}-1}] \tag{1}$$

In Eq. (1), T denotes the temperature, and V indicates the volume that might have multiple parameters. Thus, for ease of use, the problem became 2D problems by describing a new parameter $\forall$ termed "overall volume":

$$\forall = \frac{\Sigma_{i=1}^{Nvar-1} V_i}{N_{var} - 1} \tag{2}$$

Here, the thermodynamic state of all the systems is determined by the following expression:

$$x = [T, \forall] \tag{3}$$

The cost of every system can be defined using *the f* cost function.

$$cost = f(x) = f(T, \forall) \tag{4}$$

At first, the initial population can be generated by the number of N_{sys}. Based on upper and lower boundaries, the system is arbitrarily scattered in the function domain.

In coupling the Thermodynamic System, firstly, every system is combined to the closest system in the function domain. Next, the combined system exchanges heat or implements work together to reach equilibrium progressively. Eventually, the temperature gradient is reduced, the scope of the optimization technique is well analyzed, as well as the optimum value is found. Afterwards, the system is coupled, and heat exchange and work are executed, leading to changes in the volume and temperature of the system. The volume and temperature are distinct. However, the pressure can be similar. Moreover, the piston in the two systems moves freely. Hence every system implements

either negative or positive work. In every thermodynamic procedure, the pressure in two systems is equivalent:

$$\Delta E = Q - W \tag{5}$$

In Eq. (5), Q represents the amount of heat transferred, and W denotes the work done by every system. Because of the conservation law of energy, the heat absorbed from the system is equivalent to the heat lost by the others ($Q_1 = -Q_2$):

$$\Delta E_1 + W_1 = -(\Delta E_2 + W_2) \tag{6}$$

The energy change for an ideal gas is evaluated as $\Delta E = mc_v \Delta T$, and the thermodynamic work by $W = P\Delta\forall$. By replacing the term into Eq. (6) and divide with P, Eq. (7) can be attained:

$$\frac{m_1 c_v}{P}\left(T_{eq} - T_1\right) + \left(\forall_{eq} - \forall_1\right) = -\frac{m_2 c_v}{P}\left(T_{eq} - T_2\right) - \left(\forall_{eq} - \forall_2\right) \tag{7}$$

In Eq. (7), the subscript eq denotes the thermodynamic state at equilibrium. Consider that $m_1 = m_2$ and for easiness $\alpha = (m_1 c_v)/P = (m_2 c_v)/P = 1$, Eq. (7) is formulated by:

$$\left(T_{eq} - T_1\right) + \left(\forall_{eq} - \forall_1\right) = \left(T_2 - T_{eq}\right) + \left(\forall_2 - \forall_{eq}\right) \tag{8}$$

Alternatively, the ideal gas law is represented as:

$$P\forall = nRT \tag{9}$$

The overall molar mass at the equilibrium state is equivalent to the sum of molar mass according to the conservation law of mass:

$$n_1 + n_2 = 2n_{eq} \tag{10}$$

By substituting Eqs. (9) with (10),

$$\frac{\forall_1}{T_1} + \frac{\forall_2}{T_2} = 2\frac{\forall_{eq}}{T_{eq}} \tag{11}$$

By integrating Eqs. (8) and (11), the following expression can attain T_{eq}:

$$T_{eq} = \frac{T_1^2 T_2 + T_1 T_2^2 + T_1 T_2 \forall_1 + T_1 T_2 \forall_2}{T_1 T_2 + T_2 \forall_1 + T_1 \forall_2} \tag{12}$$

In addition, $\forall_{eq}$ is derived by:

$$\forall_{eq} = \frac{T_1^2 \forall_2 + T_1 T_2 \forall_1 + T_1 T_2 \forall_2 + T_1 \forall_1 \forall_2 + T_1 \forall_2^2 + T_2^2 \forall_1 + T_2 \forall_1^2 + T_2 \forall_1 \forall_2}{2(T_1 T_2 + T_2 \forall_1 + T_1 \forall_2)} \tag{13}$$

By using the ideal gas law and the first law of thermodynamics the equilibrium state of hypothetical system is evaluated. If every system attains equilibrium instantly, the state of two systems would be equivalent, and they won't exchange energy. In such cases, the function's domain won't be explored further, and the optimum solution cannot be attained. To avoid that, the subsequent relationship is suggested for overall volume and new temperature:

$$T_{i,new} = \frac{T_i + T_{eq}}{2} \tag{14}$$

$$\forall_{i,new} = \frac{\forall_i + \forall_{eq}}{2} \tag{15}$$

Now the term i signifies the system number. When the $\forall$ values are encompassed of multiple parameters like the case in a three or higher-dimension problem, V_i is attained by:

$$V_{i,new} = V_i + \forall_{i,new} \tag{16}$$

The two systems grow nearer to equilibrium at every phase, which causes the whole domain to be examined progressively and the optimum point to be found. The relationships between entropy and cost function are determined by:

$$cost = \frac{1}{entropy} \equiv \frac{temperature}{heat} \tag{17}$$

Thus, the study aims to diminish the cost function leads to maximizing the entropy of the system (S):

$$S_{new} - S_{eq} \leq 0 \tag{18}$$

By transforming entropy into a cost function, the condition is transformed to:

$$f\left(T_{new'} \forall_{new}\right) - f\left(T_{eq'} \forall_{eq}\right) \geq 0 \tag{19}$$

