scientific reports

Check for updates

OPEN

Explainable artificial intelligence-based cyber resilience in internet of things networks using hybrid deep learning with improved chimp optimization algorithm

Sarah A. Alzakari[1], Mohammed Aljebreen[2], Nazir Ahmad[3], Sultan Alahmari[4], Othman Alrusaini[5], Ali Alqazzaz[6], Hassan Alkhiri[7] & Yahia Said[8]✉

The rapid growth of the Internet of Things (IoT) has driven new research into artificial intelligence (AI)-based methods for detecting anomalies. With its advanced capabilities, AI can automate tasks, analyze large datasets, and accurately identify vulnerabilities. The lack of transparency in cybersecurity systems makes it difficult to explain critical decisions and associated risks clearly. Machine learning (ML)-based intrusion detection systems (IDS) excel in threat detection but encounter threats due to limited transparency and scarce attack data, specifically in IoT. This paper presents the Explainable Artificial Intelligence for Cyber Resilience Using a Hybrid Deep Learning and Optimization Algorithm (XAICR-HDLOA) approach to improve cyber threat detection and interpretation in IoT environments. Min-max normalization is initially applied to standardize feature scales, followed by the Bald Eagle Search (BES) model for selecting key features. Moreover, the hybrid Convolutional Neural Networks-Bidirectional Gated Recurrent Unit (CNN-BiGRU) model is employed for cyberattack classification. Furthermore, the Improved Chimp Optimizer Algorithm (IChoA) is implemented for the hyperparameter tuning process. Finally, SHAP is applied to improve model interpretability, increasing trust and reliability in cybersecurity. Simulations on the Edge-IIoT and BoT-IoT datasets highlight the efficiency of the XAICR-HDLOA approach, achieving high accuracy of 98.41% and 98.25%, outperforming existing methods.

Keywords Data normalization, Deep learning, Explainable artificial intelligence, Cybersecurity, Dimensionality reduction, Internet of things

The IoT has recently developed, allowing ubiquitous computing and sensing to link many things to the Internet[1]. An IoT system links with other crucial structures in the smart city environment, like telecommunication networks, smart homes, and smart airports, to give citizens diverse benefits that can improve their lives[2]. An IoT structure is a cyber-physical system (CPS) that contains physical and computational abilities that permit interactions, data sharing, and connections between devices and machines to upgrade efficiency and functionality without requiring a person in the loop[3]. Consequently, such gadgets' memory and power limits make this system susceptible to hacking and contribute to cyber threats. Thus, cybersecurity has become vital to keep these systems operational[4]. Cybersecurity is securing gadgets, data, and systems against illegal usage

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. [2]Department of Computer Science, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia. [3]Department of Computer Science, Applied College at Mahayil, King Khalid University, Abha, Saudi Arabia. [4]King Abdul Aziz City for Science and Technology (KACST), Cybersecurity Institute, Riyadh, Kingdom of Saudi Arabia. [5]Department of Engineering and Applied Sciences, Applied College, Umm Al-Qura University, Makkah, Saudi Arabia. [6]Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Bisha 67714, Saudi Arabia. [7]Department of Computer Science, Faculty of Computing and Information Technology, Al-Baha University, Bisha, Saudi Arabia. [8]Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia. ✉email: Yahia.said@nbu.edu.sa

or unauthorized access and retaining information privacy, availability, and integrity. At the same time, cyber defensive mechanisms develop at the network, application, data, and host levels. The Internet has become a crucial device in people's everyday lives, and the number of systems connected to the Internet extends too[5].

The development of mobile devices, computer servers, and networks has considerably increased Internet utilization. Explainable AI (XAI) increases attention to several applications due to having multiple benefits, like being a trustworthy, highly transparent, and interpretable method[6]. AI methods are being developed every day with more advanced aspects. AI has also become a stage where human brains can quickly interface with machines. Nevertheless, these are frequently susceptible to method bias, absence of code, and trust concerns[7]. To address such hazards and maintain the AI methods transparent, the development of XAI provides a significant understanding of the method without any confusion when making decisions or embracing solutions. Consequences of XAI in the existing businesses can switch the traditional AI methods, and make a significant impact with better improvement and advancement in the manufacturing, production, wealth management, financial sectors, and supply chain[8]. Recently, XAI technology has been of extensive interest to both academia and industry. The development of this technology has attained considerable success, and trustworthy decisions have been made using these methods[9]. The XAI application in cybersecurity might be a double-edged sword: it can significantly enhance cybersecurity methods and assist in addressing novel threats to AI applications. It is also Explainable to the attacker who can pose severe security attacks[10]. AI methods, intense learning (DL) and ML methodologies can give impressive performances on benchmark datasets in various applications of the cybersecurity field.

This paper presents the Explainable Artificial Intelligence for Cyber Resilience Using a Hybrid Deep Learning and Optimization Algorithm (XAICR-HDLOA) approach to improve cyber threat detection and interpretation in IoT environments. Min-max normalization is initially applied to standardize feature scales, followed by the Bald Eagle Search (BES) model for selecting key features. Moreover, the hybrid Convolutional Neural Networks-Bidirectional Gated Recurrent Unit (CNN-BiGRU) model is employed for cyberattack classification. Furthermore, the Improved Chimp Optimizer Algorithm (IChoA) is implemented for the hyperparameter tuning process. Finally, SHAP is applied to improve model interpretability, increasing trust and reliability in cybersecurity. Simulations of the XAICR-HDLOA approach are performed under the Edge-IIoT and Bot-IoT datasets. The key contribution of the XAICR-HDLOA approach is listed below.

- The XAICR-HDLOA approach applies min-max normalization to standardize feature scales, ensuring consistent data processing and improving the model's capability for handling diverse input data. This technique enhances the accuracy of subsequent algorithms by removing biases caused by varying feature ranges. It contributes to the overall efficiency of the model, making it more reliable in IoT environments.
- The XAICR-HDLOA approach effectively employs the BES model to choose the most relevant features, optimizing the classification process. Detecting key features mitigates dimensionality and improves the model's performance. This approach confirms that only the most informative data is used, improving the accuracy and efficiency of threat detection in IoT systems.
- The XAICR-HDLOA approach implements the hybrid CNN-BiGRU classification approach, utilizing the power of CNNs for feature extraction and BiGRU units for capturing temporal dependencies. This integration improves the model's ability to detect complex patterns and identify threats. It significantly enhances the robustness and precision of the IDS in dynamic IoT environments.
- The XAICR-HDLOA approach employs the IChoA technique to fine-tune model parameters, optimizing the search for optimal solutions. Adjusting hyperparameters more effectively improves the overall performance and accuracy of the model. This results in more precise predictions and enhanced efficiency in real-time threat detection within IoT systems.
- The novelty of the XAICR-HDLOA approach is in its unique combination of advanced techniques, incorporating a hybrid CNN-BiGRU model with BES for efficient feature selection and IChoA for optimization. This approach is specifically designed to improve intrusion detection in resource-constrained IoT environments. By seamlessly combining these methods, the model balances performance, accuracy, and computational efficiency, making it appropriate for real-time IoT applications.

Comprehensive literature review on cybersecurity in IoT and IIoT networks

Birahim et al.[11] developed an innovative IDS utilizing PSO and an ensemble ML method associating DT, KNN, and RF methods to improve the precision and dependability of intrusion detection in WSN. The projected method accomplishes substantial growth by integrating OTE-Tomek models to balance the data, and proposes utilizing XAI models like SHAP and LIME. Narkedimilli et al.[12] projected a lightweight and scalable curriculum learning structure developed with XAI models, comprising LIME. The presented method utilizes an innovative neural network (NN) structure employed at each phase of Curriculum Learning. Sturdiness is accomplished through staged learning, where the technique iteratively upgrades itself by extracting lower-relevance aspects and enhancing performance. The workflow comprises edge-optimized pruning and quantization to safeguard portability that can be employed in the edge-IoT gadgets. An ensemble method integrating random forest (RF), XGBoost, and the staged learning base continues to improve generalization. Naif Alatawi[13] projected an innovative IDS structure, which incorporates sophisticated ML models containing transfer learning (TL), feature engineering, and ensemble learning, to improve recognition precision, interpretability, and adaptability. The ensemble learning module integrates different classifiers, for instance, DT, KNN, and LR, by utilizing their unique capabilities to increase recognition rates. Pre-training techniques applied to TL are connected to cybersecurity datasets and fine-tuned on the combined dataset. Izuazu et al.[14] projected the eXplainable cyber-threat detection framework (XC-TDF) as an innovative solution to overcome these tasks. The projected model improves sturdiness against adversarial threats and noise by applying adversarial training and regularisation correspondingly, and

also upgrades transparency over an XAI component. Patel et al.[15] developed X-NET, an XAI-based system data security method for medical care 4.0 applications. For comparison purposes, five diverse kinds of conventional feature extraction models are employed together with logistic regression (LR), naive bayes (NB), and Insight. Using X-AI models like SHAP and LIME has substantially boosted X-NET's performance and dependability. Baral et al.[16] developed a novel, wide-ranging structure for real-world IoT threat recognition and response that utilizes XAI, LLM, and ML. Combining XAI models like LIME and SHAP with a model-independent structure guarantees this structure's flexibility through several ML models. Furthermore, integrating LLM improves the accessibility and interpretability of recognition decisions and human-understandable explanations of identified attacks. Tripathy et al.[17] projected a structure that depends on XAI for safeguarding user IoT applications in smart cities. At the initial stage of protocol execution, the participants switch authenticated data over the blockchain (BC)-based AKA process. Simultaneously, this model implements the Python-based SHAP structure to interpret and explain the core aspects guiding decision-making.

In[18], a BC-enabled XAI is projected to improve the decision-making ability of cyber-attack recognition in the context of Smart Medical care Methods. Initially, this model utilizes BC to store and validate data among different cloud vendors by applying a Clique Proof-of-Authority (C-PoA) consensus. Then, a new DL-based threat-hunting method is developed by relating Parallel Stacked LSTM (PSLSTM) techniques with a multi-head attention mechanism for enhanced threat recognition. Zeghida et al.[19] developed ML and DL models such as RF, support vector machine (SVM), convolutional neural network (CNN), CNN with long short-term memory (CNN-LSTM) for detecting various cyberattacks. Also, an explainable AI framework using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) is utilized to provide transparent and interpretable intrusion detection. Nandanwar and Katarya[20] presented a robust DL technique named AttackNet, which utilizes an adaptive CNN and gated recurrent unit (CNN-GRU) methodology for accurate detection and classification of botnet attacks. Reddy et al.[21] developed an advanced intrusion prevention framework using graph-residual adversarial network (GRANet) approach incorporated with hawk-bee stride finder (HBSF) optimization. Nandanwar and Katarya[22] proposed a robust CNN Neural Network–bidirectional long short-term memory (CNN-BiLSTM) with TL-BiLSTM. Nandanwar and Katarya[23] presented a robust DL-based IDS named cyber-sentinet for CPS in industrial IoT environments. The model integrates SHAP to improve interpretability while accurately detecting diverse cyber-attacks. Kauhsik et al.[24] proposed a methodology by using ML and DL techniques. Nandanwar and Katarya[25] developed a secure and efficient healthcare data management system by integrating BC technology, smart contracts, non-interactive zero-knowledge proof (NIZKP), and inter-planetary file system (IPFS) with an IDS to ensure data confidentiality, integrity, and privacy in IoT-enabled healthcare environments. Attique et al.[26] developed a transparent and data-efficient IDS for IIoT environments by using BiLSTM and a self-adaptive attention mechanism (S-AAM). Additionally, SHAP from explainable artificial intelligence (XAI) are integrated to improve model interpretability and trustworthiness. Bajpai and Patankar[27] presented a self-configuring intrusion detection framework for BC networks using Adaptive Goal Target Optimization with Deep BiLSTM (AGLSTM). The model also incorporates rider-cheetah hybrid optimization (RCHO) and synthetic minority oversampling technique (SMOTE) to improve feature learning and data balance. Elsaid et al.[28] proposed an optimized IDS by integrating grey wolf optimization (GWO) with DL models such as GRU and LSTM (GRU-GWO) and (LSTM-GWO).

Summary of related works with identification of research gaps and limitations

Table 1 summarises the existing studies on cyberthreat detection.

The limitations of the existing studies are in the reliance on specific, often limited, datasets, which may not fully capture the diverse and growing nature of IoT threats. Several models also encounter threats in terms of scalability, specifically when applied to resource-constrained devices in real-time IoT environments. Though the XAI models are utilized, transparency and interpretability remain constrained, particularly with complex ensemble or DL models. Furthermore, a research gap exists in addressing the robustness of these systems against adversarial attacks and noise. The integration of BC for data validation is promising but still lacks comprehensive testing in large-scale, heterogeneous IoT networks. Moreover, many approaches concentrate on detection accuracy, neglecting the optimization of response times and real-time adaptability in dynamic IoT systems. Additionally, explainability and transparency in decision-making remain underexplored, limiting trust and practical deployment.

