A Secure and Privacy-Preserving IoT Cybersecurity Framework Using Feature Selection and Ensemble Deep Learning for Smart City Applications

Samah Alzanin
Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Kharj, Saudi Arabia
s.alzanin@psau.edu.sa (corresponding author)

Mohammed Alonazi
Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
mn.alonazi@psau.edu.sa

Received: 11 January 2026 | Revised: 27 January 2026 | Accepted: 7 February 2026

 DOI: https://doi.org/10.48084/etasr.17472

ABSTRACT

The rapid proliferation of Internet of Things (IoT) applications in recent years has played a significant role in the advancement of smart cities. Sustainable cities and communities are a main goal of the Sustainable Development Goal (SDG), which aims to make urban areas safe, resilient, and sustainable by 2030. Smart cities employ IoT-based technologies, communication systems, and intelligent applications to optimize operational efficiency and improve both service delivery and citizens' quality of life. Cybersecurity has become a problematic issue in IoT environments, needing effective addressing of the persistent cyberthreats. Intrusion Detection Systems (IDSs) are required to protect data, and the utilization of Artificial Intelligence (AI) subfields such as Machine Learning (ML) and Deep Learning (DL) is proven to be among the most effective solutions. In this paper, an Artificial Intelligence-Driven Cybersecurity Detection using Ensemble Models (AIDCD-EM) framework in smart city applications is proposed. Initially, the Z-score normalization is used for data normalization. For dimensionality reduction, the AIDCD-EM utilizes the enhanced Mutual Information Feature Selection (MIFS) method and an ensemble classification comprising the Bi-directional Gated Recurrent Unit (BiGRU), autoencoder (AE), and Graph Convolutional Networks (GCNs), is utilized for cyberattack classification. The experimental valuation of the AIDCD-EM model highlighted superior accuracy values of 99.60% and 99.55% when investigated under the ToN-IoT and Edge-IIoT datasets.

Keywords-artificial intelligence; smart city; cybersecurity; internet of things; ensemble models; mutual information

I. INTRODUCTION

The development of smart cities represents a significant transformation in urban environments, propelled by the integration of advanced technologies, including the IoT, big data analytics, and AI [1]. These technologies enable the creation of intelligent, sustainable, and highly interconnected urban ecosystems [2, 3]. Nevertheless, the growing deployment of connected devices and digital infrastructures broadens the potential attack surface [4], as the highly interconnected nature of smart city systems presents substantial cybersecurity challenges [5], including risks arising from large-scale data acquisition, and complexities in managing smart infrastructure [6]. Therefore, safeguarding the security and privacy of smart city systems is vital to protect critical structures and ensure the confidentiality of citizens' data [7]. Compromising IoT gadgets might be employed to launch large-scale distributed DoS (DDoS) attacks, besides specific methods [8], overcoming them with unreasonable traffic and containing the same to legal consumers [9]. Cybersecurity has experienced enormous shifts in operations and its technology in the computing context, and Data Science (DS) is driving the change. ML and DL, as core components of AI, play a significant role in extracting insights from data [10].

In this article, an Artificial Intelligence-Driven Cybersecurity Detection using Ensemble Models (AIDCD-EM) model is proposed. The key contributions are:

- Initially, Z-score normalization is used for pre-processing to improve learning efficiency and consistent feature scaling.
- The most informative features are chosen based on MIFS for choosing while also removing the repeating features from the dataset. This also mitigates the overhead problem and preserves distinct features.
- The ensemble model captures patterns, graph-based associations, and latent feature representations with high detection accuracy and robustness for a diverse and growing range of IoT cyberattacks while maintaining efficiency in resource-constrained environments.

II. LITERATURE REVIEW

Authors in [11] developed a Deep Gated Recurrent Unit (D-GRU) as an AI-based structure. Authors in [12] introduced the Bald Eagle Search Optimizer with a Hybrid DL-based botnet detection (BESO-HDLBD) model. This approach employs HDL, which is an incorporation of Bidirectional Long Short-Term Memory (BiLSTM), CNN, and attention theory. Authors in [13] developed an innovative method for recognizing cyber threats in IoT settings by integrating Deep Belief Networks (DBNs), AEs, and Self-Organizing Maps (SOMs). Authors in [14] introduced methods such as Deep LSTM and Deep Neural Networks (DNNs). A Deep Sparse AE (DSAE) was also utilized. Authors in [15] projected an Improved Tunicate Swarm Algorithm (ITSA)-based Feature Selection (FS) method for reducing size. Moreover, the attention LSTM-NN (ALSTM-NN) methodology was utilized to classify and detect cyberthreats. Authors in [16] developed a convertible method that utilizes DL models. Authors in [17] presented a new feature reduction method to use the Modified Lemrus Optimization Algorithm (MLOA) to select the optimum feature set. Additionally, the FDIA recognition method was implemented. Authors in [18] addressed anomaly recognition in smart city settings derived from the IoT.