The abovementioned conditions cause the algorithm to move towards the optimization function. When the criterion isn't satisfied, a counter is determined for counting the number of failed attempts, then the thermodynamic condition is upgraded:

$$T_{i,new} = \frac{T_i + T_{eq}}{2j_{new}} \tag{20}$$

$$\forall_{i,new} = \frac{\forall_i + \forall_{eq}}{2j_{new}} \tag{21}$$

Whereas $i_{new} = j + 1$. Because of the abovementioned relationship, in these cases, the system move towards the balance state at the lower speed than Eqs. (14) and (15) for satisfying the abovementioned conditions (Eq. (19)). This process is same as inspecting the second law of thermodynamics and preventing blind search of the function domain.

Lastly, many coupled systems reached equilibrium and accumulated at any point, excluding a few that might be trapped. This technique has the benefit that only some systems have an equivalent state. Hence the cost of every system differs from the thermodynamic equilibrium method.

The fitness function (FF) employed in the TEA-FS approach was modelled to maintain a balance amongst the classification accuracy (maximum), and the number of selected features in each solution (minimum) acquired through the selected attributes, Eq. (22) signifies the FF for evaluating solutions.

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|R|}{|C|} \tag{22}$$

Here $|R|$ signifies the cardinality of the selected subset, and $|C|$ is the total features in the database; $\gamma r(D)$ designates the classification error rate of a given classifier. β and α were 2 parameters concerning the significance of subset length and classification quality. $\in [1,0]$ and $\beta = 1 - \alpha$.

3.2 Cyberattack Detection Using ELM Model

To detect cyberattacks, the BAOML-CADC technique used the ELM model. Huang et al. developed an ELM to increase overall performance and prevent time-consuming backward iterative

training [23]. ELM is a Single hidden Layer Feedforward Network (SLFN) with random weight and biases amongst the hidden and input layers. ELM is commonly applied in different areas because of its quick training speed and easier realization. Fig. 2 depicts the framework of ELM. Data of N instances are provided $(i_p, 0_p)(p = 1, 2, \ldots, N)$, where input matrix $\mathrm{I} = [i_1, i_2, \ldots, i_N]^T$ and output matrices $0 = [0_1, 0_2, \ldots, 0_N]$, whereas $i_p = [i_{p1}, i_{p2}, \ldots, i_{pn}]^T$, $o_p = [0_{p1}, 0_{p2}, \ldots, 0_{pm}]^T$ and n and m characterize the input and output dimensions, correspondingly. Initially, the ELM arbitrarily allows the bias $D = [d_1, d_2, \ldots, d_q] \in R^{N \times L}$ and the weight $E = [E_1, E_2, \ldots, E_q] \in R^{n \times L}$ that links the hidden and input layers. In contrast, L signifies the hidden node count with activation function $h(x)$, $E_q = [E_{q1}, E_{q2}, \ldots, E_{qn}]^T$ represents the weight connecting vector which connects n input node with q-th hidden nodes.

$$M = h\,(IE + D) \tag{23}$$

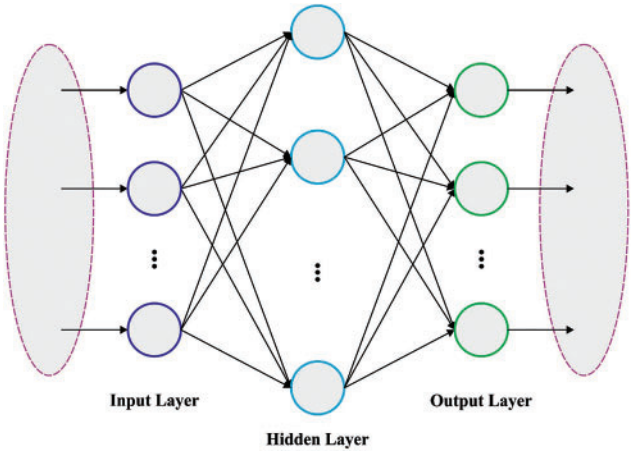

Figure 2: ELM structure

The following formula can be expressed as output matrix O.

$$MF = O \tag{24}$$

Whereas $F = [F_1, F_2, \ldots, F_L]^T$ signifies an output weight matrix that connects output to the hidden layers. The output weight matrices F are calculated using the least-square method.

$$F = M^{+}O \tag{25}$$

In Eq. (25), M^{+} signifies the Moore-Penrose (MP) generalized inverse of M matrix. When $M^T M$ is non-singular, then $M^{+} = (M^T M)^{-1} M^T$. In such cases, L is lesser than N. If MM^T is non-singular, then $M^{+} = M^T (MM^T)^{-1}$. In such cases, L is higher than N.