The proposed methodology

This paper proposes the XAICR-HDLOA approach. The main objective of the XAICR-HDLOA approach is to enhance cyber threat detection and interpretation in IoT environments. To accomplish this, the XAICR-HDLOA approach has data normalization, BES-based feature selection, a hybrid of CNN-BiGRU models, and parameter tuning using IChoA. Figure 1 represents the entire procedure of the XAICR-HDLOA approach.

Data normalization: min-max normalization

Initially, the XAICR-HDLOA approach applies the min-max normalization approach to standardize feature scales during the data normalization process[29]. This model is chosen due to its simplicity and efficiency in scaling features within [0, 1]. This confirms that all features contribute equally to the model, preventing dominance by those with larger scales. Compared to other techniques, such as Z-score normalization, this technique does not assume a normal distribution, making it more appropriate for diverse datasets, particularly those with skewed or non-normal distributions. Furthermore, it is computationally efficient and easy to implement, making it ideal for real-time applications in IoT systems. Its capability to preserve the relationships between the original data values improves the model's performance. It balances simplicity and effectiveness, specifically in environments with resource constraints.

Ref.	Techniques	Metrics	Findings
11	PSO, RF, DT, KNN, SMOTE-Tomek, LIME, SHAP	Accuracy, Precision, Recall, F1-Score	The proposed IDS outperforms existing methods, ensuring robust WSN security.
12	Curriculum Learning, LIME, NN, Staged Learning, Quantization & Pruning, Ensemble Model (RF, XGBoost)	Accuracy	The framework presents high accuracy and robustness, ensuring reliable IoT network security.
13	Ensemble Learning, TL, Feature Engineering, LIME, SHAP	Accuracy, Detection Rate Increase, Training Time Reduction	The framework improves detection accuracy, adaptability, and transparency for IDS in diverse environments.
14	XC-TDF, Adversarial Training, Regularization, XAI	Accuracy, Precision, Recall, F1-Score, MCC	XC-TDF enhances robustness, transparency, and accuracy, resisting adversarial attacks and noise.
15	X-NET, X-AI, NB, LR, Perceptron	Accuracy, Precision, Recall, F1-Score, ROC	X-NET enhances security and performance for remote patient monitoring utilizing X-AI models.
16	ML, XAI, SHAP, LIME, LLM, Gemini, OPENAI	Standard Metrics	The framework improves IoT attack detection and response by utilizing XAI and LLM for better accuracy and insights.
17	Consumer IoT, XAI, IDS, BC, SHAP, Python-based Framework	Benchmark Metrics	The framework improves IoT security in smart cities utilizing XAI, SHAP, and BC.
18	BC, XAI, C-PoA, PSLSTM, Multi-head attention	Accuracy, Precision, Recall, F1-Score	Enhances decision support for cybersecurity analysts in smart healthcare systems.
19	RF, SVM, CNN-LSTM, SHAP, LIME	Accuracy, Precision, Recall, F1-Score	The model attained highest accuracy of 99.9%; XAI enhanced model interpretability.
20	Adaptive CNN-GRU, DL, Botnet Attack Classification	Accuracy, Loss, Precision, Recall	AttackNet attained 99.75% accuracy, outperforming existing methods; future work will address real-time scalability and dataset diversity.
21	GRANet, HBSF, Swarm Intelligence Tuning	Precision, Recall, Detection Rate	The framework presents superior detection.
22	Hybrid CNN-BiLSTM, TL	Testing and Training Accuracy, Loss	The model attains 99.52% accuracy, surpassing existing methods.
23	DL, IDS, SHAP	Accuracy, Precision, Recall, Loss	The model attained high detection performance with interpretability.
24	ML, DL	Security Enhancement, Vulnerability Identification	The review identifies gaps and suggests combined techniques to enhance IoT security.
25	BC, Smart Contracts, NIZKP, IPFS, IDS	Security Efficiency, Privacy Preservation, Storage Cost Reduction, System Scalability	The model improves healthcare security and privacy with cost savings.
26	BiLSTM, S-AAM, SHAP, XAI	Accuracy, Dataset Efficiency, Interpretability	Achieved 99.92% and 96.54% accuracy.
27	AGLSTM, RCHO, SMOTE	Accuracy, Sensitivity and Specificity Rate, Training Percentage, K-Fold Validation	AGLSTM performs well across datasets.
28	GRU-GWO, LSTM-GWO	Detection Accuracy, Dimensionality Reduction (DR), Computational Efficiency	Enhanced accuracy and efficiency with GRU-GWO and LSTM-GWO models.

Table 1. Summary of the BC-enabled XAI framework for cyber threat detection.

A min-max scaling method was applied to certify data consistency and uniformity across measures. This method regularizes data by regulating the least and most significant values to 0 to 1, respectively, with every intermediate value measured evenly in this collection. The primary purpose of standardization is to avert the uneven amplification of input features, which might badly impact the learning procedure. Moreover, using standardized data is vital in NNs as it aids in decreasing error propagation. The mathematical formulation is given in Eq. (1).

$$x_{scaled} = \frac{x_{real} - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Dimensionality reduction: BES approach

For dimensionality reduction, the BES model is employed to select the most relevant features[30]. This model is chosen for its effectiveness in selecting the most relevant features while maintaining high classification performance. Unlike conventional methods such as PCA, which mitigate dimensions based purely on variance, BES utilizes a nature-inspired optimization technique to detect key features directly affecting the model's predictive capabilities. This technique is highly efficient in intrinsic datasets, averting the loss of crucial data during reduction. The model also adapts well to nonlinear relationships and averts overfitting by concentrating on the most relevant features for the task. Its capability in handling massive, high-dimensional datasets makes it ideal for resource-constrained IoT environments, where computational efficiency is crucial. Overall, BES balances accuracy and efficiency, ensuring optimal performance for IDS. Figure 2 illustrates the BES technique.

As a stochastic optimizer model, BES is stimulated naturally and established in the group's behaviours. BES derives from the joint hunting approaches bald eagles use while hunting prey. The population of bald eagles matches their assault on prey over three distinct tactics, such as choosing the search region, recognizing the prey, and diving to take the prey.

Choose search scope

The eagle group presents data in this phase to maintain the swarm's unity. The behaviour of tth group in the search space is expressed below:

$$p_i^{(c+1)_s} = p_*^c + \alpha \cdot r_1 \cdot (p_{mean}^c - p_i^c) \tag{2}$$

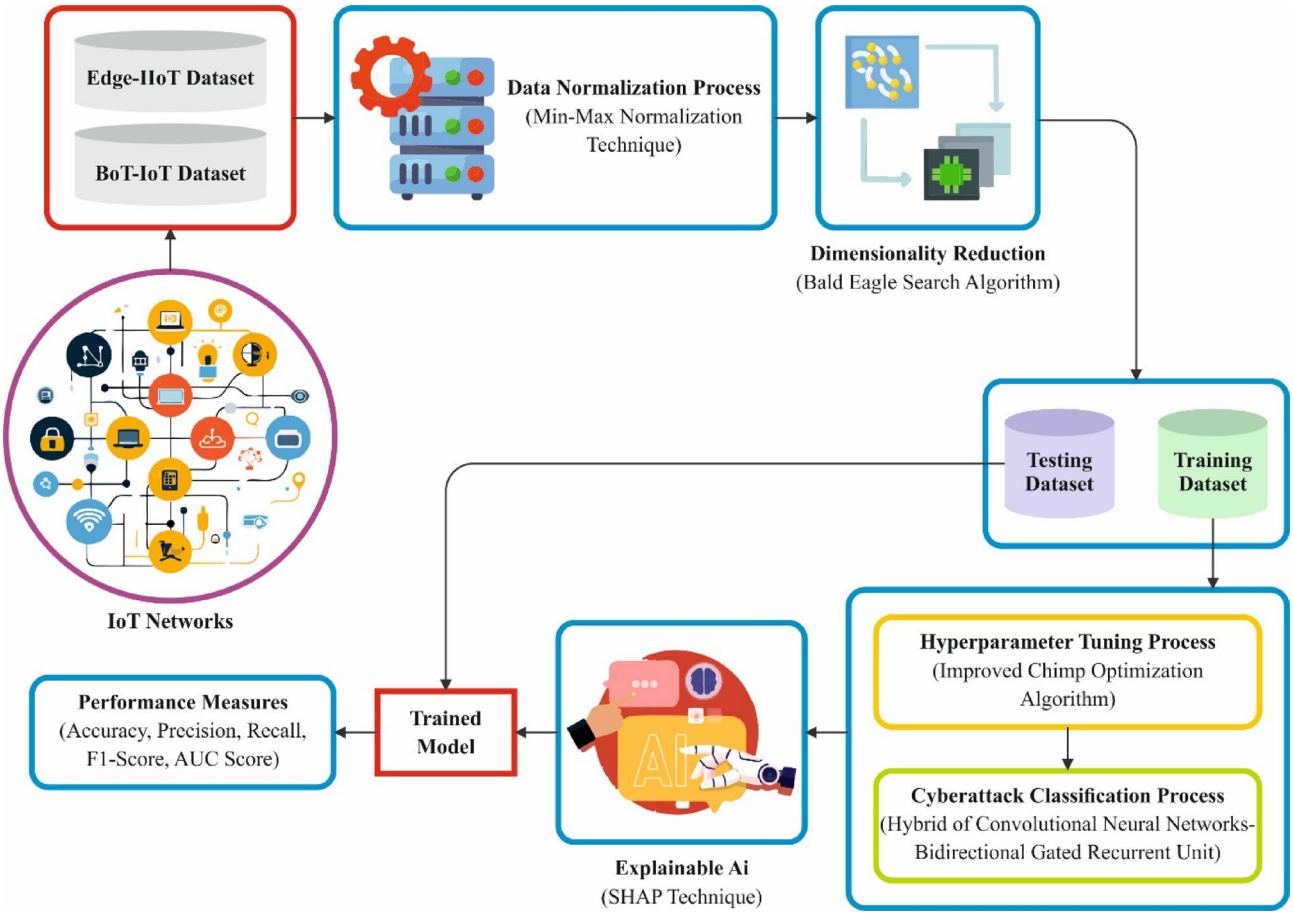

Fig. 1. Overall process of XAICR-HDLOA approach.

Here, $\alpha \in (1.5,2), r_1 \in [0,1]$, $p_{meon}^t = \frac{1}{N}\sum_{i=1}^{N} p_i^t, p_*^t$ signifies the finest individual, N represents the dimension of the swarm, p_i^t, and $p_i^{(t+1)_s}$ epitomize an ith individuals before and after the searching space range. Over the progress of Eq. (2), a population $p^{(t+1)_s} = [p_1^{(t+1)_s}, p_2^{(t+1)_s}, \ldots, p_N^{(t+1)_s}]$ is moulded.

Pick quarry
Here, the flock of eagles utilizes chain rubrics for spiral upgrades.

$$\theta_i^{tz} = a \cdot \pi \cdot randr_i^{tz} = \theta_i^{tz} + R \cdot rand$$
$$xr_i^{tz} = r_i^{rz} \cdot \sin\left(\theta_i^{tz}\right) y^{\beta_i} = r_i^{tz} \cdot \cos\left(\theta_i^{tz}\right) \tag{3}$$
$$x_i^{tz} = xr_i^{tz} / \max_{1 \le i \le N} (|xr_i^{tz}|)\, y_i^{tz} = yr_i^{tz} / \max_{1 \le i \le N} (|yr_i^{tz}|)$$

$$p_i^{(t+1)_z} = p_i(+x_i \cdot \left(p_i^{(t+1)_s} - p_{mean}^t\right) + y_i^{rz} \cdot \left(p_i^{(t+1)_s} - p_{i+1}^{(t+1)_s}\right) \tag{4}$$

While $a \in (0,5)$and R $\in (052\, are$ employed for controlling the dimension of the spiral. $p_{i+1}^{(t+1)_s}$ denotes the bald eagle whose population pis is situated at $the\, i+1-th$ location, and $p_i^{(t+1)_z}$ signifies the discrete upgrade by Eq. (4).