The limitations of the reviewed works include high computational complexity, dependence on labelled datasets, and restricted adaptation. Also, various methods concentrate on particular devices, thus mitigating generalization across diverse IoT networks. There is a research gap in developing lightweight, adaptive, and scalable models.

III. MATERIALS AND METHODS

The proposed AIDCD-EM system comprises various steps namely data normalization, MIFS-based FS, a classification model ensemble, and tuning. Figure 1 depicts the overall flow of the AIDCD-EM approach.

A. Data Normalization: Z-Score

Z-score normalization is initially applied to convert the input data into an appropriate format for processing. It is also called standardization, a statistical model employed in cybersecurity for IoT environments to measure features, ensuring equality across dissimilar ranges of data. It converts data by dividing the standard deviation and subtracting the mean, which results in a distribution with 0 and 1 for the mean and standard deviation. In IoT systems, where data originate from varied sources with fluctuating scales, Z-score aids in enhancing anomaly recognition and the accuracy of intrusion detection. It permits DL models to converge faster and create more exact security forecasts.

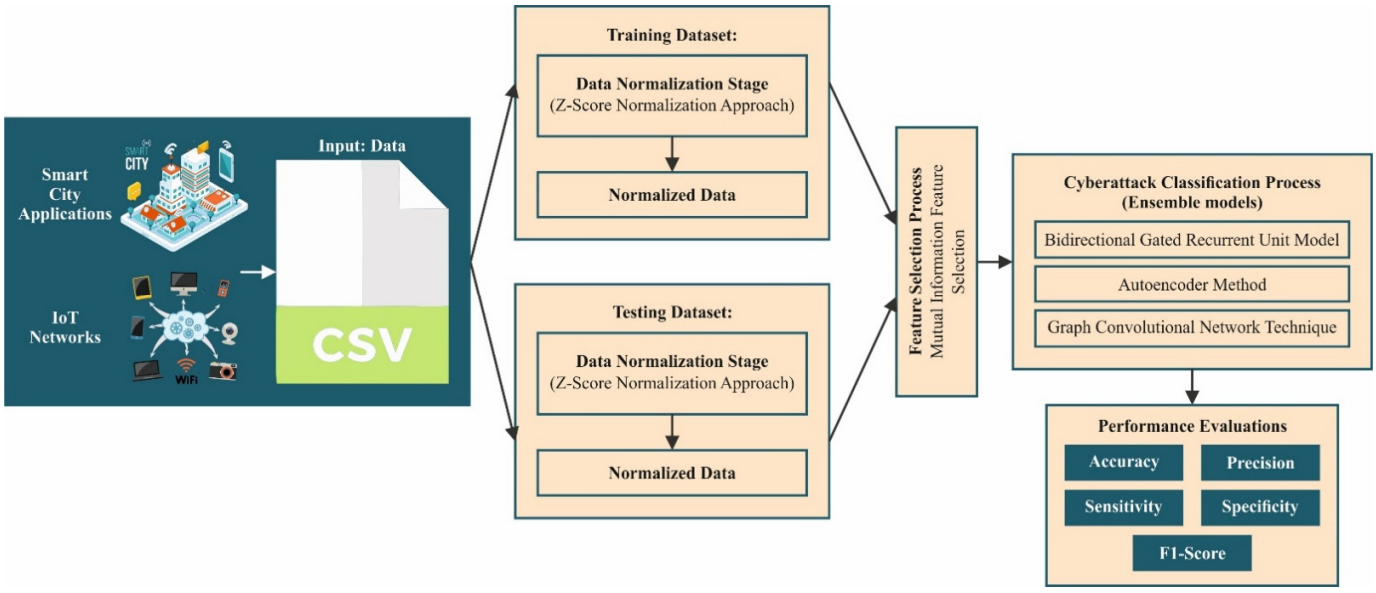

Fig. 1. Workflow of the AIDCD-EM approach.