Algorithm 1 ELM

Input: h(x): activation function, I: L: the number of nodes in the hidden layer, input matrix of N samples, O: output matrix of N instances
Output: $f(x) = h(IE + D)F$
1 Initialize E weight and D bias randomly
2 Evaluate matrix $M = h(IE + D)$
3 Evaluate matrix F based on Eq. (25)

3.3 BC-Based Integrity Verification

Finally, a BC-based integrity verification technique is developed to defend against the misrouting attack. Over the last few decades, BC technology has provided privacy and security in different networks [24]. Despite the remarkable features of BC, it is still susceptible to fraudulent activity. The malicious entity might carry out fraudulent and invalid transactions using different technologies, namely double-spending attacks. BC can be fused with ML algorithms to resolve these problems in this work. The dataset of bitcoin transactions has been utilized in the fundamental process, and the presented ML approach has been trained on the database. The pattern of transactions saved in the dataset can be analyzed for added applications. Simultaneously, the transaction can be done on the Ethereum network. The pattern of this transaction is the same as that of bitcoin transactions saved in the bitcoin transaction data. The transaction pattern is analysed and compared to the bitcoin transaction patterns. Once the pattern matches these transactions, the new transaction is categorized as malicious or legitimate. A double-spending attack has been performed in the fundamental process for additional testing of the robustness of the model.

The BC is a building block of the reliable authentication system. The major objective is to provide a solution where each flow made from the controller is stored in an immutable and verifiable dataset. The BC involves a sequence of blocks connected through hash value. In the BC system, the user contains a pair of keys, such as a public key represents the irreplaceable address and a private key to sign the BC transaction. The client sign a transaction using the private key and transfers it to another one in the network for authentication. If the transmission block gets confirmed, then it is added to BC. Once stored, the data in the presented block can't be modified without changing each subsequent block. Also, the data is presented in the host simultaneously. Thus, the modification is rejected by the peer host. Now, a private BC is presented in contradiction to public BC. The private BC determines who should participate in the network and represented action in addition to permission allocated to the identifiable applicant. Hence, the requirement has been limited for a consensus mechanism, including Proof of Work.

4 Results and Discussion

In this section, the performance validation of the BAOML-CADC method can be performed through a benchmark dataset [25], which has 1000 distinct classes of events. The dataset comprises multiclass (Attack, No event, and Natural) and binary (Attack and Natural) labels.

Table 1 demonstrates the experimental results offered by the BAOML-CADC method on the binary class dataset. In Fig. 3, a detailed $sens_y$ and $spec_y$ outcomes of the BAOML-CADC model on the binary dataset are provided. The outcomes implied the improved values of $sens_y$ and $spec_y$ under all classes. For instance, with SD-1 class, the BAOML-CADC model has provided $sens_y$ and $spec_y$ of 98.44% and 99.99%, respectively. On the other hand, with SD-2 class, the BAOML-CADC approach has offered $sens_y$ and $spec_y$ of 98.89% and 99.97% singly. Meanwhile, with the SD-10 class, the BAOML-CADC method has correspondingly rendered $sens_y$ and $spec_y$ of 97.56% and 99.50%. Moreover, with the SD-15 class, the BAOML-CADC approach has presented $sens_y$ and $spec_y$ of 96.83% and 99.87%, correspondingly.

Table 1: Results analysis of BAOML-CADC approach under binary class database

Binary dataset	Acc_y	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}
SD-1	99.79	96.46	98.44	99.99	97.39
SD-2	98.22	96.76	98.89	99.97	96.81
SD-3	99.36	99.32	97.88	98.57	98.69
SD-4	99.43	97.02	98.20	99.53	98.81
SD-5	98.73	97.53	97.62	98.69	96.96
SD-6	98.83	97.19	97.79	99.12	98.22
SD-7	99.49	97.79	96.75	99.87	97.03
SD-8	97.92	97.61	97.84	98.86	97.11
SD-9	99.77	96.84	96.98	99.52	99.02
SD-10	99.22	97.36	97.56	99.50	98.18
SD-11	99.07	98.50	97.50	99.11	98.51
SD-12	99.48	99.39	97.77	98.73	98.25
SD-13	99.21	97.70	97.42	99.18	98.31
SD-14	99.25	97.97	97.11	98.89	98.69
SD-15	99.29	98.85	96.83	99.87	97.56
Average values	**99.14**	**97.75**	**97.64**	**99.29**	**97.97**

