An Explainable and Resilient Intrusion Detection System for Industry 5.0

Danish Javeed, *Student Member, IEEE*, Tianhan Gao, Prabhat Kumar, *Member, IEEE*, and Alireza Jolfaei, *Senior Member, IEEE*

***Abstract*—Industry 5.0 is a emerging transformative model that aims to develop a hyperconnected, automated, and data-driven industrial ecosystem. This digital transformation will boost productivity and efficiency throughout the production process but will be more prone to new sophisticated cyber-attacks. Deep learning-based Intrusion Detection Systems (IDS) have the potential to recognize intrusions with high accuracy. However, these models are complex and are treated as a black box by developers and security analysts due to the inability to interpret the decisions made by these models. Motivated by the challenges, this paper presents an explainable and resilient IDS for Industry 5.0. The proposed IDS is designed by combining bidirectional long short-term memory networks (BiLSTM), a bidirectional-gated recurrent unit (Bi-GRU), fully connected layers and a softmax classifier to enhance the intrusion detection process in Industry 5.0. We employ the SHapley Additive exPlanations (SHAP) mechanism to interpret and understand the features that contributed the most in the decision of the proposed cyber-resilient IDS. The evaluation of the proposed model using the explainability can ensure that the model is working as expected. The experimental results based on the CICDDoS2019 dataset confirms the superiority of the proposed IDS over some recent approaches.**

***Index Terms*—Deep learning (DL), cyber-attacks, explainable artificial intelligence, intrusion detection system (IDS), Industry 5.0.**

I. Introduction

THE FIFTH industrial revolution also known as Industry 5.0 is considered as a next-level advancement. Its goal is to combine the human expert's creativity with effective, intuitive, and explicit machinery to bring forth manufacturing solutions that are more user-friendly and resource-efficient than those of Industry 4.0 [1]. It provides a narrative intent to facilitate the users and organizations. Industry 5.0 is expected to have a huge impact on consumer technology, accelerating innovation and revolutionizing the way products are conceived, manufactured, and supplied to customers. Consumer Electronic (CE) devices play a crucial role in such an industry through collecting data from the sensors and machines, monitoring and controlling them remotely [2]. Consumer-grade sensors and cameras are used to capture environmental data, i.e., temperature, air quality, and humidity, which can then be utilized to optimize industrial operations. Additionally, CE devices like smartphones, wearables, and tablets function as interfaces between machines and operators, offering real-time data and alarms on the status of the equipment. As a result, academia, industry, and individuals are endeavoring to integrate rapid commercialization flow while paying slight attention to the safety and security of Industry 5.0 devices and networks [3]. For instance, autonomous robots in manufacturing plant can be remotely hijacked and controlled by cybercriminals to frighten the company. Even with the availability of traditional security measures like authentication, encryption, access control, and data confidentiality, Industry 5.0 network have proven vulnerable to network attacks, necessitating the need for an extra layer of security. One commonly used strategy is to develop and deploy Intrusion Detection Systems (IDSs) for connected Industry 5.0 systems [4]. However, the variety of cyber attacks makes traditional IDS less effective. Thus, it is crucial to design an effective and reliable system in line with contemporary criteria. The IDS tracks online activity in real-time and spots unusual behavior. In the recent years, Deep Learning (DL)-based IDS became a trending research topic for researchers around the globe and they proposed numerous DL-based IDS to protect such industries against cyber threats. The authors of [5] proposed an IDS to identify threats and safeguard the network from them. However, for early detection, the proposed IDS must be updated frequently and should include the patterns and characteristics of new potential attacks.

DL-based IDS provides an efficient performance. However, these models lack explainability and interpretability, i.e., comprehending the underlying data proof of the prediction decisions for the behavior of the designed model [6]. Consequently, the decision lacks trust and their output cannot be further used to optimize the behaviour and reasoning offered by the sophisticated algorithm. The Explainable AI (XAI)-based IDS gives methodical and comprehensible justifications for its behaviors that users can follow. For instance, the authors of [7] designed a comprehensible architecture

Manuscript received 4 March 2023; revised 17 April 2023 and 25 May 2023; accepted 5 June 2023. Date of publication 7 June 2023; date of current version 26 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52130403, and in part by the China Fundamental Research Funds for the Central Universities under Grant N2017003. *(Corresponding author: Tianhan Gao.)*

Danish Javeed and Tianhan Gao are with the Software College, Northeastern University, Shenyang 110169, China (e-mail: 2027016@stu.neu.edu.cn; gaoth@mail.neu.edu.cn).