B. FS Process: MIFS

For dimensionality reduction, the AIDCD-EM model utilizes the MIFS system. The MIFS is a prevalent method for FS that can efficiently choose the appropriate features irrespective of data distribution, which makes it suitable for earlier detection conditions, whereas data possesses no adequate attack patterns. For two distinct variable quantities, Mutual Information (MI) [19] is a quantity of how much data

is exchanged between variables. The computation of MI is specified by:

$$\mathrm{MI}(U,V) = H(U) - H(U|V) = \sum_{v\epsilon V}\sum_{u\epsilon U} p\,(u,v) log\,\frac{p(u,v)}{p(u)p(v)} \quad (1)$$

where $p(u)$ and $p(v)$ describe the marginal distributions of u and v, $H(U)$ denotes the entropy of U, $H(U|V)$ refers to the Conditional Entropy (CE) of specified U and V, and $p(u,v)$ denotes the combined distribution of u and v. The entropy $H(U)$ is measured by (2) and $H(U|V)$ is calculated measured by (3):

$$H(U) = -\sum_{u_i\in U} p\,(u_i) log\big(p(u_i)\big) \quad (2)$$

$$H(U|V) = -\Sigma_{v_j\in V} p\big(y_j\big)\Sigma_{u_i\in U} p\big(u_i|v_j\big) log\left(p\big(u_i|v_j\big)\right) \quad (3)$$

The overall mathematical expression for the linear integrations of Shannon data relationship is:

$$J(U_k) = I\big(U_{k;}V\big) - \beta\sum_{U_{j\epsilon S}} I\left(U_{j;}U_k\right) + \gamma\sum_{U_{j\epsilon S}} I\left(U_{j;}U_k|V\right) \quad (4)$$

This mathematical form is collected with relevance and redundancy terms, denoted by (5) and (6), whereas the summation of marginal redundancy exhibits the redundancy term, shown in (7), and conditional redundancy, denoted in (8). These are weighed by variables β and γ with values ranging in [0, 1].

$$I(U_k;V) \quad (5)$$

$$\beta\sum_{U_j\epsilon S} I\,(U_j;U_k) + \gamma\sum_{U_j\epsilon S} I\,(U_j;U_k|\mathrm{V}) \quad (6)$$

$$\beta\sum_{U_j\epsilon S} I\left(U_j;U_k\right) \quad (7)$$

$$\gamma\sum_{U_j\epsilon S} I\left(U_j;U_k|\mathrm{V}\right) \quad (8)$$

The MI among the candidate features U_k and the classes V is described by $I(U_k;V)$, whereas the conditional MI among U_k and alternative features U_j at the chosen sets S specifies that the classes V can be defined by $I(U_j;U_k|V)$.

C. Ensemble Classification Model

The Bi-GRU, AE, and GCN models are used for cyberattack classification. The ensemble utilizes stacking to send the combined features to the meta-classifier. The GCN efficiently models IoT network traffic, thus enabling the capture of inter-device associations. Diverse features are also used for enhancing accuracy, robustness, and generalization.

1) BiGRU Model

Bi-GRU can take the slight variations among dissimilar time steps and can satisfactorily join the frequency and time-domain signals of the dual dynamic actions. The efficiency and simplicity of Bi-GRU [20] allow it to handle signal data in the real world under restricted source situations. As a result, Bi-GRU is designated to examine the changing connection of the sequences of local features.

$$r_t = \sigma\cdot(W_r x_t + U_r h_{t-1}) \quad (9)$$

$$z_t = \sigma\cdot(W_z x_t + U_z h_{t-1}) \quad (10)$$

$$\tilde{h}_t = \tanh(r_t\cdot U h_{t-1} + W x_t) \quad (11)$$

$$h_t = (1-z_t)\cdot h_{t-1} + z_t\cdot h_{t-1} \quad (12)$$

where x_{t-1} and x_t are features at times $t-l$ and t, r_t is the reset gate controlling ignored past info, z_t is the update gate controlling retained past info, $\tilde{h}_t$ is the candidate hidden state, and h_t is the hidden output, with σ and tanh as activation functions and W, U matrices as trainable parameters.

Bi-GRU is a bidirectional recurrent model where each output depends on the forward-and-backward propagating GRU, $\vec{h}_t$, and $\overleftarrow{h}_t$, at time step t, demonstrated as:

$$\vec{h}_t = GRU\big(x_t,\vec{h}_{t-1}\big) \quad (13)$$

$$\overleftarrow{h}_t = GRU\big(x_t,\overleftarrow{h}_{t-1}\big) \quad (14)$$

$$h_t = \alpha_t\cdot\vec{h}_t + \beta_t\cdot\overleftarrow{h}_t + c_t \quad (15)$$

whereas *GRU* is the calculation procedure of GRU, $\overleftarrow{h}_t$, and $\vec{h}_t$ represent backward and forward HL outputs of GRU, α_t and β_t denote the weighting outputs of the equivalent HLs, and c_t refers to the bias of the HL consistent with h_t.