Dive to catch prey
If the bald eagle has eliminated in the optimum location of its prey, then it will protect its novel position by finding a hyperbolic route.

$$\theta_i^t = a \cdot \pi \cdot randr_i^t = \theta_i^t$$
$$xr_i^t = r_i^t \cdot \sinh\left(\theta_i^t\right) yr_i^t = r_i^t \cdot \cosh\left(\theta_i^t\right) \tag{5}$$

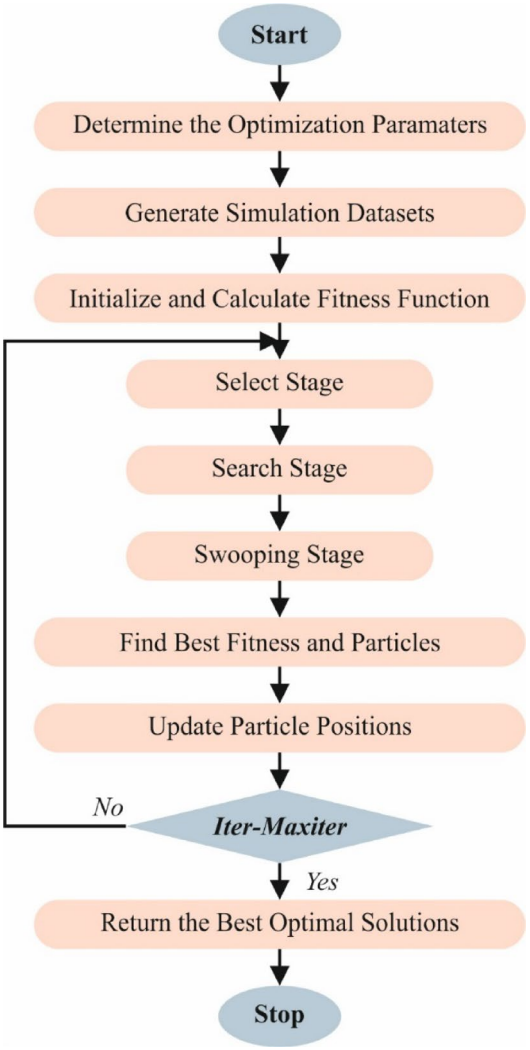

Fig. 2. Workflow of the BES model.

$$x_{\mathrm{i}}^{\mathrm{tz}} = xr_{\mathrm{i}}^{\mathrm{tz}} / \max_{1 \le i \le N} (|\, xr\,|)\, yl_i^c = yr_i^{\mathfrak{c}/\max(1 \mathcal{Y}^{\Gamma\, f_i}|)}$$

$$p_i^{t+1} = rand \cdot\ p_*^{tm} + xl_{\mathrm{i}}^t \cdot \left(p_i^{(t+1)_z} - c_1 \cdot\ p_{meon}^t\right) + yl_{\mathrm{i}}^t \cdot \left(p_i^{(t+1)_z} - c_2 \cdot\ p_*^{tm}\right) \tag{6}$$

Here, $sinh\,()$ and $\cosh\,()$ represent the hyperbolic functions, P^{tm} is the finest solution, and $c_1, c_2 \in (1{,}2)$. The BES concludes the upgrading of dispersion and aggregation over Eqs. (2)- (6) and discovers an optimum solution to an issue over any iterations.

In the BES approach, the fitness function (FF) employed is intended to balance the number of selected features (least) and the accuracy of classification (highest) attained by consuming these chosen features. Equation (7) characterizes the FF to estimate the solution.

$$Fitness = \alpha\, \gamma_{\ R}\,(D) + \beta\, \frac{|R|}{|C|} \tag{7}$$

Here, $\gamma_{\ R}\,(D)$ signifies an assumed classifier's classification error rate. $|R|$ refers to the cardinality of the chosen subset; $|C|$ represents the total number of features in the dataset, α and β represent a binary parameter corresponding to the impact of classifier excellence and subset length. $\in$ [1,0] and $\beta = 1 - \alpha$ ..

Hybrid classification process: CNN-BiGRU model

Moreover, the hybrid of CNN-BiGRU model is employed for cyberattack classification[31]. This model was chosen for its capability of efficiently capturing both spatial and temporal dependencies in data. BiGRU shows efficiency in capturing sequential data by processing data in both forward and backwards directions, while CNNs outperform in extracting raw data, and detecting intrinsic patterns. This integration allows the model

to handle spatial patterns effectively (from CNNs) and temporal relationships (from BiGRUs), making it highly appropriate for intrusion detection in dynamic IoT environments. Unlike conventional classification models, this hybrid approach can better adapt to real-time, evolving threats. It also gives superior accuracy in detecting complex attack patterns, enhancing anomaly robustness. The model's capability to process diverse data types makes it more effective than single-method models, ensuring high performance even in resource-constrained settings. Figure 3 represents the infrastructure of CNN-BiGRU technique.

An ID-CNN is utilized to extract the complete features of an input feature. The Bi-GRU technique was employed to mine the links between input features. First, the feature subset is used as an input after feature selection. Then, CNN is employed to extract the links among features, combining manifold attributes to absorb the coupling features between them. Furthermore, the removed features are input to extract the sequential feature. Lastly, the backwards and forward outputs were combined and produced. CNN-BiGRU can efficiently mine both global and local feature data.

The main aim of CNN is to mine noticeable features from input data. A classic CNN structure includes numerous layers, such as convolutional, pooling, and fully connected (FC). The convolutional layer is fundamental in extracting features, where convolution kernels take appropriate features from input data. The abstraction level of the mined feature increases with the number of convolution kernels employed. The FC layers compress the pooling neurons into a 1D vector method, enabling more convenient data processing. Conversely, CNNs were ineffectual in seizing the time-based dependencies. So, it is vital to incorporate the recurrent NN (RNN) methods and unite CNN with Bi-GRU models to enhance the performance.

GRU is a kind of RNN method generally employed for handling successive data that tackles long-term dependency problems. When equated to conventional RNNs, GRU presents a gating device that allows it to acquire, forget, or retain data efficiently. GRU includes dual gates, such as reset and update. The update gate manages how much historical data is to be recollected. In contrast, the reset gate aids the system in defining how much past data wants to be disregarded, enabling the handling of short-term dependency. The mathematical formulations for every GRU gate unit are formulated in Eqs. (8)- (11):

$$r_t = \sigma\ (W_r x_t + U_r h_{t-1}) \tag{8}$$

$$z_t = \sigma\ (W_z x_t + U_z h_{t-1}) \tag{9}$$

$$\widetilde{h_t} = tanh\,(W_h x_t + U_h\,(r_t \odot\ h_{t-1})) \tag{10}$$

$$h_t = z_t \odot\ \widetilde{h_t} + (1 - z_t) \odot\ h_{t-1} \tag{11}$$

Here, x_t signifies the input to the hidden layer (HL) at tth time ; h_t means the present output at tth time ; z_t and r_t signify the update gate and reset gate, correspondingly; W_z and U_z represent the weight coefficient for reset gate; $\widetilde{h_t}$ indicates the unit of candidate memory at tth time ; U_r and W_r signifies the weight coefficient for update gate; σ denotes an activation function. As a unidirectional structure of RNN, GRU naturally spreads states in a forward route. However, Bi-GRU holds dual GRU methods with reverse ways, permitting it to capture long-term dependencies and global data widely.

Parameter tuning: IChoA technique

To fine-tune the hyperparameter values of CNN-BiGRU model, the IChoA is utilized[32]. This model is chosen for its efficient search for optimal hyperparameters, giving a robust solution for complex optimization problems. Unlike conventional gradient-based methods, IChoA is a nature-inspired, population-based approach that doesn't depend on gradient information, making it appropriate for models with non-differentiable or highly non-convex objective functions. It effectually explores the solution space by replicating the social behaviour of

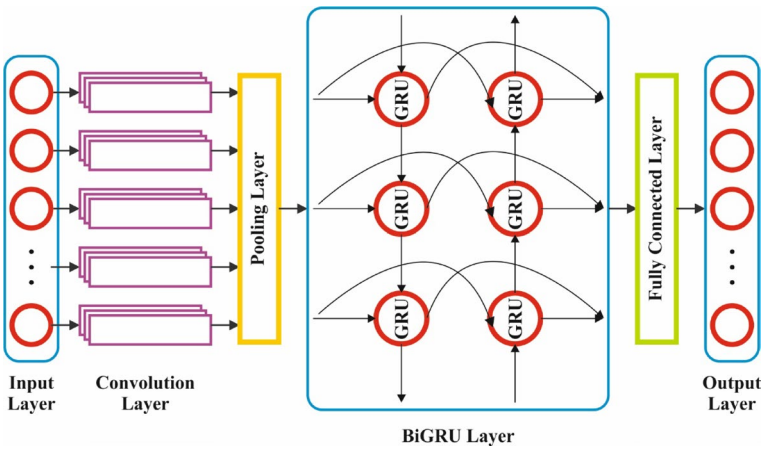

Fig. 3. Architecture of CNN-BiGRU model.

chimpanzees, ensuring better global optimization. Compared to other optimization techniques, such as grid search or genetic algorithms, IChoA presents faster convergence and avoids local optima, resulting in more accurate and efficient model performance. Its adaptability to various types of data and model structures makes it an ideal choice for improving the performance of IDS in resource-constrained IoT environments. Overall, IChoA improves both the accuracy and computational efficiency of the model. Figure 4 demonstrates the working flow of the IChoA approach.

The ChoA method stems from the chimp's hunting behaviour. The algorithm sets the chimp's locations and picks the four lowest-fitness chimps, such as Chaser, Attacker, Barrier, and Driver, signifying the top four optimum solutions: $(X_{attacker},\ X_{barrier},\ X_{chaser},\ and\ X_{driver})$.

In the phase of chimps chasing and driving prey, each chimp alters its location depending on the prey's position in the hunting procedure. It is mathematically formulated below:

$$\begin{cases} d = |c \cdot\ X_{prey}(t) - m \cdot\ X_{chimp}(t)| \\ X_{chimp}(t+1) = X_{prey}\,(t) - a \cdot\ d \\ a = f(2r_1 - 1) \\ c = 2r_2 \\ f = 2\left(1 - \left(\frac{t}{t_{\max}}\right)\right) \end{cases} \tag{12}$$

Correspondingly, t and $t_{\max}$ mean the present and maximum iteration count. d signifies the distance between the prey and the chimp. X_{prey} and X_{chimp} denote the locations of the prey and the chimp, respectively. r_l and r_2 represent randomly generated numbers within the interval of [0, 1]. f refers to a factor of convergence that linearly reduces from 2 to 0 throughout the iteration procedure. a and c denote the constant vectors, $and\ m$ refers to the chaotic vector.

Fig. 4. Working flow of the IChoA technique.

In an attack stage, chimps discover the prey's position and encircle it with the Chaser, Attacker, Barrier, and Driver. Then, it introduces a synchronized attack on the victim. It is mathematically given below:

$$\begin{cases} d_{attacker} = |c_1 \cdot X_{attacker} - m_1 \cdot X| \\ d_{barrier} = |c_2 \cdot X_{barrier} - m_2 \cdot X| \\ d_{chraser} = |c_3 \cdot X_{chaser} - m_3 \cdot X| \\ d_{driver} = |c_4 \cdot X_{driver} - m_4 \cdot X| \end{cases} \tag{13}$$

While X_1, X_2, X_3, and X_4 represent the upgraded location vectors of Attacker, Barrier, Chaser, and Driver, X denotes the location vector. The other chimp individual's locations are defined equally by four optimum chimp locations. $X(t+1)$ signifies their upgraded location vector, and the formulation is given below:

$$X(t+1) = \frac{X_1 + X_2 + X_3 + X_4}{4} \tag{14}$$

Generally, the ChOA experiences problems like a tendency to meet local goals and slow convergence velocity. The population was initialized utilizing Logistic mapping, and the separate chimp location upgrade model was enhanced. The Spiral functions were presented in the Choa, and this function improved the model's performance by developing the searching space and harmonizing global and local hunts. The IChOA has enhanced the ChOA from 4 dissimilar sizes.

SPM chaotic map

ChoA randomly implements the initialization of the population, which might lead to an uneven population spread and affect the model's performance. The chaotic map might efficiently produce a significantly expanded initial population. When equated to a standard chaotic map, the SPM chaotic map delivers a population with sturdier arbitrariness and even distribution. It is mathematically computed below:

$$x(t+1) = \begin{cases} mod(\frac{x(t)}{\eta} + \mu \sin(\pi\, x(t) + r, 1), 0 \le x(t) < \eta \\ mod(\frac{\frac{x(t)}{\eta}}{0.5-\eta} + \mu \sin(\pi\, x(t) + r, 1), \eta \le x(t) < 0.5 \\ mod(\frac{(1-x(t))}{\eta} + \mu \sin(\pi\ (t))) + r, 1), 0.5 \le x(t) < 1-\eta \\ mod(\frac{(1-x(t))}{\eta} + \mu \sin(\pi\ (1-x(t)) + r, 1), 1-\eta \le x(t) < 1 \end{cases} \tag{15}$$

Here, r and $x(t)$ represent randomly generated numbers 0 and 1.