Prabhat Kumar is with the Department of Software Engineering, LUT School of Engineering Science, LUT University, 53850 Lappeenranta, Finland (e-mail: prabhat.kumar@lut.fi).

Alireza Jolfaei is with the College of Science and Engineering, Flinders University, Adelaide, SA 5001, Australia (e-mail: alireza.jolfaei@flinders.edu.au).

Digital Object Identifier 10.1109/TCE.2023.3283704

for IoT setups to track customer sentiment. They base their approach on merging enterprise data and IoT to model consumer sentiment, which improves customer prioritization and aids in problem-solving. Likewise, the authors in [8] proposed an XDL-based model to design an efficient IDS for Internet of Medical Things (IoMT) networks. Research on XDL-based IDS is still in its infancy, especially for IoT-enabled Industrial networks. Therefore, the proposed work designed an explainable and resilient IDS to protect such industries against evolving threats.

A. Contribution

The contributions of this research are as follows:

- A novel explainable and cyber-resilient IDS is designed by combining bidirectional long short-term memory networks (BiLSTM), a bidirectional-gated recurrent unit (Bi-GRU), fully connected layers, and a softmax classifier to enhance the attack detection process in Industry 5.0.
- The SHapley Additive exPlanations (SHAP) mechanism is employed to interpret and understand the decision made by the proposed DL-based sophisticated IDS. As a result, the explanation will help the security analyst to interpret the traffic features from the CICDDoS2019 dataset and the output can be further used to optimize and develop new algorithms for DL-based IDS. Furthermore, the experimental results based on the CICDDoS2019 dataset confirm the superiority of the designed IDS over some recent threat detection techniques.

The remainder of this work is structured as follows; Section II discuss the related work. The proposed detection scheme is elaborated in Section III. The experimental details are provided in Section IV. Section V discuss the result analysis. Finally, the conclusion along with the future remarks are presented in Section VI.

II. Related Work

Over the last decade, DL and ML-based approaches have demonstrated their utility in detecting anomalous entities in traditional IoT-based networks. The authors of [9] proposed a DL-based IDS which is capable to encounter the existence of threats in IoT networks. The model is based on a CNN classifier to obtain desired security objective. The authors trained and evaluated their proposed framework on the BoT-IoT dataset that comes with a huge variety of security threats and is considered an ideal choice to train IDS. The system has achieved 92.46% accuracy when evaluated on diverse performance metrics. The authors of [10] proposed a Stacked Denoising Auto-encoder Support Vector Machine (SDAE-SVM)-based model to detect threats in large-scale industrial networks. The authors used the KDD-CUP99 dataset for training and testing purposes. Their proposed system shows competitive strength against a diverse variety of potential security threats and achieved 97.83% accuracy. In [11], the authors employed Natural Language Processing (NLP) and Multi-Layer Perceptron (MLP) to differentiate between crucial and non-crucial posts on the Dark Web.

Another Deep Neural Network (DNN)-based DDoS attack detection framework is presented in [12]. The authors employed the CICDDoS2019 dataset for experimentation and achieved an accuracy of 94.57%. Further, a cognitive computing-based-IDS is proposed in [13]. The authors combined Gated Recurrent Unit (GRU) and Binary Bacterial Foraging Optimization (BBFO) for efficient intrusion detection. Their proposed scheme is trained and evaluated with the CICIDS2017 dataset and achieved an accuracy of 98.45%. The authors of [14] designed a Generative Adversarial Networks (GAN) based IDS. The authors trained their model using the CICIDS2017 dataset and achieved 88.70% accuracy. Another intrusion detection scheme using Aquila Optimizer (AQO) is proposed to combat botnet attacks in IoT-based smart environments. NSL-KDD and CICIDS2017 datasets are used for model training. The system significantly proves its effectiveness in terms of threat detection [15].