2) AE Classifier

AEs [21] have become helpful in many fields, including object detection, image classification, image reconstruction, and Natural Language Processing (NLP). An AE contains dual sub-networks: a decoder and an encoder. For all input data x, the decoding $D_\theta\colon \mathcal{Z}\to\mathcal{X}$ rebuilds $\hat{x}\in\mathcal{X}$ from the variable of the encoding latent area z:

$$\hat{x} = D_\theta(z) = D_\theta\left(\mathcal{E}_\varphi(x)\right) \quad (16)$$

where φ and θ signify the parameters of the encoding and decoding systems. The aim of training the AE is to discover the functions $\mathcal{E}_\varphi(\cdot)$ and $D_\theta(\cdot)$, which are calculated by a function of the reconstruction loss ℓ_{REC}, to reduce the change between the reconstructed data $\hat{x}$ and the input data x:

$$\underset{\varphi,\theta}{\operatorname{argmin}}\ \mathbb{E}[\ell_{REC}(x, D_\theta(\mathcal{E}_\varphi(x)))] \quad (17)$$

whereas $\mathbb{E}[\cdot]$ signifies the expectant and data distribution. An autoencoder has a mirrored encoder–decoder structure, with many variants based on architecture and training.

3) GCN Model

In the multilayer GCN [22] structure, an undirected graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ contains edges $(v_i,v_j)[\mathcal{E}(1\le i,j\le N)N$ and nodes $v_i[\mathcal{V}(1\le\ i\le N)$. The connection among nodes v_i and v_j is characterized by the matrix of adjacency $A[\mathbb{R}^{N\times N}$. In the lth layer, once the node features are converted to the $(l+1)$th layer, they are upgraded utilizing the succeeding layer-by-layer rule of propagation:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{\frac{1}{2}}H^{(l)}W^{(l)}\right) \quad (18)$$

whereas $\tilde{A} = A + I_N$ denotes the adjacent matrix of an undirected graph $\mathcal{G}$ using additional self-connections. I_N represents the identity matrix, $\widetilde{D} = \sum_j \tilde{A}$, W signifies the layer-particular trainable weighting matrix, and $\sigma(\cdot)$ is the activation function. $H^{(l)}[\mathbb{R}^{N\times D}$ represents the activation matrix within the lth layer, and $H^{(0)}$ symbolizes the input of the ANN.

Let the lth $\mathfrak{F}_{l)}^{aph}$ convolution layer obtain the collection of node features $H^{(l)} = \left\{h_1^{(l)}, h_2, \cdots, h_N^{(l)}\right\}$, $h_i^{(l)}[\mathbb{R}^{F_l}$ represents the input and provides a novel collection, $H^{(l+1)} = \left\{h_1^{(l+1)}, h_2^{(l+1)}, \cdots, h_N^{(l+1)}\right\}$, $h_i^{(l+1)}[\mathbb{R}^{F_{l+1}}$, for the $(l+1)$th layer. F_{l+1} and F_l represent feature counts in all nodes at the $(l+1)$th and lth layer, respectively. The layer-by-layer rule of propagation is stated by:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N(i)} \frac{1}{c_{ij}} W^{(l)} h_j^{(l)}\right) \quad (19)$$

where $\mathcal{N}(i)$ represents the set of neighbors of the i^{th} node, c_{ij} depicts the chosen standardization constant for the edge (v_i, v_j) described as the square root product of node degrees (for example: $c_{ij} = \sqrt{|\mathcal{N}(j)|}\sqrt{|\mathcal{N}(i)|}$), and $W^{(l)} \in \mathbb{R}^{F_{l+1}\times F_l}$.

The graph structure $\mathcal{G}$ originates from an unstructured network of ISSM. The edges and nodes of the network are considered as the edges and nodes of graph $\mathcal{G}$.

IV. EXPERIMENTAL ANALYSIS

The performance study of the proposed AIDCD-EM methodology is investigated for the the ToN-IoT dataset [23]. The system run on a Python 3.6.5 with an i5-8600k CPU, 4GB GPU, 16GB RAM, 250GB SSD, and 1TB HDD, using the following parameter values: 0.01 learning rate, ReLU, 50 epochs, 0.5 dropout, and batch size 5. This dataset contains 119,957 samples, including 78,369 normal instances and various cyberattacks such as MiTM, DoS, DDoS, password, injection, XSS, ransomware, and backdoor attacks. This diverse distribution supports effective evaluation of IoT intrusion detection models. Out of the total 42 features, only 23 were considered. Resampling techniques were used for class imbalance handling, while regularization and early stopping were used to mitigate overfitting. Also, efficient training and testing sets were used.