Nonlinear convergence factor

The convergence factor f doesn't efficiently balance local exploitation and global exploration throughout population location upgrade. A slow declining convergence factor in the initial phases of model iteration permits the populations to discover the global optimum more efficiently. The quicker diminishing convergence factor advantages the technique in hunting for the local optimum solutions.

$$f = \cos\left(\pi\ \left(\frac{t}{t_{\max}}\right)\right) + 1 \tag{16}$$

T-distribution and opposition-based learning perturbation tactic

The attacker's location affects the accuracy and efficacy of the model's optimizer. However, dependency on this location can lead to other chimpanzees collecting everywhere, obstructing the search of different areas in the search space. T-distribution and opposition-based learning were employed to interrupt an attacker's location to evade convergence stagnation. Its mathematical formulation is given below:

$$X_{op}(t) = ub + r \oplus (lb - X_{attacker}(t)) \tag{17}$$

$$X_t(t) = \begin{cases} X_{attacker}(t) + X_{attacker}(t) \cdot t\_dis(t+1), & t \le \frac{t_{max}}{2} \\ X_{attacker}(t) + X_{attacker}(t) \cdot t\_dis\left(t^2\right), & t \le \frac{t_{max}}{2} \end{cases}$$

Here, X_{op} and X_t signify the locations after the attacker, ub and lb mean the searching space limits, r refers to a randomly generated vector among [0, 1], $and\ f(x)$ indicates the fitness value of location x. The function of t_dis is employed to create a random number that obeys a t-distribution. The t distribution looks like a Cauchy distribution, with a higher distribution of randomly generated numbers beneficial to global exploration. In future phases, the freedom grades increase more quickly, so the t-distribution approaches the Gaussian distribution with randomly generated values concentrated around the mean. If the chosen fitness value location is less than an attacker's, it signifies a higher solution. It is mathematically computed below:

$$X_{best}(t) = \begin{cases} X_{op}(t) & , f(X_{op}(t)) < f(X_t(t)) \\ X_t(t) & , f(X_{op}(t)) \ge f(X_t(t)) \end{cases} \tag{18}$$

$$X_{attacker}(t) = \begin{cases} X_{best}(t) & , f(X_{best}(t)) < f(X_{attacker}(t)) \\ X_{attacker}(t) & , f(X_{best}(t)) \ge f(X_{attacker}(t)) \end{cases} \tag{19}$$

The IChoA originates an FF to accomplish enhanced performance of classification. It defines a positive number to denote the better outcome of the candidate solution. The reduction of the classifier rate of error is measured as FF, as set in Eq. (20).

$$fitness\left(x_i\right) = ClassifierErrorRate\left(x_i\right) = \frac{no.\ of\ misclassified\ samples}{Total\ no.\ of\ samples} \times 100 \tag{20}$$

XAI process: SHAP

Finally, the SHAP is integrated to enhance interpretability, offering insights into model decisions for improved trust and reliability in cybersecurity. The SHAP method is one of the effective models for illustrating the forecasts of ML methods[33]. The values of SHAP depend upon the values of Shapley from game theory and were employed to allocate every participant's cooperation to the joint outcomes equally. The values of SHAP have numerous main properties, such as efficacy, additivity, and symmetry. These features certify that a characteristic contribution is equivalent to an alteration between the mean and the predicted value, and that data collection from a single method is comparable to the forecasts of every model combined. The mathematical formulation for computing the value of SHAP is given as follows:

$$y_i = y_{base} + f\left(x_{i1}\right) + f\left(x_{i2}\right) + \cdots + f\left(x_{ij}\right) \tag{21}$$

While y_{base} denotes an average value of the target variable, $f\left(x_{ij}\right)$ means a SHAP value of x_{ij}. The SHAP values formulation includes multiplying the minimal donation of every feature by the equivalent weight and totalling. This model considers the contribution of features in every sample and displays the effect's negativity and positivity. Here, the SHAP model depends on game theory to deduce and examine the technique. The model owns consistency and local accuracy, which can efficiently construe the outcomes of ML forecast methods.

Result analysis and discussion

The experimental evaluation of the XAICR-HDLOA approach is examined under the Edge-IIoT dataset[34]. The dataset encompasses 56,000 samples with 12 classes as defined in Table 2.

Dataset features overview with a comprehensive list of 62 features and a highlighted selection of 47 key attributes for analysis

The dataset comprises a total of 62 features, including frame.time, ip.src_host, ip.dst_host, arp.dst.proto_ipv4, arp.opcode, arp.hw.size, arp.src.proto_ipv4, icmp.checksum, icmp.seq_le, icmp.transmit_timestamp, icmp.unused, http.file_data, http.content_length, http.request.uri.query, http.request.method, http.referer, http.request.full_uri, http.request.version, http.response, http.tls_port, tcp.ack, tcp.ack_raw, tcp.checksum, tcp.connection.fin, tcp.connection.rst, tcp.connection.syn, tcp.connection.synack, tcp.dstport, tcp.flags, tcp.flags.ack, tcp.len, tcp.options, tcp.payload, tcp.seq, tcp.srcport, udp.port, udp.stream, udp.time_delta, dns.qry.name, dns.qry.name.len, dns.qry.qu, dns.qry.type, dns.retransmission, dns.retransmit_request, dns.retransmit_request_in, mqtt.conack.flags, mqtt.conflag.cleansess, mqtt.conflags, mqtt.hdrflags, mqtt.len, mqtt.msg_decoded_as, mqtt.msg, mqtt.msgtype, mqtt.proto_len, mqtt.protoname, mqtt.topic, mqtt.topic_len, mqtt.ver, mbtcp.len, mbtcp.trans_id, mbtcp.unit_id, Attack_label, and Attack_type.

From the overall features, 47 key features were selected for analysis, such as frame.time, arp.opcode, arp.hw.size, icmp.checksum, icmp.seq_le, icmp.transmit_timestamp, icmp.unused, http.content_length, http.request.method, http.referer, http.request.version, http.response, http.tls_port, tcp.ack, tcp.checksum, tcp.connection.fin, tcp.connection.rst, tcp.connection.syn, tcp.connection.synack, tcp.dstport, tcp.flags, tcp.len,

Edge-IIoT dataset	
Type of Event	**Data Record**
"Normal"	5000
"DDoS-UDP"	5000
"DDoS-ICMP"	5000
"SQL injection"	5000
"DDoS-TCP"	5000
"Password"	5000
"DDoS-HTTP"	5000
"Uploading"	5000
"Backdoor"	5000
"XSS"	5000
"Ransomware"	3000
"Fingerpriniting"	3000
Total Record	**56,000**

Table 2. Details on Edge-IIoT dataset.

tcp.seq, tcp.srcport, udp.port, udp.stream, udp.time_delta, dns.qry.name.len, dns.qry.qu, dns.qry.type, dns.retransmission, dns.retransmit_request, dns.retransmit_request_in, mqtt.conack.flags, mqtt.conflag.cleansess, mqtt.conflags, mqtt.hdrflags, mqtt.len, mqtt.msgtype, mqtt.proto_len, mqtt.protoname, mqtt.topic_len, mqtt.ver, mbtcp.len, mbtcp.trans_id, mbtcp.unit_id, and Attack_type. The chosen features represent critical protocol-specific parameters across layers, capturing diverse behavioral patterns crucial for distinguishing benign and malicious traffic. This multidimensional selection ensures comprehensive coverage of temporal, structural, and content-based indicators relevant to intrusion detection.

Analysis and results highlighting key findings and performance evaluation

Figure 5 displays the classifier performances of the XAICR-HDLOA approach on the Edge-IIoT dataset. Figure 5a and b exhibits the confusion matrices through precise identification and classification of all 12 class labels on a 70% of training set (TRASE) and 30% of testing set (TESSE). Figure 5c illustrates the PR study, which enhanced performance over 12 classes. Finally, Fig. 5d demonstrates the ROC outcome, illustrating capable solutions with great ROC values for 12 distinct classes.

Table 3; Fig. 6 indicate the overall attack detection results of the XAICR-HDLOA approach under the Edge-IIoT dataset with 70% TRASE and 30% TESSE. The performances exemplify that the XAICR-HDLOA approach suitably acknowledged varied class labels. On 70%TRASE, the XAICR-HDLOA approach presents an average $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score}, and Kappa of 98.37%, 90.13%, 89.69%, 89.87%, 94.40%, and 94.47%, correspondingly. Followed by, based on 30%TESSE, the XAICR-HDLOA approach provides an average $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score}, and Kappa of 98.41%, 90.42%, 90.01%, 90.19%, 94.57%, and 94.64%, respectively.

Figure 7 depicts the TRA $accu_y$ (TRAAY) and validation $accu_y$ (VLAAY) performances of the XAICR-HDLOA approach below the Edge-IIoT dataset. The values of $accu_y$ are computed across a period of 0–25

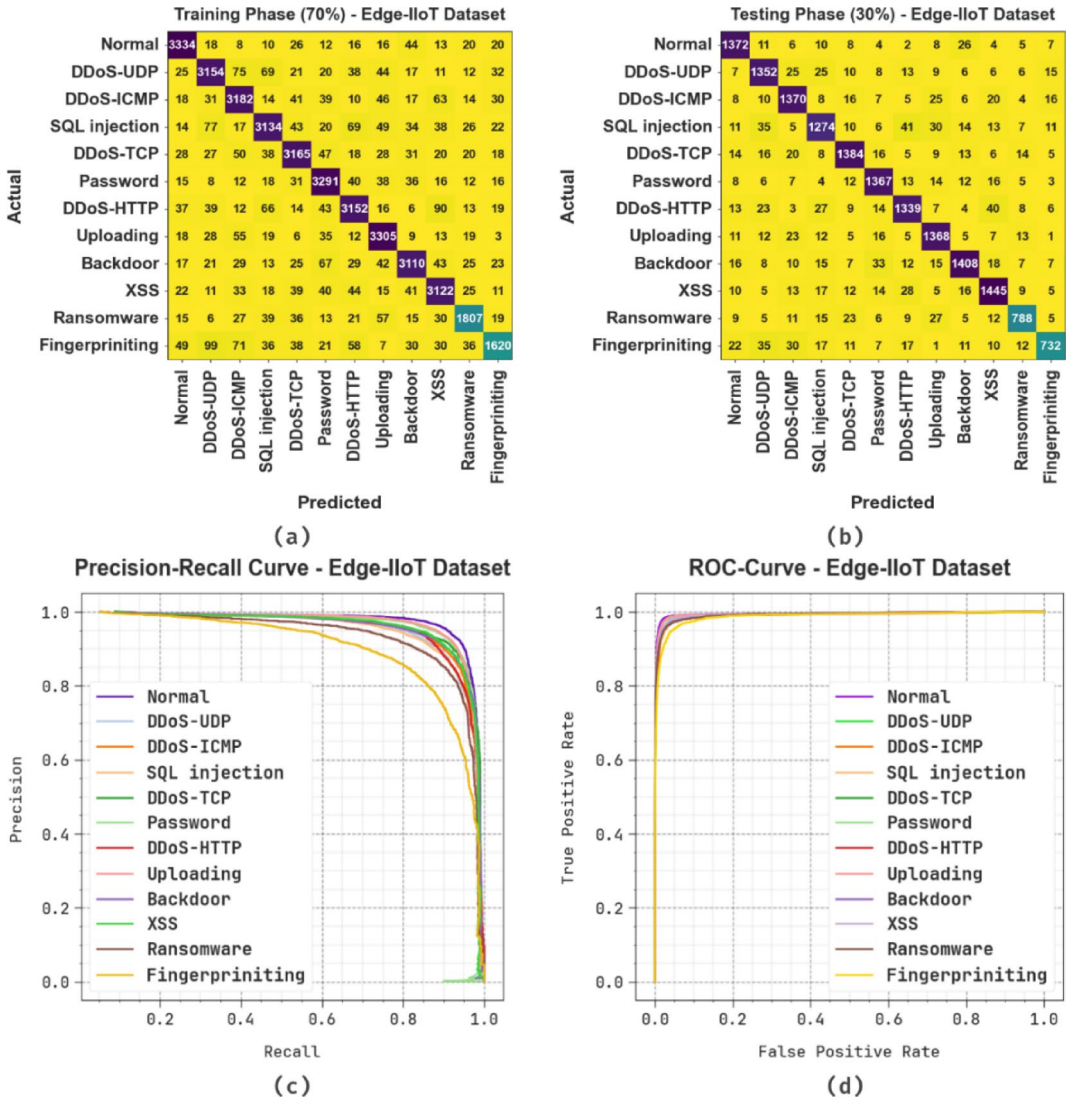

Fig. 5. Edge-IIoT dataset **(a-b)** confusion matrices and **(c-d)** curves of PR and ROC.