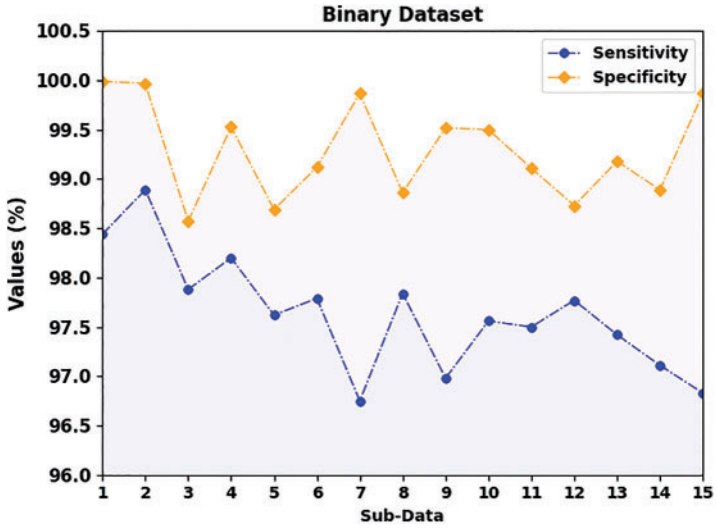

Figure 3: $Sens_y$ and $spec_y$ analysis of BAOML-CADC system under binary class database

Fig. 4 presents a comprehensive $prec_n$ and F_{score} outcomes of the BAOML-CADC method on the binary dataset. The outcomes exhibited improved values of $prec_n$ and F_{score} under all classes. For example, with SD-1 class, the BAOML-CADC algorithm has provided $prec_n$ and F_{score} of 96.46% and 97.39%, correspondingly. In contrast, with SD-2 class, the BAOML-CADC technique has offered $prec_n$ and F_{score} of 96.76% and 96.81%, correspondingly. In the meantime, with the SD-10 class, the BAOML-CADC model has provided $prec_n$ and F_{score} of 97.36% and 98.18% correspondingly. Likewise, with the

SD-15 class, the BAOML-CADC methodology has correspondingly provided $prec_n$ and F_{score} of 98.85% and 97.56%.

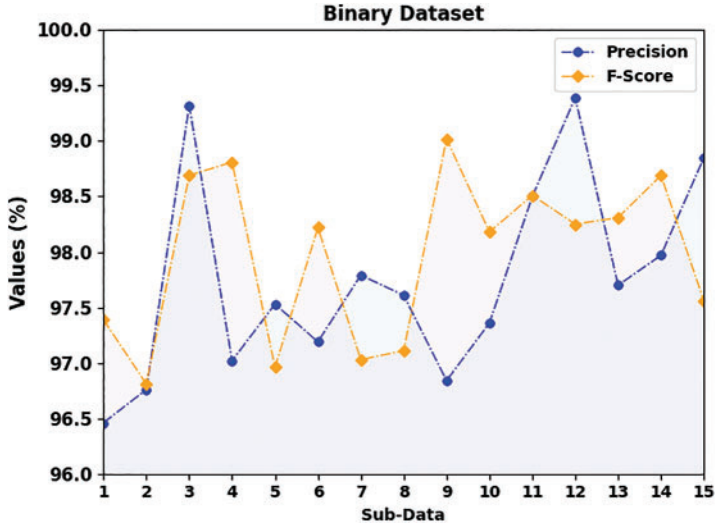

Figure 4: $Prec_n$ and F_{score} analysis of BAOML-CADC system under binary class database

In Fig. 5, detailed $accu_y$ outcomes of the BAOML-CADC approach on the binary dataset are provided. The outcomes displayed the improved value of $accu_y$ under all classes. For example, with SD-1 class, the BAOML-CADC method has rendered *an* $accu_y$ of 99.79%. On the other hand, with SD-2 class, the BAOML-CADC approach has provided *an* $accu_y$ of 98.22%. In the meantime, with the SD-10 class, the BAOML-CADC technique has provided an $accu_y$ of 99.22%. Additionally, with the SD-15 class, the BAOML-CADC methodology has provided an $accu_y$ of 99.29%.

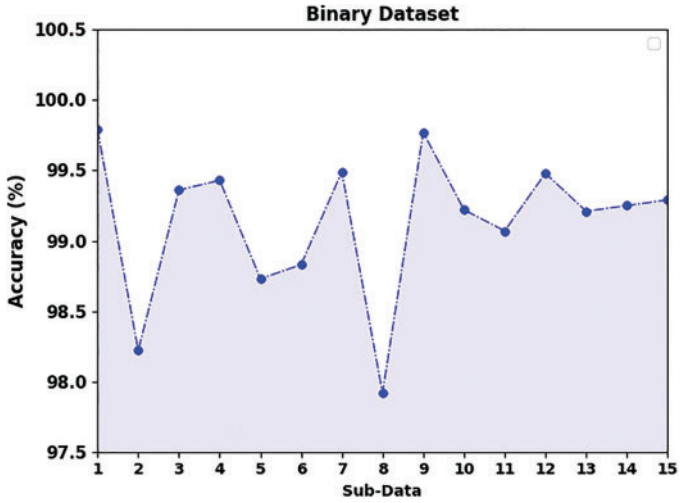

Figure 5: $Accu_y$ analysis of BAOML-CADC system under binary class database

In Table 2, the experimental results offered by the BAOML-CADC model on the Multiclass class dataset are represented. In Fig. 6, a brief $sens_y$ and $spec_y$ outcomes of the BAOML-CADC methodology on the Multiclass dataset are offered. The figure exhibited the improved values of $sens_y$