Figure 2 demonstrates the confusion matrices on the training and test datasets (split ratio: 70:30) of the AIDCD-EM method on the ToN-IoT dataset. Table I exhibits the comparison results of the AIDCD-EM approach with existing techniques [9, 25, 26] on the ToN-IoT dataset, showcasing its superiority.

Table II portrays the ablation study evaluation of the AIDCD-EM approach. The AIDCD-EM approach illustrated highest performance with an $accu_y$ of 99.60%, $prec_n$ of 90.31%, $sens_y$ of 96.91%, and $spec_y$ of 99.77%, respectively. The BiGRU+MIF recorded an accuracy of 97.71%, while AE+MIF and GCN+MIF attained 98.50% and 99.08% values.

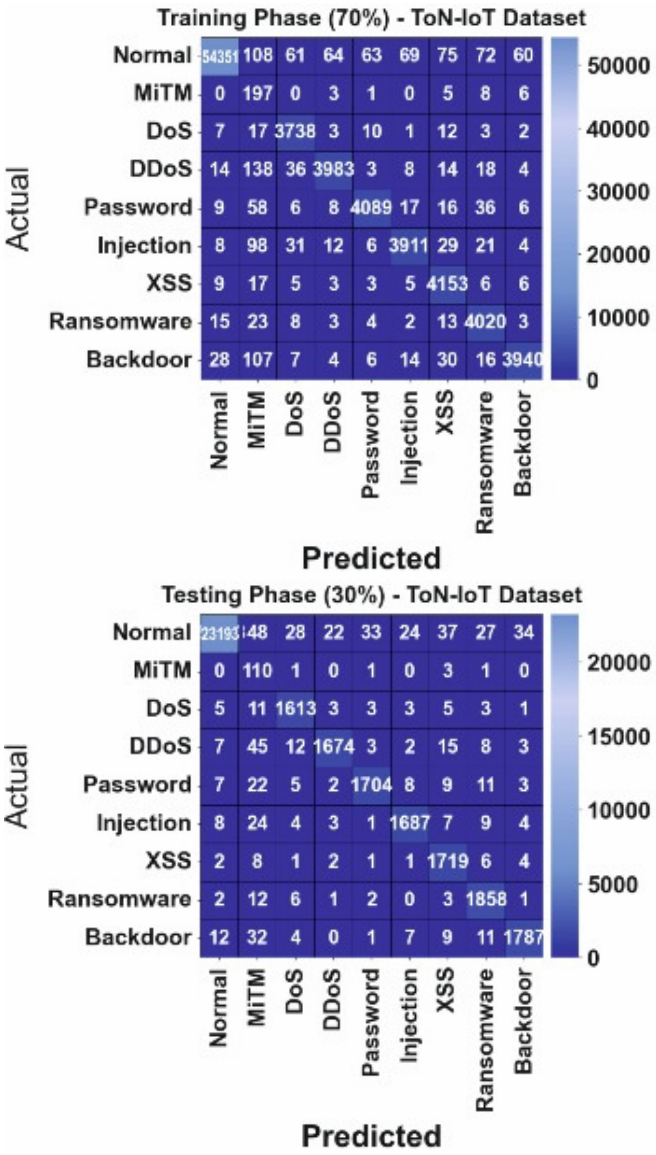

Fig. 2. Confusion matrix of the proposed model on the ToN-IoT dataset.

TABLE I. COMPARISON OUTCOME OF THE AIDCD-EM MODEL ON THE TON-IOT DATASET [9, 25, 26]

Method	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	Error rate
ET	98.05	89.51	96.27	95.91	1.95
DFF-CHI	85.61	90.55	94.64	96.48	14.39
DFF	94.74	87.48	92.75	91.37	5.26
DL-IDS Metavers-IoT	97.29	88.75	95.59	99.31	2.71
CNN-Metaverse-IDS	98.88	90.05	94.02	95.84	1.12
ICA+LOF	93.55	86.77	92.08	90.60	6.45
CHI2+HBOS	97.31	80.96	95.83	92.01	2.69
Bi-LSTM	90.34	89.48	94.11	91.36	9.66
RNN Algorithm	93.72	88.43	89.80	91.00	6.28
BHS-ALOHDL	94.61	81.90	95.53	98.82	5.39
AIDCD-EM	99.60	90.31	96.91	99.77	0.40