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$	AUC_{Score}	Kappa
TRASE (70%)						
Normal	98.82	92.82	94.26	93.53	96.77	96.84
DDoS-UDP	98.14	89.63	89.65	89.64	94.32	94.39
DDoS-ICMP	98.18	89.11	90.78	89.94	94.85	94.92
SQL injection	98.09	90.21	88.46	89.33	93.75	93.81
DDoS-TCP	98.35	90.82	90.69	90.75	94.90	94.95
Password	98.47	90.21	93.15	91.66	96.07	96.13
DDoS-HTTP	98.19	89.88	89.88	89.88	94.44	94.51
Uploading	98.53	90.23	93.84	92.00	96.42	96.49
Backdoor	98.43	91.74	90.30	91.02	94.76	94.84
XSS	98.30	89.48	91.26	90.36	95.12	95.19
Ransomware	98.72	89.06	86.67	87.85	93.03	93.09
Fingerpriniting	98.24	88.38	77.33	82.48	88.38	88.45
Average	**98.37**	**90.13**	**89.69**	**89.87**	**94.40**	**94.47**
TESSE (30%)						
Normal	98.69	91.41	93.78	92.58	96.47	96.53
DDoS-UDP	98.24	89.06	91.23	90.13	95.07	95.13
DDoS-ICMP	98.35	89.95	91.64	90.79	95.32	95.39
SQL injection	97.97	88.97	87.44	88.20	93.21	93.28
DDoS-TCP	98.52	91.84	91.66	91.75	95.43	95.50
Password	98.62	91.26	93.18	92.21	96.16	96.23
DDoS-HTTP	98.19	89.93	89.69	89.81	94.35	94.41
Uploading	98.45	90.12	92.56	91.32	95.79	95.85
Backdoor	98.42	92.27	90.49	91.37	94.86	94.92
XSS	98.30	90.48	91.51	90.99	95.26	95.34
Ransomware	98.71	89.75	86.12	87.90	92.78	92.84
Fingerpriniting	98.49	90.04	80.88	85.22	90.19	90.25
Average	**98.41**	**90.42**	**90.01**	**90.19**	**94.57**	**94.64**

Table 3. Attack detection outcome of the XAICR-HDLOA approach under the Edge-IIoT dataset.

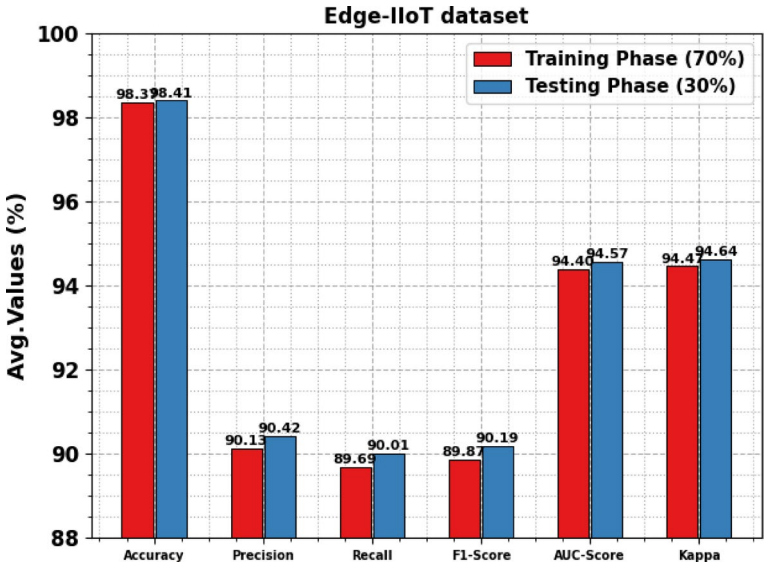

Fig. 6. Average outcome of XAICR-HDLOA approach under the Edge-IIoT dataset.

Fig. 7. $Accu_y$ curve of XAICR-HDLOA approach under the Edge-IIoT dataset.

Fig. 8. Loss curve of XAICR-HDLOA approach under the Edge-IIoT dataset.

epochs. The figure underscored that the values of TRAAY and VLAAY present a growing tendency to indicate the proficiency of the XAICR-HDLOA approach with higher performance across numerous repetitions. In addition, the TRAAY and VLAAY values remain close through the epochs, indicating diminishing overfitting and expressing superior performance of the XAICR-HDLOA approach, which guarantees steady calculation on unseen samples.

Figure 8 shows the TRA loss (TRALO) and VLA loss (VLALO) graph of the XAICR-HDLOA approach below the Edge-IIoT dataset. The loss values are computed throughout 0–25 epochs. The values of TRALO and VLALO demonstrate a diminishing tendency, which indicates the proficiency of the XAICR-HDLOA approach in corresponding to a tradeoff between data fitting and generalization. The successive dilution in loss values also assures the superior performance of the XAICR-HDLOA approach and tunes the calculation results over time.

In Table 4; Fig. 9, a detailed comparison analysis of the XAICR-HDLOA approach is reported. The performances demonstrated that the SVM, RF, FFNN, MLP, and K-NN models have shown ineffectual detection results with the least $accu_y$ of 92.31%, 92.91%, 93.60%, 94.73%, and 95.14%, respectively. In the meantime, the CNN model has exhibited considerable performance with $accu_y$ of 96.84%, $prec_n$ of 86.74%, $reca_l$ of 87.19%, and $F1_{score}$ of 83.64%. Furthermore, the XGBoost model has accomplished reasonable outcomes with $accu_y$

Edge-IIoT dataset				
Technique	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
RF	92.91	88.55	83.77	84.12
K-NN Algorithm	95.14	85.16	84.47	85.49
CNN Classifier	96.84	86.74	87.19	83.64
XGBoost	97.09	83.80	85.96	85.29
FFNN Method	93.60	87.14	89.04	84.57
MLP Model	94.73	87.67	84.42	85.39
SVM Method	92.31	83.37	89.86	83.01
XAICR-HDLOA	98.41	90.42	90.01	90.19

Table 4. Comparative outcome of XAICR-HDLOA approach under the Edge-IIoT dataset with existing models.

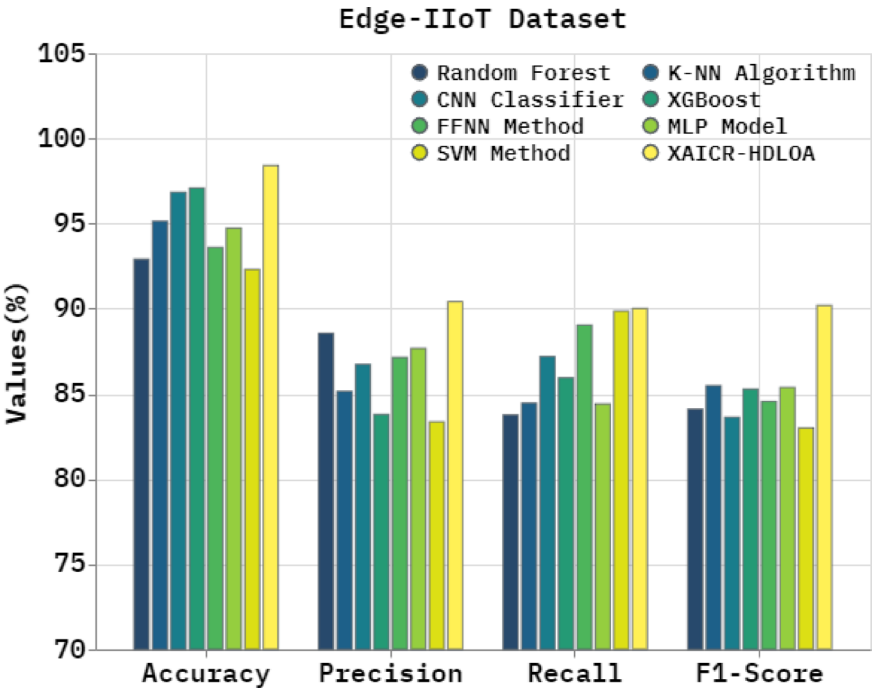

Fig. 9. Comparative outcome of XAICR-HDLOA approach under the Edge-IIoT dataset with existing models.

of 97.09%, $prec_n$ of 83.80%, $reca_l$ of 85.96%, and $F1_{score}$ of 85.29%. Finally, the XAICR-HDLOA approach demonstrates superior performance with an increased $accu_y$ of 98.41%, $prec_n$ of 90.42%, $reca_l$ of 90.01%, and $F1_{score}$ approach of 90.19%.

Table 5; Fig. 10 illustrate the computational time (CT) analysis of the XAICR-HDLOA approach with existing techniques under the Edge-IIoT dataset. The RF model takes 5.42 s, the kNN technique requires 7.16 s, and the CNN classifier needs 6.62 s for completion. Other techniques such as XGBoost, with a CT of 7.98 s, and the FFNN method, which takes 4.78 s, illustrate varying time requirements. The MLP model has a CT of 6.39 s, while the SVM method processes in 6.22 s. The XAICR-HDLOA approach outperforms with the lowest CT of 3.48 s, indicating its efficiency in processing tasks compared to the other methods. This reduced CT is crucial for real-time intrusion detection in time-sensitive IIoT applications, enhancing overall system responsiveness and scalability.

Table 6; Fig. 11 indicates the ablation analysis of the XAICR-HDLOA methodology under the Edge-IIoT dataset. The BES model achieved an $accu_y$ of 96.86%, $prec_n$ of 88.50%, $reca_l$ of 88.31%, and $F1_{score}$ of 88.44%, while IChoA slightly improved the results with an $accu_y$ of 97.37%, $prec_n$ of 89.05%, $reca_l$ of 88.82%, and $F1_{score}$ of 89.04%. Further enhancement was seen in the CNN-BiGRU model, which reached an $accu_y$ of 97.88%, $prec_n$ of 89.82%, $reca_l$ of 89.44%, and $F1_{score}$ of 89.54%. The XAICR-HDLOA approach outperformed all baselines, delivering an $accu_y$ of 98.41%, $prec_n$ of 90.42%, $reca_l$ of 90.01%, and $F1_{score}$ of 90.19%, highlighting its superior capability in learning complex patterns and improve classification performance.

Table 7 specifies the computational efficiency of the XAICR-HDLOA methodology under the Edge-IIoT dataset. The XAICR-HDLOA methodology indicated the lowest computational cost with 10.77 GFLOPs and the

Edge-IIoT dataset	
Technique	**CT (sec)**
RF	5.42
K-NN Algorithm	7.16
CNN Classifier	6.62
XGBoost	7.98
FFNN Method	4.78
MLP Model	6.39
SVM Method	6.22
XAICR-HDLOA	3.48

Table 5. CT analysis of XAICR-HDLOA approach under the Edge-IIoT dataset over existing techniques.

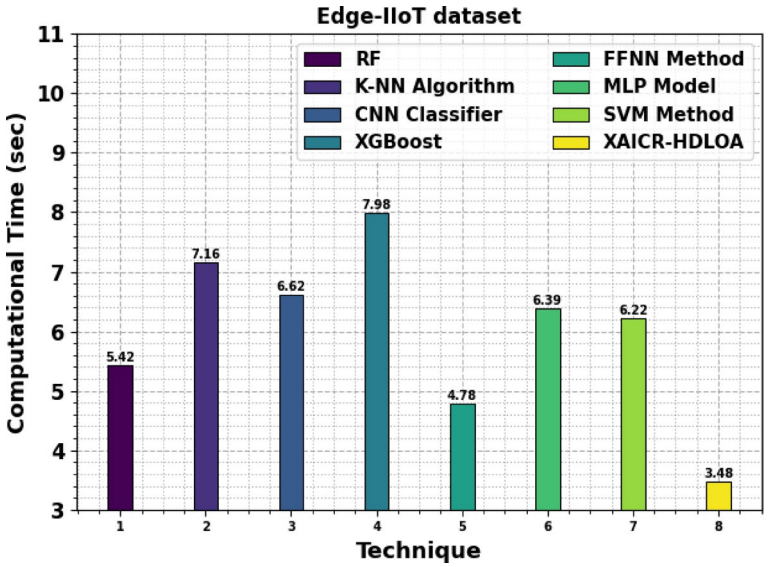

Fig. 10. CT analysis of XAICR-HDLOA approach under the Edge-IIoT dataset over existing techniques.