and $spec_y$ under all classes. For example, with SD-1 class, the BAOML-CADC algorithm has provided $sens_y$ and $spec_y$ of 93.30% and 95.07%, correspondingly. Instead, with SD-2 class, the BAOML-CADC model has presented $sens_y$ and $spec_y$ of 91.04% and 94.03%, correspondingly. Meanwhile, with the SD-10 class, the BAOML-CADC method has provided $sens_y$ and $spec_y$ of 93.05% and 95.57%, correspondingly. Also, with the SD-15 class, the BAOML-CADC model has provided $sens_y$ and $spec_y$ of 90.99% and 95.36%, correspondingly.

Table 2: Results analysis of proposed BAOML-CADC method on multiclass dataset

Multiclass dataset	Acc_y	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}
SD-1	93.64	79.32	93.30	95.07	82.16
SD-2	94.67	82.35	91.04	94.03	80.83
SD-3	94.32	83.45	89.06	92.88	83.90
SD-4	94.05	81.66	91.05	94.19	82.43
SD-5	92.91	80.70	90.64	92.62	82.21
SD-6	92.81	84.05	93.57	94.73	80.81
SD-7	92.45	80.33	92.76	94.00	85.14
SD-8	93.70	83.28	90.66	95.24	81.29
SD-9	92.47	81.09	92.09	95.31	80.59
SD-10	94.05	83.39	93.05	95.57	80.16
SD-11	93.39	80.37	91.91	93.97	82.96
SD-12	93.09	82.16	89.71	92.60	82.06
SD-13	93.11	83.51	90.53	94.27	80.25
SD-14	91.16	85.02	89.86	94.89	80.68
SD-15	91.77	79.36	90.99	95.36	83.77
Average values	**93.17**	**82.00**	**91.35**	**94.32**	**81.95**

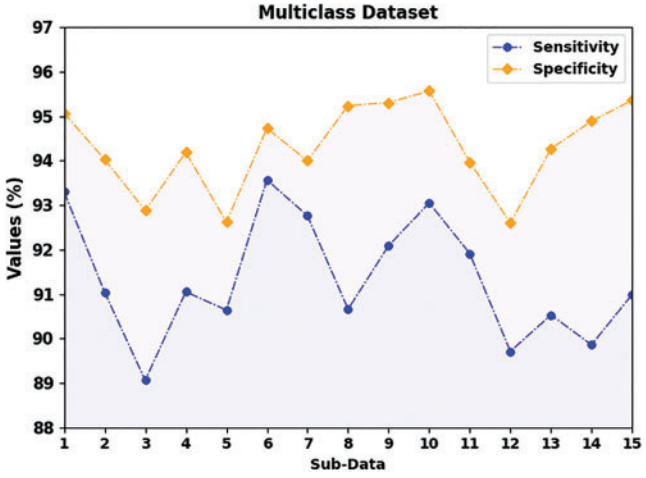

Figure 6: $Sens_y$ and $spec_y$ analysis of BAOML-CADC system under multiclass database

In Fig. 7, detailed $prec_n$ and F_{score} outcomes of the BAOML-CADC method on the Multiclass dataset are provided. The figure exhibited the improved values of $prec_n$ and F_{score} under all classes. For instance, with SD-1 class, the BAOML-CADC model has provided $prec_n$ and F_{score} of 79.32% and 82.16%, correspondingly. Otherwise, with SD-2 class, the BAOML-CADC technique has presented $prec_n$ and F_{score} of 82.35% and 80.83%, correspondingly. In the meantime, with the SD-10 class, the BAOML-CADC model has provided $prec_n$ and F_{score} of 83.39% and 80.16% correspondingly. Besides, with the SD-15 class, the BAOML-CADC method has presented $prec_n$ and F_{score} of 79.36% and 83.77%, correspondingly.