TABLE II. ABLATION STUDY ANALYSIS OF THE AIDCD-EM METHOD ON THE TON-IOT DATASET

Method	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$
BiGRU+MIF (with FS without AE and GCN)	97.71	88.73	94.70	97.88
AE+MIF (with FS without AE and BiGRU)	98.50	89.24	95.39	98.64
GCN+MIF (with FS without BiGRU and GCN)	99.08	89.80	96.11	99.24
AIDCD-EM (ensemble classifiers with the MIFS process)	99.60	90.31	96.91	99.77

The AIDCD-EM method was also investigated under the Edge-IIoT dataset [24]. This dataset contains 56,000 records, including regular traffic and various attack types. Out of the 62 total features, only 38 were considered. Figure 3 represents the classifier outputs of the AIDCD-EM approach on the Edge-IIoT dataset. Table III examines the comparison results of the AIDCD-EM approach on the Edge-IIoT dataset with existing methods [10, 25, 26], showcasing its superiority.

Table IV depicts the ablation study assessment of the AIDCD-EM model on the Edge-IIoT dataset. The AIDCD-EM model illustrated optimum results with an $accu_y$ of 99.55%, $prec_n$ of 96.95%, $sens_y$ of 97.29%, and $spec_y$ of 99.76%. Moreover, BiGRU+MIF achieved an accuracy of 97.80%, while AE+MIF improved to 98.50%, and GCN+MIF reached 99.03%.

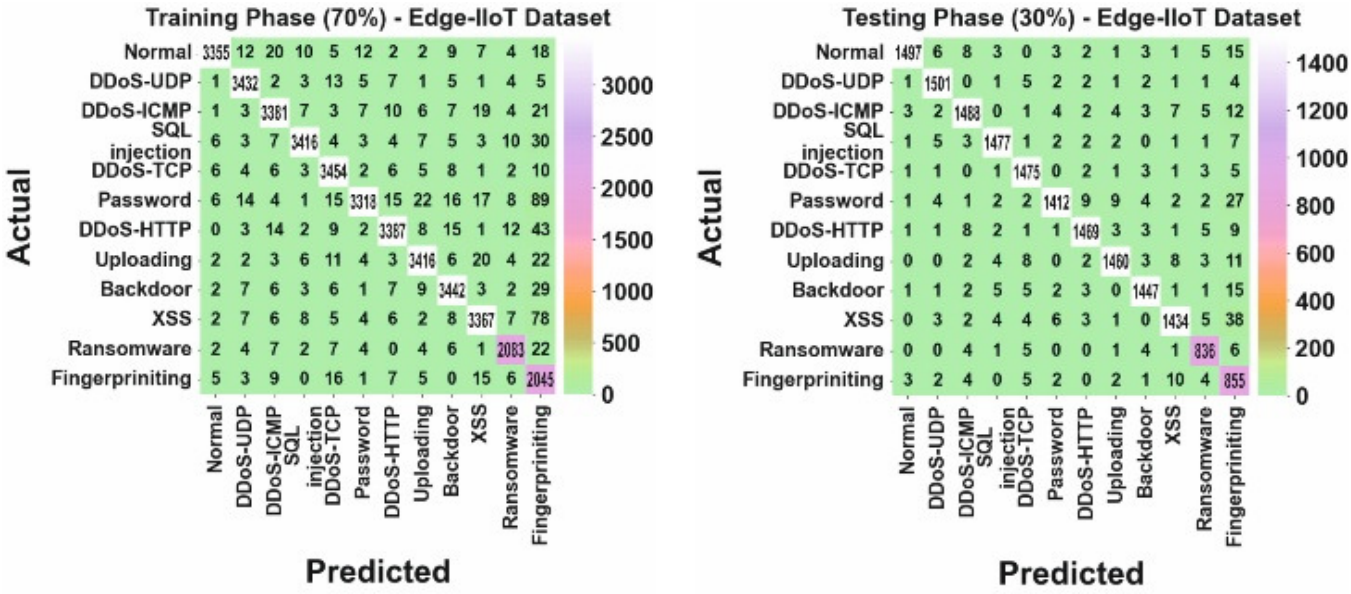

Fig. 3. Confusion matrices of the proposed method on the Edge-IIoT dataset (70:30 split ratio).