Edge-IIoT dataset				
Technique	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
BES	96.86	88.50	88.31	88.44
IChoA	97.37	89.05	88.82	89.04
CNN-BiGRU	97.88	89.82	89.44	89.54
XAICR-HDLOA	98.41	90.42	90.01	90.19

Table 6. Ablation study results comparing XAICR-HDLOA method under the Edge-IIoT dataset over existing techniques.

least memory usage at 1061 MB, highlighting its suitability for resource-constrained environments. In contrast, EfficientNet required 24.8 GFLOPs and 1354 MB, while MobileNetV2 and ShuffleNet consumed 23.42 GFLOPs and 2874 MB, and 18.53 GFLOPs and 3015 MB, respectively. GhostNet and MnasNet also illustrated higher demands, with 19.21 GFLOPs and 1876 MB, and 23.41 GFLOPs and 1881 MB. Although CGENet exhibited moderate efficiency at 22.97 GFLOPs and 1507 MB, the XAICR-HDLOA model clearly outperforms all others in both computational load and memory footprint, making it ideal for real-time IoT-based IDS.

Overview of dataset features including 35 attributes with a focused selection of 8 key features for analysis

The experimental validation of the XAICR-HDLOA approach is examined under the BoT-IoT dataset[35]. The dataset consists of 2056 samples with 5 class labels as defined in Table 8.

The dataset comprises a total of 30 features including pkSeqID, stime, flgs, proto, saddr, sport, daddr, dport, pkts, bytes, state, ltime, seq, dur, mean, stddev, smac, dmac, sum, min, max, soui, doui, sco, dco, spkts, dpkts,

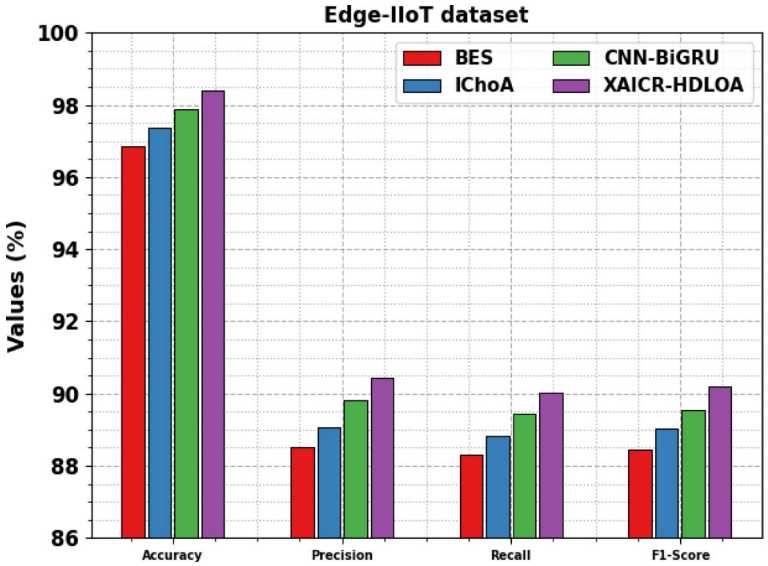

Fig. 11. Ablation study results comparing XAICR-HDLOA method under the Edge-IIoT dataset over existing techniques.

BoT-IoT dataset		
Model	**FLOPs (G)**	**GPU (M)**
EfficientNet	24.8	1354
GhostNet	19.21	1876
MnasNet	23.41	1881
MobileNetV2	23.42	2874
ShuffleNet	18.53	3015
CGENet	22.97	1507
XAICR-HDLOA	10.77	1061

Table 7. Comparison of computational efficiency and memory usage of XAICR-HDLOA technique under the Edge-IIoT dataset.

BoT-IoT dataset	
Classes	**No. of Samples**
"DDoS"	500
"DoS"	500
"Recon"	500
"Theft"	79
"Normal"	477
Total	**2056**

Table 8. Details on BoT-IoT dataset.

sbytes, dbytes, rate, srate, drate, attack, category, and subcategory. Out of these, the eight key features proto, saddr, sport, daddr, dport, pkts, bytes, and attack have been carefully chosen to ensure effectual evaluation and accurate detection. The features are chosen for their direct relevance to traffic flow dynamics and classification. These attributes effectually capture source-destination behavior, protocol type, and volumetric patterns, which are significant in detecting and differentiating attack signatures.

Findings and performance evaluation emphasizing key analytical results

Figure 12 shows the classifier performances of the XAICR-HDLOA approach on BoT-IoT dataset. Figure 12a and b displays the confusion matrices through specific identification and classification of all 5 class labels on a 70%TRASE and 30%TESSE. Figure 12c exhibits the PR examination, which showed lower performance over five

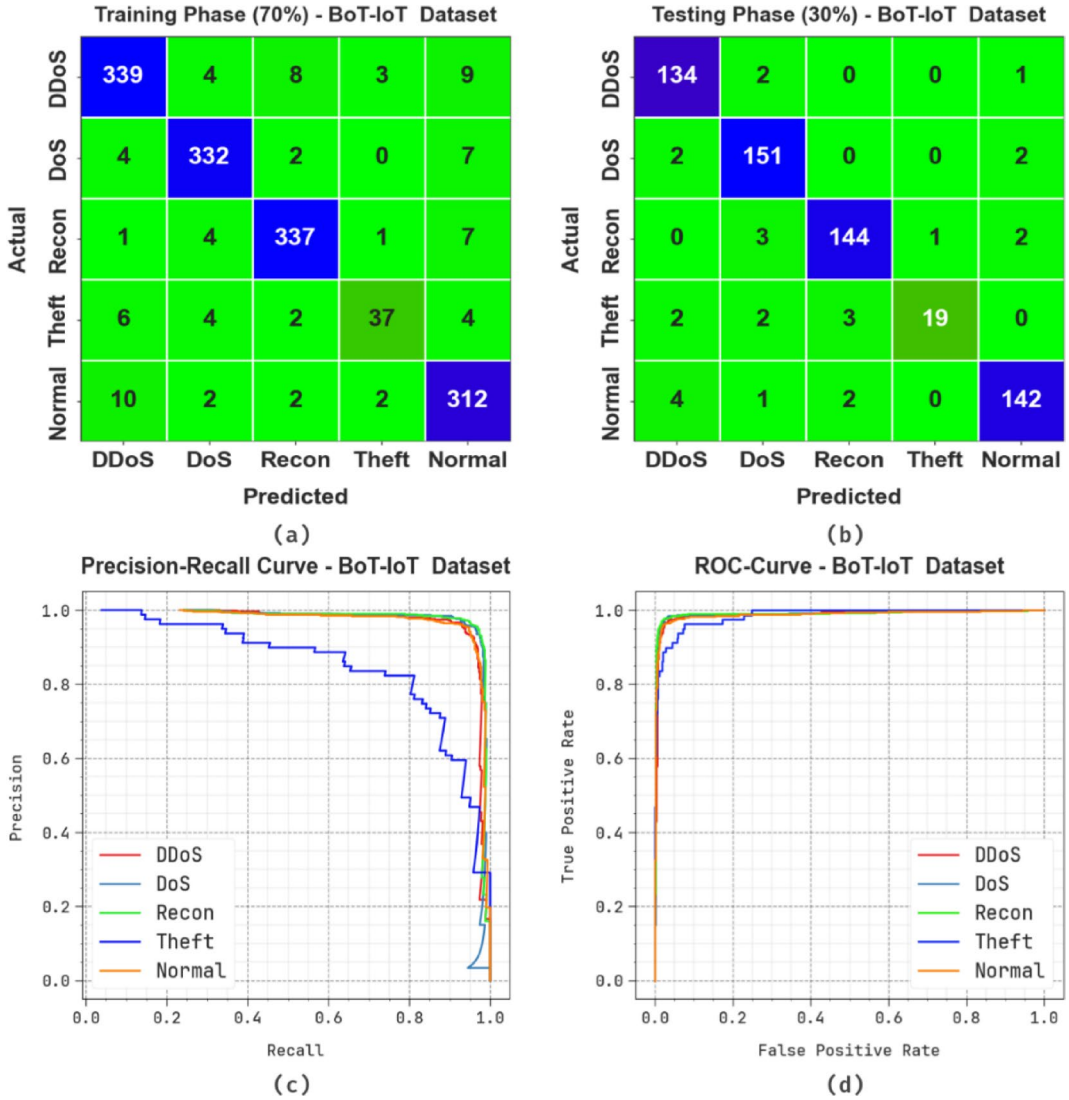

Fig. 12. BoT-IoT dataset **(a-b)** confusion matrices and **(c-d)** curves of PR and ROC.

classes. Eventually, Fig. 12d signifies the ROC study, which represents a skilful solution with great ROC values for five different classes.

Table 9; Fig. 13 imply an attack detection solution of the XAICR-HDLOA approach below the Bot-IoT dataset using 70% TRASE and 30% TESSE. The performances suggest that the XAICR-HDLOA approach can accurately recognize different classes. Based on 70%TRASE, the XAICR-HDLOA approach presents an average $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score}, and Kappa of 97.72%, 92.84%, 90.17%, 91.33%, 94.35%, and 94.41%, respectively. Afterwards, using 30%TESSE, the XAICR-HDLOA approach presents an average $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score}, and Kappa of 98.25%, 95.52%, 91.92%, 93.42%, 95.39%, and 95.46%, subsequently.

Figure 14 depicts the TRAAY and VLAAY performances of the XAICR-HDLOA approach below BoT-IoT dataset. The values of $accu_y$ are computed through a period of 0–25 epochs. The figure underscored that the values of TRAAY and VLAAY present a cumulative tendency indicating the competency of the XAICR-HDLOA approach with higher performance through multiple repetitions. Moreover, the TRAAY and VLAAY values remain close across the epochs, indicating decreased overfitting and maximum performance of the XAICR-HDLOA approach, ensuring reliable calculation on unnoticed samples.

Figure 15 demonstrates the TRALO and VLALO graphs of the XAICR-HDLOA approach below the Bot-IoT dataset. The loss values are computed across a period of 0–25 epochs. The values of TRALO and VLALO represent a declining tendency, indicating the capacity of the XAICR-HDLOA approach to equalize a tradeoff between data fitting and generalization. The succeeding dilution in values of loss and securities improves the performance of the XAICR-HDLOA approach and tunes the calculation results gradually.

Table 10; Fig. 16 show a detailed comparison of the XAICR-HDLOA approach[36–41]. The performances illustrated that the GANs + AE, Decision Tree (DT), NB, Bi-LSTM, and GBC models displayed inefficient detection solutions with minimum $accu_y$ of 87.61%, 89.14%, 89.60%, 90.80%, and 94.29%, respectively. Meanwhile, the GSOM technique exhibited substantial outcome with $accu_y$ of 92.21%, $prec_n$ of 94.15%, $reca_l$

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$	AUC_{Score}	Kappa
TRASE (70%)						
DDoS	96.87	94.17	93.39	93.78	95.72	95.77
DoS	98.12	95.95	96.23	96.09	97.48	97.54
Recon	98.12	96.01	96.29	96.15	97.50	97.56
Theft	98.47	86.05	69.81	77.08	84.69	84.77
Normal	97.01	92.04	95.12	93.55	96.35	96.40
Average	**97.72**	**92.84**	**90.17**	**91.33**	**94.35**	**94.41**
TESSE (30%)						
DDoS	98.22	94.37	97.81	96.06	98.07	98.14
DoS	98.06	94.97	97.42	96.18	97.84	97.90
Recon	98.22	96.64	96.00	96.32	97.46	97.53
Theft	98.70	95.00	73.08	82.61	86.45	86.53
Normal	98.06	96.60	95.30	95.95	97.12	97.20
Average	**98.25**	**95.52**	**91.92**	**93.42**	**95.39**	**95.46**

Table 9. Attack detection outcome of XAICR-HDLOA approach under BoT-IoT dataset.