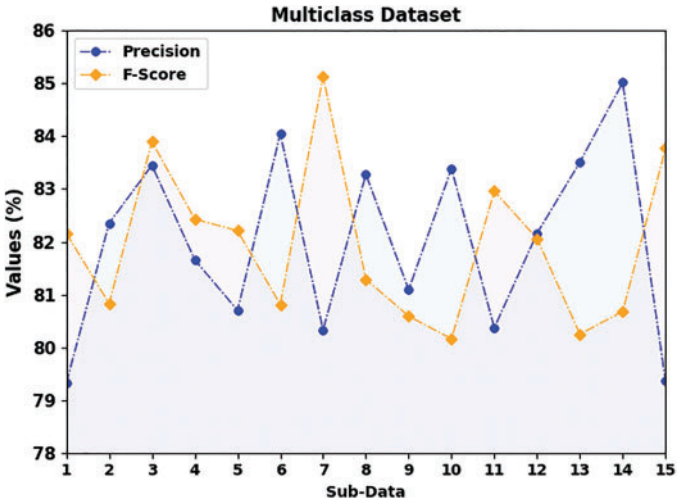

Figure 7: $Prec_n$ and F_{score} analysis of BAOML-CADC system under multiclass database

In Fig. 8, a complete $accu_y$ outcomes of the BAOML-CADC method on the Multiclass dataset are provided. The results show the improved value of $accu_y$ under all classes. For example, with SD-1 class, the BAOML-CADC method has provided an $accu_y$ of 93.64%. On the other hand, with SD-2 class, the BAOML-CADC approach has shown *an* $accu_y$ of 94.67%. In the meantime, with the SD-10 class, the BAOML-CADC algorithm has rendered *an* $accu_y$ of 94.05%. Furthermore, with the SD-15 class, the BAOML-CADC technique has an $accu_y$ of 91.77%.

Table 3 presents a comparative analysis of the BAOML-CADC with existing approaches [26]. The comprehensive comparative examination of the BAOML-CADC method with existing models on binary class is described in Fig. 9. The outcomes exhibited that the RF technique has reached poor performance with *an* $accu_y$ of 80.61% while the JRip model has managed to obtain slightly improved outcomes. Additionally, the Adaboost+JRip and k-nearest neighbour (KNN) methods have accomplished closer performance with $accu_y$ of 95.56% and 95.49%, respectively. Although the BDLE-CAD technique has resulted in reasonable outcomes with $accu_y$ of 98.63%, the BAOML-CADC technique surpassed the other ones with *an* $accu_y$ of 99.14%.

The comprehensive comparative examination of the BAOML-CADC model with existing methods on multiclass is given in Fig. 10. The outcomes show that the RF method has reached poor performance with *an* $accu_y$ of 80.51% while the JRip algorithm has managed to gain slightly improved outcomes. Also, the Adaboost+JRip and KNN methodologies have accomplished closer performance with $accu_y$ of 91.44% and 87.66%, correspondingly. Although the BDLE-CAD method has resulted

in reasonable outcomes with *an accu*$_y$ of 92.63%, the BAOML-CADC approach surpassed the other ones with an *accu*$_y$ of 93.17%.

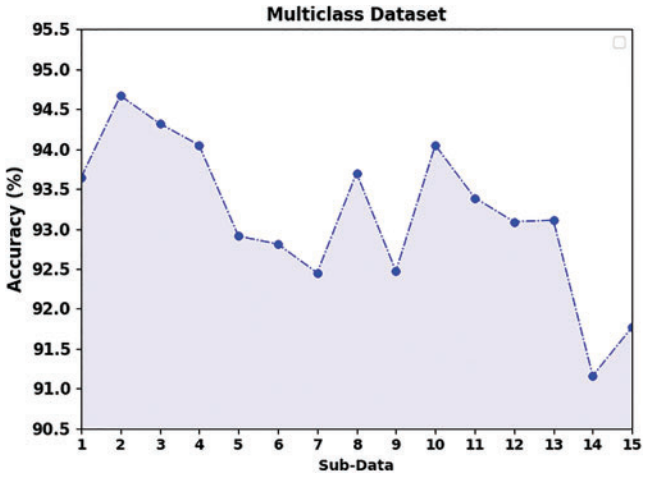

Figure 8: *Accu*$_y$ analysis of BAOML-CADC system under multiclass database

Table 3: Comparative analysis of BAOML-CADC approach under binary and multiclass datasets

Methods	Binary class	Multi-class
BAOML-CADC	99.14	93.17
BDLE-CAD	98.63	92.63
KNN	95.49	87.66
Random forest	80.61	80.51
JRip	90.10	90.09
Adaboost+JRip	95.56	91.44