TABLE III. COMPARISON ANALYSIS OF THE AIDCD-EM APPROACH ON THE EDGE-IIOT DATASET [10, 25, 26]

Method	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	Error rate
XGB	81.81	93.06	93.72	98.16	18.19
LGBM	79.46	90.76	91.25	97.98	20.54
TabNet	77.56	89.64	89.69	94.76	22.44
1D CNN Method	99.28	92.43	93.08	98.41	0.72
GB	90.7	90.23	90.67	99.22	9.30
LSTM Model	98.44	89	89.13	94.08	1.56
AE	98.5	89.07	90.65	89.86	1.50
GNB	99.31	92.56	93.09	95.6	0.69
Attention-Hybrid DL	89.22	96.29	93.62	90.28	10.78
GRU Method	97.54	90.33	93.97	92.68	2.46
AIDCD-EM	99.55	96.95	97.29	99.76	0.45

TABLE IV. ABLATION STUDY ASSESSMENT OF THE AIDCD-EM METHODOLOGY ON THE EDGE-IIOT DATASET

Method	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$
BiGRU+MIF (with FS without AE and GCN)	97.80	95.08	95.14	97.94
AE+MIF (with FS without AE and BiGRU)	98.50	95.83	95.86	98.51
GCN+MIF (with FS without BiGRU and GCN)	99.03	96.45	96.53	99.21
AIDCD-EM (ensemble classifier with MIFS)	99.55	96.95	97.29	99.76

V. CONCLUSION

The proposed AIDCD-EM model for smart city applications is presented and examined in this paper. The AIDCD-EM model relies on improving the effectual detection of cyberattacks in an IoT network. The model comprises Z-score normalization, Mutual Information Feature Selection (MIFS), and an ensemble classification, combining the Bi-directional Gated Recurrent Unit (BiGRU), autoencoder (AE), and Graph Convolutional Networks (GCNs). The experimental valuation of the AIDCD-EM method highlighted superior accuracy values of 99.60% and 99.55% when investigated under the ToN-IoT and Edge-IIoT datasets, respectively.

The limitations of the proposed system include its reliance on pre-collected and labeled datasets and restrictions on some IoT devices and real-time deployment. Future work should concentrate on adaptive, online learning approaches and lightweight anomaly detection models to enhance scalability, privacy, and real-time responsiveness.

DATA AVAILABILITY STATEMENT

The utilized datasets can be found in [23, 24].

ACKNOWLEDGMENT

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/33090).

REFERENCE

[1] T. V. Hoang, "Impact of Integrated Artificial Intelligence and Internet of Things Technologies on Smart City Transformation," *Journal of Technical Education Science*, vol. 19, no. Special Issue 01, pp. 64–73, Feb. 2024, https://doi.org/10.54644/jte.2024.1532.

[2] U. Chatterjee, G. S. Bhunia, D. Mahata, and U. Singh, "Smart Cities and Their Role in Enhancing Quality of Life," in *Quality of Life*, CRC Press, 2021, pp. 127–143.

[3] M. Kalinin, V. Krundyshev, and P. Zegzhda, "Cybersecurity Risk Assessment in Smart City Infrastructures," *Machines*, vol. 9, no. 4, Apr. 2021, Art. no. 78, https://doi.org/10.3390/machines9040078.

[4] T. Nishitha and A. Khare, "Smart Contract-Enhanced Residual GRU with Merkle-Damgard Cryptography for IoT Attack Detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19331–19336, Feb. 2025, https://doi.org/10.48084/etasr.8860.

[5] V. Baladari, "Adaptive Cybersecurity Strategies: Mitigating Cyber Threats and Protecting Data Privacy," *Journal of Scientific and Engineering Research*, vol. 7, no. 8, pp. 279–288, Dec. 2020, https://doi.org/10.5281/zenodo.15044844.

[6] K. Thakur, "Analysis of Denial of Services (DOS) Attacks and Prevention Techniques," *International Journal of Engineering Research & Technology*, vol. 4, no. 7, Jul. 2015, https://doi.org/10.17577/IJERTV4IS070164.

[7] R. H. Ali, S. F. M. Alazawy, A. Mustafa, and K. R. Erzaij, "Smart City Feasibility Study using IoT and Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17494–17500, Oct. 2024, https://doi.org/10.48084/etasr.8714.

[8] C. P. Singh, R. Yamaganti, and L. S. Umrao, "A privacy-preserving and secure framework using blockchain-based quantum-inspired complex convolutional neural network for IoT-driven smart cities," *Peer-to-Peer Networking and Applications*, vol. 19, no. 1, Nov. 2025, Art. no. 3, https://doi.org/10.1007/s12083-025-02168-5.