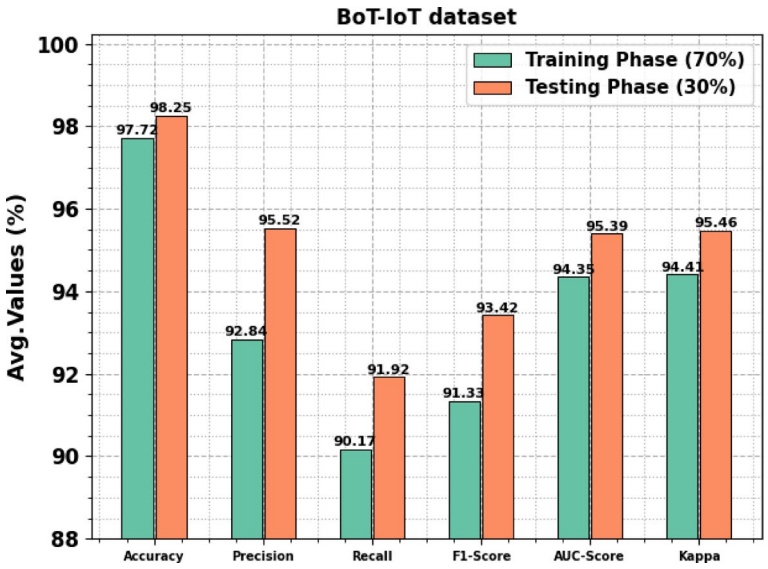

Fig. 13. Average outcome of XAICR-HDLOA approach under BoT-IoT dataset.

of 86.71%, and $F1_{score}$ of 90.44%. In addition, the CGANs + FNN technique has obtained judicious models with $accu_y$ of 97.98%, $prec_n$ of 91.00%, $reca_l$ of 87.13%, and $F1_{score}$ of 87.51%. Lastly, the XAICR-HDLOA approach represents maximum performance with superior $accu_y$ of 98.25%, $prec_n$ of 95.52%, $reca_l$ of 91.92%, and $F1_{score}$ of 93.42%. Hence, the XAICR-HDLOA approach is applied for improved cyber resilience in the IoT environment.

Table 11; Fig. 17 demonstrates the CT analysis of XAICR-HDLOA approach with existing models. The GBC model requires a CT of 13.84 s, while the DT technique takes 9.97 s. The GSOM method requires 10.44 s, and the GAN integrated with autoencoders method needs a CT of 10.09 s. The CGANs + FNN methodology acquired 13.72 s, while the NB method requires 10.51 s. The Bi-LSTM method has a computation time of of 9.55 s. Finally, the XAICR-HDLOA approach outperforms with a CT of 5.00 s, highlighting its efficiency compared to the existing models.

Table 12; Fig. 18 specifies the ablation study of the XAICR-HDLOA technique under the BoT-IoT dataset. The XAICR-HDLOA technique attains an $accu_y$ of 98.25%, $prec_n$ of 95.52%, $reca_l$ of 91.92%, and $F1_{score}$ of 93.42%. In comparison, the BES method achieves an $accu_y$ of 96.41%, $prec_n$ of 93.66%, $reca_l$ of 90.02%, and $F1_{score}$ of 91.44%, while IChoA shows an $accu_y$ of 97.15%, $prec_n$ of 94.35%, $reca_l$ of 90.63%, and $F1_{score}$ of 92.21%. The CNN-BiGRU model achieves an $accu_y$ of 97.73%, $prec_n$ of 94.98%, $reca_l$ of 91.23%, and $F1_{score}$ of 92.91%. The consistently higher values of the XAICR-HDLOA method highlight its efficiency in improving classification accuracy and robustness for intrusion detection.

Table 13 demonstrates the superior computational efficiency of the XAICR-HDLOA approach under the BoT-IoT dataset. The EfficientNet model consumes 24.8G FLOPs and 1354 M GPU memory, while GhostNet,

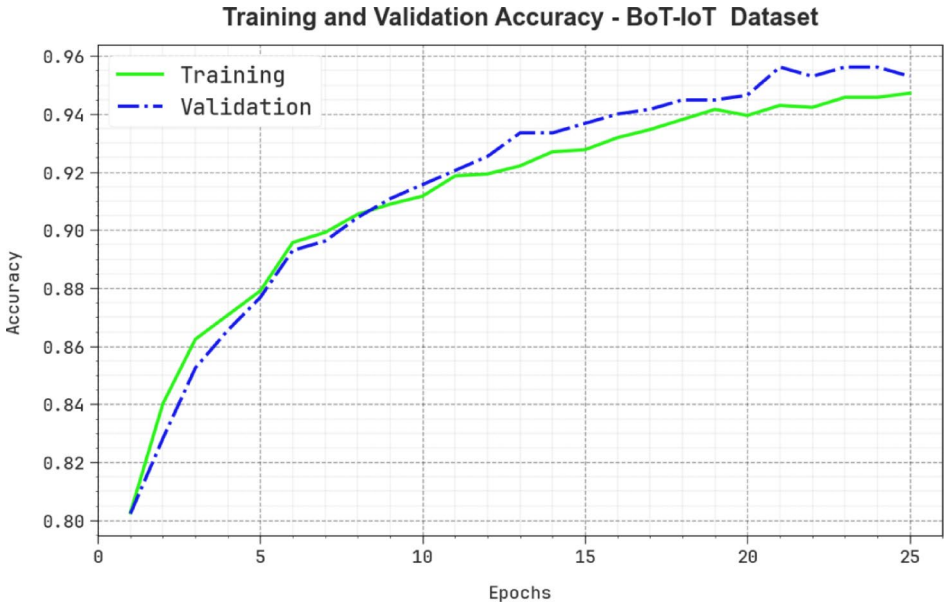

Fig. 14. $Accu_y$ curve of XAICR-HDLOA approach under BoT-IoT dataset.

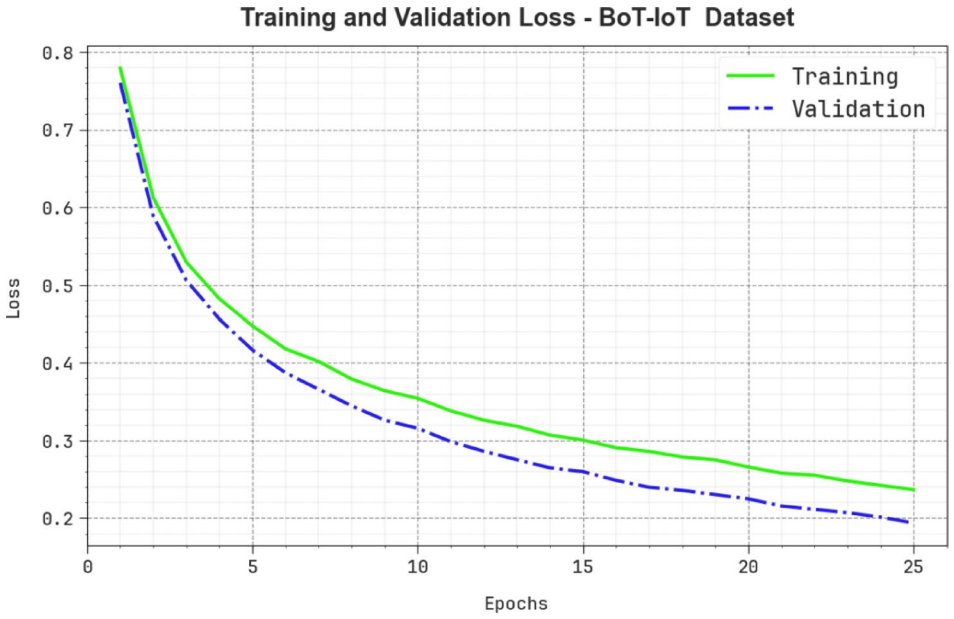

Fig. 15. Loss curve of XAICR-HDLOA approach under BoT-IoT dataset.

MnasNet, and MobileNetV2 require 19.21G, 23.41G, and 23.42G FLOPs with 1876 M, 1881 M, and 2874 M GPU memory, respectively. ShuffleNet, despite having the lowest FLOPs among the compared models at 18.53G, uses the highest GPU memory at 3015 M. CGENet reports 22.97G FLOPs and 1507 M GPU usage. But the XAICR-HDLOA model illustrates the lowest FLOPs of 10.77G and minimal GPU memory usage of 1061 M. The significantly lower resource demands of the XAICR-HDLOA model highlight its suitability for deployment in resource-constrained edge and IoT environments without losing performance.

Conclusion

In this paper, the XAICR-HDLOA approach is proposed. The main objective of the XAICR-HDLOA approach is to improve cyber threat detection and interpretation in IoT environments. To accomplish this, the XAICR-HDLOA approach applies the min-max normalization approach to standardize feature scales during the data normalization. The BES model selects the most relevant features for dimensionality reduction. Moreover, the hybrid of CNN-BiGRU model is used for the cyberattack classification. To fine-tune the hyperparameter values

BoT-IoT dataset				
Technique	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
GBC Model	94.29	91.85	88.63	87.75
DT	89.14	91.55	88.10	84.72
GSOM Method	92.21	94.15	86.71	90.44
GANs + AE	87.61	87.90	88.84	90.13
CGANs + FNN	97.98	91.00	87.13	87.51
NB	89.60	89.88	89.84	88.41
Bi-LSTM	90.80	87.70	87.13	85.95
XAICR-HDLOA	98.25	95.52	91.92	93.42

Table 10. Comparative outcome of XAICR-HDLOA approach under BoT-IoT dataset with existing models.

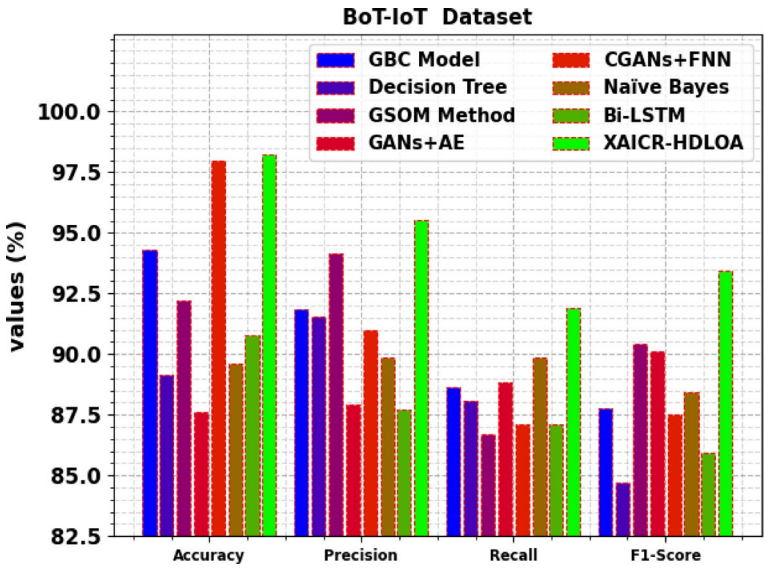

Fig. 16. Comparative outcome of XAICR-HDLOA approach under BoT-IoT dataset with existing models.

BoT-IoT dataset	
Technique	CT (sec)
GBC Model	13.84
DT	9.97
GSOM Method	10.44
GANs + AE	10.09
CGANs + FNN	13.72
NB	10.51
Bi-LSTM	9.55
XAICR-HDLOA	5.00

Table 11. CT analysis of XAICR-HDLOA approach under BoT-IoT dataset with existing models.

of CNN-BiGRU model, the IChoA is utilized. Finally, the SHAP is integrated to enhance interpretability, offering insights into model decisions for improved trust and reliability in cybersecurity. A wide range of simulations is performed to ensure the significance of the XAICR-HDLOA approach under the Edge-IIoT dataset. The performance validation of the XAICR-HDLOA approach portrayed a superior accuracy value of 98.41% and 98.25% over existing techniques under the Edge-IIoT and BoT-IoT datasets. The limitations of the XAICR-HDLOA approach include reliance on specific datasets, which may not fully capture the diversity of real-world IoT environments and cyberattacks. Furthermore, the model's performance may degrade in extremely resource-constrained devices with restricted computational power and memory. The study also assumes that all IoT devices are equally secure, which may not reflect real-world vulnerabilities. Future work may explore the

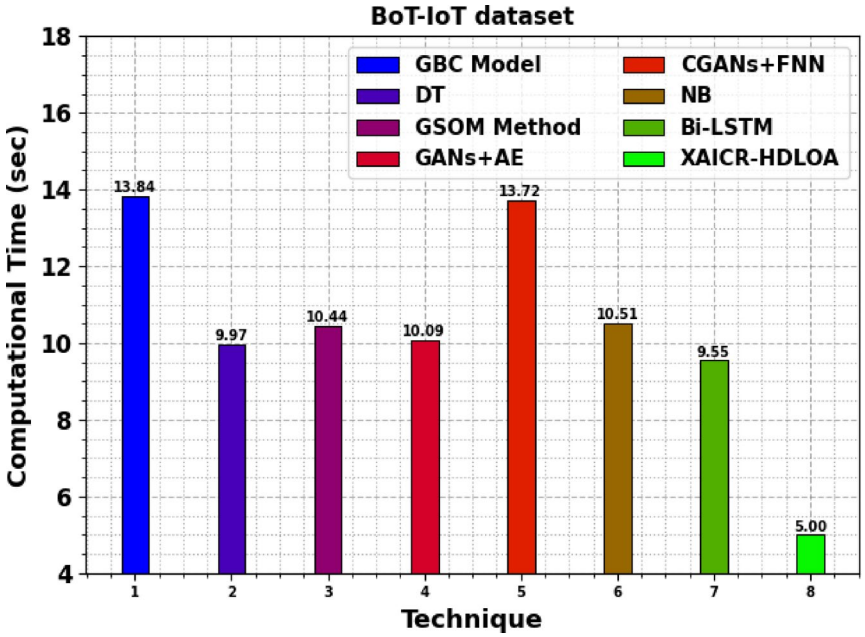

Fig. 17. CT analysis of XAICR-HDLOA approach under BoT-IoT dataset with existing models.