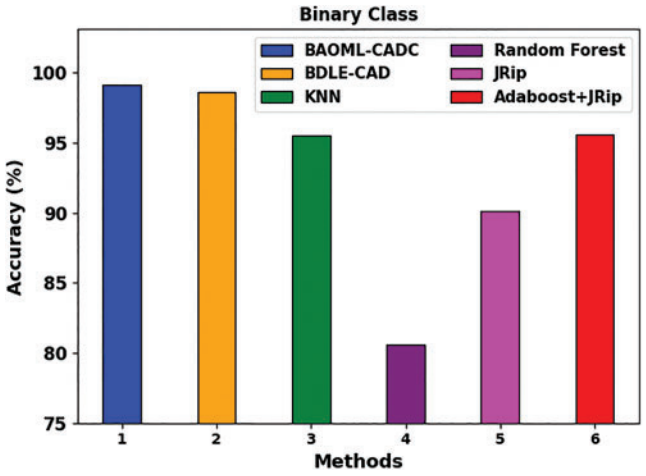

Figure 9: *Accu*$_y$ analysis of BAOML-CADC approach under binary class dataset

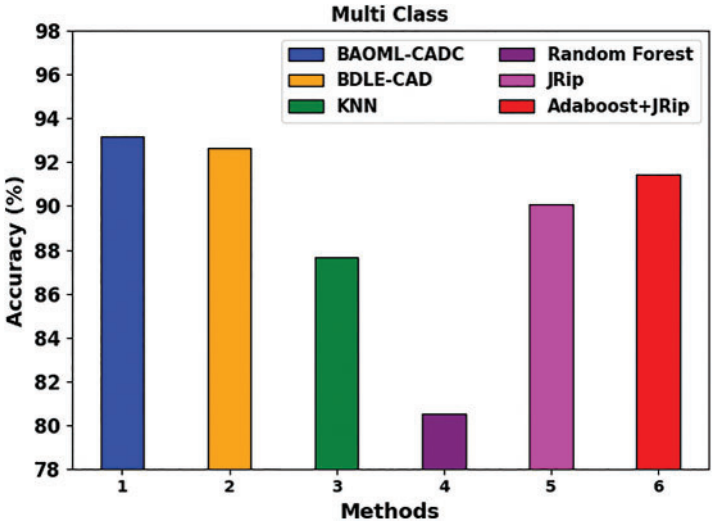

Figure 10: $Accu_y$ analysis of BAOML-CADC approach under multiclass dataset

These results demonstrated the supremacy of the BAOML-CADC model on cyberattack classification.

5 Conclusion

In this study, we have developed a new BAOML-CADC technique for Cyberattack Detection and Classification process. In the BAOML-CADC technique, the major focus lies in identifying cyberattacks. Initially, the presented BAOML-CADC technique applied the TEA-FS approach for the optimal choice of features. For cyberattack recognition, the BAOML-CADC technique used the ELM model. In addition, a BC-based integrity verification technique is developed to protect against the misrouting attack. The experimental validation of BAOML-CADC technique is tested on a benchmark cyberattack dataset. The obtained values implied the improved performance of the BAOML-CADC algorithm over other models with maximum accuracy of 93.17%. In the future, the performance of BOAML-CADC technique can be improved using a hyperparameter tuning process.

Funding Statement: This work was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, through the Research Groups Program Grant No. (RGP-1443-0051).

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

References

[1] D. G. Roy and S. N. Srirama, "A blockchain-based cyber attack detection scheme for decentralized internet of things using the software-defined network," *Software: Practice and Experience*, vol. 51, no. 7, pp. 1540–1556, 2021.

[2] M. Abdel-Basset, N. Moustafa and H. Hawash, "Privacy-preserved cyberattack detection in industrial edge of things (IeoT): A blockchain-orchestrated federated learning approach," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7920–7934, 2022.

[3] J. Zhang, L. Pan, Q. L. Han, C. Chen, S. Wen *et al.,* "Deep learning based attack detection for cyber-physical system cybersecurity: A survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 377–391, 2021.

[4] O. Ajayi, M. Cherian and T. Saadawi, "Secured cyber-attack signatures distribution using blockchain technology," in *IEEE Int. Conf. on Computational Science and Engineering (CSE) and IEEE Int. Conf. on Embedded and Ubiquitous Computing (EUC)*, New York, NY, USA, pp. 482–488, 2019.

[5] V. Kelli, P. Sarigiannidis, V. Argyriou, T. Lagkas and V. Vitsas, "A cyber resilience framework for NG-IoT healthcare using machine learning and blockchain," in *CC-IEEE Int. Conf. on Communications*, Montreal, QC, Canada, pp. 1–6, 2021.

[6] T. V. Khoa, Y. M. Saputra, D. T. Hoang, N. L. Trung, D. Nguyen *et al.,* "Collaborative learning model for cyberattack detection systems in IoT industry 4.0," in *IEEE Wireless Communications and Networking Conf. (WCNC)*, Seoul, Korea (South), pp. 1–6, 2020.

[7] M. Dehghani, T. Niknam, M. Ghiasi, N. Bayati and M. Savaghebi, "Cyber-attack detection in dc microgrids based on deep machine learning and wavelet singular values approach," *Electronics*, vol. 10, no. 16, pp. 1914, 2021.