[9] T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method," *Sensors*, vol. 22, no. 3, Feb. 2022, https://doi.org/10.3390/s22031154.

[10] T. Hasan, A. Hossain, M. Ansari, and T. Syed, "Enhanced Intrusion Detection in IIoT Networks: A Lightweight Approach with Autoencoder-Based Feature Learning," in *10th International Conference on Internet of Things, Big Data and Security*, Mar. 2026, pp. 207–214, https://doi.org/10.5220/0013203700003944.

[11] P. M. Kumar, B. P. Kavin, A. Jagathpally, and T. Shahwar, "Transforming the cybersecurity space of healthcare IoT devices using Deep Learning," in *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, Houston, TX, USA, Oct. 2025, pp. 1–6, https://doi.org/10.1109/ICAIC63015.2025.10849305.

[12] L. A. Maghrabi *et al.*, "Enhancing Cybersecurity in the Internet of Things Environment Using Bald Eagle Search Optimization With Hybrid Deep Learning," *IEEE Access*, vol. 12, pp. 8337–8345, 2024, https://doi.org/10.1109/ACCESS.2024.3352568.

[13] A. Bensaoud and J. Kalita, "Optimized detection of cyber-attacks on IoT networks via hybrid deep learning models," *Ad Hoc Networks*, vol. 170, Apr. 2025, Art. no. 103770, https://doi.org/10.1016/j.adhoc.2025.103770.

[14] H. Gonaygunta, G. S. Nadella, P. Pramod Pawar, and D. Kumar, "Enhancing Cybersecurity: The Development of a Flexible Deep Learning Model for Enhanced Anomaly Detection," in *2024 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA, Feb. 2024, pp. 79–84, https://doi.org/10.1109/SIEDS61124.2024.10534661.

[15] H. Alamro *et al.*, "Mathematical modelling-based blockchain with attention deep learning model for cybersecurity in IoT-consumer electronics," *Alexandria Engineering Journal*, vol. 113, pp. 366–377, Feb. 2025, https://doi.org/10.1016/j.aej.2024.11.016.

[16] V. Jaganraja and R. Srinivasan, "An agile solution for enhancing cybersecurity attack detection using deep learning privacy-preservation in IoT-smart city," *Wireless Networks*, vol. 31, no. 3, pp. 2227–2242, Mar. 2025, https://doi.org/10.1007/s11276-024-03876-1.

[17] F. A. F. Alrslani *et al.*, "Enhancing cybersecurity via attribute reduction with deep learning model for false data injection attack recognition," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 3944, https://doi.org/10.1038/s41598-024-82566-6.

[18] W. Villegas-Ch, J. Govea, and A. Jaramillo-Alcazar, "IoT Anomaly Detection to Strengthen Cybersecurity in the Critical Infrastructure of Smart Cities," *Applied Sciences*, vol. 13, no. 19, Oct. 2023, Art. no. 10977, https://doi.org/10.3390/app131910977.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, and vol. 27, no. 4, pp. 623–656, 1948.

[20] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Jul. 2014, pp. 1724–1734, https://doi.org/10.3115/v1/D14-1179.

[21] D. E. Rumelhart and J. L. McClelland, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, Cambridge, MA, USA: MIT Press, 1987, pp. 318–362.

[22] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks." arXiv, Feb. 22, 2017, https://doi.org/10.48550/arXiv.1609.02907.

[23] M. Sarhan, S. Layeghy, and & M. Portmann"CIC-ToN-IoT." 2023, [Online]. Available: https://www.kaggle.com/datasets/dhoogla/cictoniot.

[24] M. A. Ferrag, "Edge-IIoTset Cyber Security Dataset of IoT & IIoT." [Online]. Available: https://www.kaggle.com/datasets/mohamedamineferrag/edgeiiotset-cyber-security-dataset-of-iot-iiot.

[25] S. Srivastav, A. K. Shukla, S. Kumar, and P. K. Muhuri, "HYRIDE: HYbrid and Robust Intrusion DEtection approach for enhancing cybersecurity in Industry 4.0," *Internet of Things*, vol. 30, Mar. 2025, Art. no. 101492, https://doi.org/10.1016/j.iot.2025.101492.

[26] M. A. Alkhonaini *et al.*, "Sandpiper optimization with hybrid deep learning model for blockchain-assisted intrusion detection in iot environment," *Alexandria Engineering Journal*, vol. 112, pp. 49–62, Jan. 2025, https://doi.org/10.1016/j.aej.2024.10.032.