BoT-IoT dataset				
Technique	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
BES	96.41	93.66	90.02	91.44
IChoA	97.15	94.35	90.63	92.21
CNN-BiGRU	97.73	94.98	91.23	92.91
XAICR-HDLOA	98.25	95.52	91.92	93.42

Table 12. Ablation study-based comparative analysis of the XAICR-HDLOA technique under the BoT-IoT dataset.

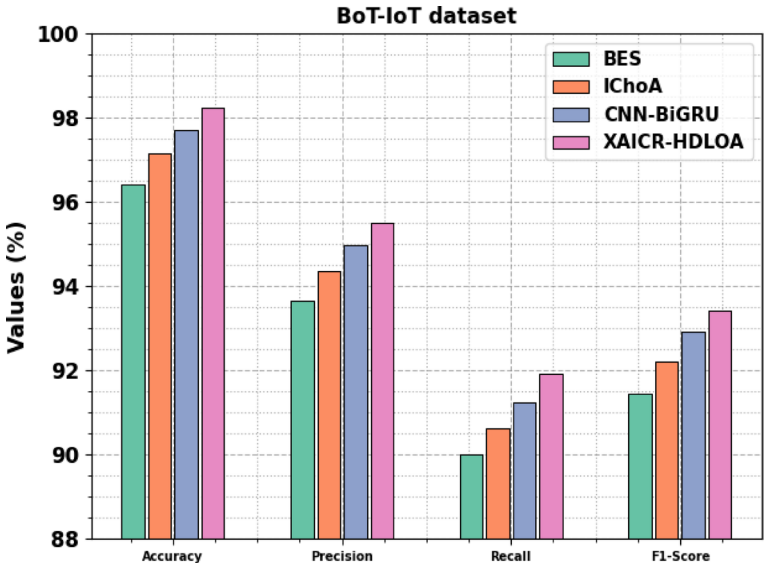

Fig. 18. Ablation study-based comparative analysis of the XAICR-HDLOA technique under the BoT-IoT dataset.

BoT-IoT dataset		
Model	**FLOPs (G)**	**GPU (M)**
EfficientNet	24.8	1354
GhostNet	19.21	1876
MnasNet	23.41	1881
MobileNetV2	23.42	2874
ShuffleNet	18.53	3015
CGENet	22.97	1507
XAICR-HDLOA	10.77	1061

Table 13. Performance comparison of XAICR-HDLOA technique in terms of computational cost and memory utilization on the BoT-IoT dataset.

application of the model to larger, more diverse datasets and investigate its performance in more heterogeneous IoT environments. Further study may optimize the model for edge devices with lesser computational resources. Moreover, integrating more advanced anomaly detection models and real-time adaptation could improve the technique's effectiveness in growing cybersecurity threats. Finally, assessing the robustness of the model against adversarial attacks remains an area for future research.

Data availability

The data supporting this study's findings are openly available at https://www.kaggle.com/datasets/mohamedamineferrag/edgeiiotset-cyber-security-dataset-of-iot-iiot and https://research.unsw.edu.au/projects/bot-iot-dataset, reference numbers[34,35].

Received: 6 February 2025; Accepted: 6 August 2025
Published online: 26 September 2025

References

1. Mishra, S. The impact of AI-based cyber security on the banking and financial sectors. *J. Cybersecurity Inform. Manag.* **14**(1), 8–19 (2024).
2. Djenouri, Y., Belhadi, A., Srivastava, G. & Lin, J. C. W. When explainable AI Meets IoT applications for supervised learning. *Cluster Comput.* **26** (4), 2313–2323 (2023).
3. Rjoub, G. et al. A survey on explainable artificial intelligence for cybersecurity. *IEEE Trans. Netw. Serv. Manage.* **20** (4), 5115–5140 (2023).
4. Sleem, A. Intelligent and secure detection of Cyber-attacks in industrial internet of things: A federated learning framework. *Full Length Article*. **7** (1), 51–51 (2023).
5. Arisdakessian, S., Wahab, O. A., Mourad, A., Otrok, H. & Guizani, M. A survey on IoT intrusion detection: federated learning, game theory, social psychology, and explainable AI as future directions. *IEEE Internet Things J.* **10** (5), 4059–4092 (2022).
6. Oseni, A. et al. An explainable deep learning framework for resilient intrusion detection in IoT-enabled transportation networks. *IEEE Trans. Intell. Transp. Syst.* **24** (1), 1000–1014 (2022).
7. Jagatheesaperumal, S. K. et al. Explainable AI over the internet of things (IoT): overview, state-of-the-art, and future directions. *IEEE Open. J. Commun. Soc.* **3**, 2106–2136 (2022).
8. Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y. & Taher, F. Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access.* **10**, 93104–93139 (2022).
9. Capuano, N., Fenza, G., Loia, V. & Stanzione, C. Explainable artificial intelligence in cybersecurity: A survey. *Ieee Access.* **10**, 93575–93600 (2022).
10. Zolanvari, M., Yang, Z., Khan, K., Jain, R. & Meskin, N. Trust xai: Model-agnostic explanations for Ai with a case study on Iiot security. *IEEE Internet Things J.* **10** (4), 2967–2978 (2021).
11. Birahim, S. A. et al. Intrusion Detection for Wireless Sensor Network using Particle Swarm Optimization based Explainable Ensemble Machine Learning Approach. *IEEE Access* (2025).
12. Narkedimilli, S. et al. Enhancing IoT Network Security through Adaptive Curriculum Learning and XAI. *arXiv preprint arXiv:2501.11618*. (2025).
13. Naif Alatawi, M. Enhancing intrusion detection systems with advanced machine learning techniques: an ensemble and explainable artificial intelligence (AI) approach. *Secur. Priv.* **8** (1), e496 (2025).
14. Izuazu, U. U., Nwakanma, C. I., Kim, D. S. & Lee, J. M. Explainable and perturbation-resilient model for cyber-threat detection in industrial control systems Networks. *Discover Internet of Things*, *5*(1), p.9. (2025).
15. Patel, N. et al. June. X-NET: Explainable AI-Based Network Data Security Framework for Healthcare 4.0. In *2024 IEEE International Conference on Communications Workshops (ICC Workshops)* (pp. 481–486). IEEE. (2024).
16. Baral, S., Saha, S. & Haque, A. November. An adaptive End-to-End IoT security framework using explainable AI and LLMs. In *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)* 469–474 (IEEE, 2024).
17. Tripathy, S. S. et al. An adaptive explainable AI framework for Securing consumer electronics-based IoT applications in fog-cloud infrastructure. *IEEE Trans. Consum. Electron.* **71**, 1889–1896 (2024).
18. Kumar, P., Javeed, D., Kumar, R. & Islam, A. N. Blockchain and explainable AI for enhanced decision-making in cyber threat detection. *Software: Practice and Experience*. (2024).
19. Zeghida, H. et al. XMID-MQTT: explaining machine learning-based intrusion detection system for MQTT protocol in IoT environment. *International Journal of Information Security*, *24*(3), p.128. (2025).
20. Nandanwar, H. & Katarya, R. Deep learning enabled intrusion detection system for Industrial IOT environment. *Expert Systems with Applications*, *249*, p.123808. (2024).
21. Reddy, D. R., Ramani, S., Mohan, D., Sahukar, L. & Ramaswamy, T. Secure iotnet: a graph-residual adversarial network integrated with Hawk-Bee optimizer for intrusion detection in IoT wireless networks. *Int. J. Data Sci. Anal.* https://doi.org/10.1007/s41060-025-00789-w (2025).

22. Nandanwar, H. & Katarya, R. TL-BILSTM iot: transfer learning model for prediction of intrusion detection system in IoT environment. *Int. J. Inf. Secur.* **23** (2), 1251–1277 (2024).
23. Nandanwar, H. & Katarya, R. Securing Industry 5.0: An explainable deep learning model for intrusion detection in cyber-physical systems. *Computers and Electrical Engineering*, *123*, p.110161. (2025).
24. Kauhsik, B., Nandanwar, H. & Katarya, R. October. Iot security: a deep learning-based approach for intrusion detection and prevention. In *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)* (pp. 1–7). IEEE. (2023).
25. Nandanwar, H. & Katarya, R. Privacy-preserving data sharing in blockchain-enabled Iot healthcare management system. *Comput. J.* bxaf065 (2025).
26. Attique, D., Hao, W., Ping, W., Javeed, D. & Kumar, P. Explainable and data-efficient deep learning for enhanced attack detection in Iiot ecosystem. *IEEE Internet Things J.* **11** (24), 38976–38986 (2024).
27. Bajpai, S. A. & Patankar, A. B. Self-configuring intrusion detection using adaptive goal target optimization based deep Bi-LSTM in blockchain networking systems. *Intell. Decis. Technol.* **19** (1), 175–193 (2025).
28. Elsaid, S. A., Shehab, E., Mattar, A. M., Azar, A. T. & Hameed, I. A. Hybrid intrusion detection models based on GWO optimized deep learning. *Discover Applied Sciences*, *6*(10), p.531. (2024).
29. Omidkar, A., Es'haghian, R. & Song, H. Using machine learning methods for Long-term technical and economic evaluation of wind power plants. *Green Energy Resour.* **3**, 100115 (2025).
30. Guo, W., Hou, Z., Dai, F., Wang, J. & Li, S. Global information enhanced adaptive bald eagle search algorithm for photovoltaic system optimization. *Energy Rep.* **13**, 2129–2152 (2025).
31. Liu, Y., Ning, C., Zhang, Q., Yuan, G. & Li, C. Research on ocean buoy attitude prediction model based on multi-dimensional feature fusion. *Frontiers in Marine Science*, *11*, p.1517170. (2024).
32. Wei, M. & Du, X. Apply a deep learning hybrid model optimized by an improved chimp optimization algorithm in $PM_{2.5}$ prediction. *Mach. Learn. Appl.* **19**, 100624 (2025).
33. Wu, Z., Cha, S., Wang, C., Qu, T. & Zou, Z. Salmon Consumption Behavior Prediction Based on Bayesian Optimization and Explainable Artificial Intelligence. *Foods*, *14*(3), p.429. (2025).
34. https://www.kaggle.com/datasets/mohamedamineferrag/edgeiiotset-cyber-security-dataset-of-iot-iiot
35. https://research.unsw.edu.au/projects/bot-iot-dataset
36. Tazrin, T., Rahman, Q. A., Fouda, M. M. & Fadlullah, Z. M. LiHEA: migrating EEG analytics to ultra-edge IoT devices with logic-in-headbands. *Ieee Access.* **9**, 138834–138848 (2021).
37. Belachew, H. M. et al. Design a Robust DDoS Attack Detection and Mitigation Scheme in SDN-Edge-IoT by Leveraging Machine Learning. *IEEE Access*. (2025).
38. Cruz Castañeda, W. A. & Bertemes Filho, P. Improvement of an Edge-IoT Architecture Driven by Artificial Intelligence for Smart-Health Chronic Disease Management. *Sensors*, *24*(24), p.7965. (2024).
39. Christopher, V. et al. Minority resampling boosted unsupervised learning with hyperdimensional computing for threat detection at the edge of internet of things. *IEEE Access.* **9**, 126646–126657 (2021).
40. Chu, H. C. & Lin, Y. J. Improving the IoT Attack Classification Mechanism with Data Augmentation for Generative Adversarial Networks. *Applied Sciences*, *13*(23), p.12592. (2023).
41. Zeeshan, M. et al. Protocol-based deep intrusion detection for Dos and Ddos attacks using unsw-nb15 and bot-iot data-sets. *IEEE Access.* **10**, 2269–2283 (2021).

Acknowledgements

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/231/46. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R716), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Ongoing Research Funding program, (ORF-2025-459), King Saud University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-3030-11". The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

Author contributions

Sarah A. Alzakari: Conceptualization, methodology development, experiment, formal analysis, investigation, writing. Mohammed Aljebreen: Formal analysis, investigation, validation, visualization, writing. Nazir Ahmad: Formal analysis, review and editing. Sultan Alahmari : Methodology, investigation. Othman Alrusaini: Review and editing.Ali Alqazzaz and Hassan Alkhiri: Discussion, review and editing. Yahia Said: Conceptualization, methodology development, investigation, supervision, review and editing.All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.