[8] N. Waheed, X. He, M. Ikram, M. Usman, S. S. Hashmi *et al.,* "Security and privacy in IoT using machine learning and blockchain: Threats and countermeasures," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1–37, 2020.

[9] A. Javadpour, P. Pinto, F. Ja'fari and W. Zhang, "DMAIDPS: A distributed multi-agent intrusion detection and prevention system for cloud IoT environments," *Cluster Computing*, vol. 7, no. 3, pp. 150.936, 2022. https://doi.org/10.1007/s10586-022-03621-3

[10] M. Hajizadeh, N. Afraz, M. Ruffini and T. Bauschert, "Collaborative cyber attack defense in SDN networks using blockchain technology," in *6th IEEE Conf. on Network Softwarization (NetSoft)*, Ghent, Belgium, pp. 487–492, 2020.

[11] N. Mhaisen, N. Fetais and A. Massoud, "Secure smart contract-enabled control of battery energy storage systems against cyber-attacks," *Alexandria Engineering Journal*, vol. 58, no. 4, pp. 1291–1300, 2019.

[12] M. Komar, V. Dorosh, G. Hladiy and A. Sachenko, "Deep neural network for detection of cyber attacks," in *IEEE First Int. Conf. on System Analysis & Intelligent Computing (SAIC)*, Kyiv, Ukraine, pp. 1–4, 2018.

[13] R. M. A. Ujjan, Z. Pervez and K. Dahal, "Snort based collaborative intrusion detection system using blockchain in SDN," in *13th Int. Conf. on Software, Knowledge, Information Management and Applications (SKIMA)*, Island of Ulkulhas, Maldives, pp. 1–8, 2019.

[14] O. O. Malomo, D. B. Rawat and M. Garuba, "Next-generation cybersecurity through a blockchain-enabled federated cloud framework," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 5099–5126, 2018.

[15] D. Said, M. Elloumi and L. Khoukhi, "Cyber-attack on P2P energy transaction between connected electric vehicles: A false data injection detection based machine learning model," *IEEE Access*, vol. 10, pp. 63640–63647, 2022.

[16] M. Ghiasi, M. Dehghani, T. Niknam, A. Kavousi-Fard, P. Siano *et al.,* "Cyber-attack detection and cyber-security enhancement in smart DC-microgrid based on blockchain technology and Hilbert Huang transform," *IEEE Access*, vol. 9, pp. 29429–29440, 2021.

[17] O. Ajayi and T. Saadawi, "Blockchain-based architecture for secured cyber-attack features exchange," in *7th IEEE Int. Conf. on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE Int. Conf. on Edge Computing and Scalable Cloud (EdgeCom)*, New York, NY, USA, pp. 100–107, 2020.

[18] A. Kumar, A. K. Singh, I. Ahmad, P. K. Singh, P. K. Verma *et al.,* "A novel decentralized blockchain architecture for the preservation of privacy and data security against cyberattacks in healthcare," *Sensors*, vol. 22, no. 15, pp. 5921, 2022.

[19] P. Kumar, R. Kumar, G. Srivastava, G. P. Gupta, R. Tripathi *et al.,* "PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2326–2341, 2021.

[20] H. Cui, X. Dong, H. Deng, M. Dehghani, K. Alsubhi *et al.,* "Cyber attack detection process in sensor of DC micro-grids under electric vehicle based on Hilbert-Huang transform and deep learning," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15885–15894, 2020.

[21] G. Liang, S. R. Weller, F. Luo, J. Zhao and Z. Y. Dong, "Distributed blockchain-based data protection framework for modern power systems against cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3162–3173, 2018.

[22] W. Pakdee and T. Sakkarangkoon, "Numerical study of an unsteady non-premixed flame in a porous medium based on the thermal equilibrium model," *Journal of Theoretical and Applied Mechanics*, vol. 59, no. 3, pp. 401–412, 2021.

[23] L. Li, Z. Liu, Y. Lu, F. Wang and S. Jeon, "Hard-rock tunnel thrust prediction with TBM construction big data using an improved two-hidden-layer extreme learning machine," *IEEE Access*, vol. 10, pp. 112695–112712, 2022.

[24] A. Derhab, M. Guerroumi, A. Gumaei, L. Maglaras, M. A. Ferrag *et al.,* "Blockchain and random subspace learning-based IDS for SDN-enabled industrial IoT security," *Sensors*, vol. 19, no. 14, pp. 3119, 2019.

[25] S. Abe, Y. Uchida, M. Hori, Y. Hiraoka and S. Horata, "Cyber threat information sharing system for industrial control system (ICS)," in *2018 57th Annual Conf. of the Society of Instrument and Control Engineers of Japan (SICE)*, Nara, Japan, pp. 374–379, 2018.

[26] M. Ragab and A. Altalbe, "A blockchain-based architecture for enabling cybersecurity in the internet-of-critical infrastructures," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1579–1592, 2022.