A DNN-based model is presented that introduces a pixel drop method to eliminate the existence of anomalies in medium to large-scale IoT-based smart networks. The framework analyzes the traffic streams to investigate suspicious entities and based on threat impressions; malicious traffic is highlighted [16]. The authors of [17] proposed an IDS for industrial environments. The authors used the power system and UNSW-NB15 dataset to evaluate the performance of their beta mixture-hidden Markov (MHMMs)-based model. The size of these datasets was reduced by the authors using Independent Component Analysis (ICA).

The authors of [18] proposed a model to detect intrusions in the IIoT network. They used UNSW-NB15 and BoT-IoT datasets for experimentation and achieved an accuracy of 91.25% for UNSW-NB15 and 98.10% for the BoT-IoT dataset. A hybrid DL autoencoder MLP along with the capabilities of automatic feature extraction is employed by the authors in [19]. They used the CICDDoS2019 dataset for experimentation and achieved a detection rate of 98.34%. Another DL-based IDS for SDN-based IoT networks is proposed in [20]. The authors utilized DNN with GRU-RNN to detect threats in such a network. Their proposed model achieved efficient results with 80.70% and 90% accuracy. However, their proposed scheme has a high FPR of 0.78%. The authors of [21] employed LSTM with fully-connected layers along with a hyper-parameters tuning method to identify normal and malicious events. They used six datasets to evaluate binary and multi-class intrusion detection scenarios. Moreover, in [22], the authors used a DNN-based scheme for identifying fraudulent activity in different IoT devices. They evaluated their proposed scheme under UNSW-NB15 and NSL-KDD datasets and achieved 92.40% and 98.60% detection rates.

III. Proposed Intrusion Detection System

In this section, we discuss the main components of the proposed explainable and resilient-centric deep learning-based DS for the Industry 5.0 network. We first describe the Proposed DL-based Cyber Threat Detection Scheme, followed by Connected Layers and Classifier. We further describe Explainable AI. Finally, we present the Proposed Network

TABLE I
TABLE OF NOTATION

Notation	**Description**
$\mathcal{I}p_t$	Input Gate
$\mathcal{F}g_t$	Forget Gate
$\mathcal{O}p_t$	Output Gate
$\mathcal{Z}_t$	Cell State
$\mathcal{Z}_t$	Candidate for the Cell State
$\mathcal{X}_t$	Current Input
$\mathcal{H}_{t-1}$	Previous Hidden State
σ	Sigmoid Function
$\mathcal{W}e$	Weight Matrix
$\mathcal{B}s$	Weight Bias
$\mathcal{U}p_t$	Update Gate
$\mathcal{R}e_t$	Reset Gate
$\mathcal{H}_t$	Final State
$\odot$	Point-wise Multiplication
$\rightarrow$	Forward Process
$\leftarrow$	Backward Process
t	Timestep

Model. The notations used in this work are mentioned in Table I.

A. Proposed DL-Based Cyber Threat Detection Scheme

1) BiLSTM: BiLSTM seems exclusively identical to its unidirectional counterpart (LSTM). The sole distinction is that the BiLSTM network connects with both the past and the future. For example, with synchronized repeat connections, a one-way LSTM may be trained to predict the dataset when it is loaded one at a time. On the rear tag, the BiLSTM additionally provides the following characters in succession, allowing us to access future information [23]. The BiLSTM consists of three gates, such that an input ($\mathcal{I}p_t$), forget ($\mathcal{F}g_t$), and output gate ($\mathcal{O}p_t$) along with a cell state ($\mathcal{Z}_t$) and a candidate for the cell state ($\mathcal{C}_t$). The $\mathcal{I}p_t$ keeps the state of the cell updated. The following equations control the operations to update the $\mathcal{C}_t$ for forward ($\rightarrow$) and backward ($\leftarrow$) process respectively [23]:

$$\overrightarrow{\mathcal{C}_t} = \tanh\left(\left(\overrightarrow{\mathcal{W}e_c\mathcal{H}_{t-1}}\right) * \left(\overrightarrow{\mathcal{W}_c\mathcal{X}_t}\right) + \overrightarrow{\mathcal{B}s_c}\right) \tag{1}$$

$$\overrightarrow{\mathcal{Z}_t} = \left(\overrightarrow{\mathcal{F}g_t\mathcal{H}_{t-1}}\right) + \left(\overrightarrow{\mathcal{I}p_t\mathcal{C}_t}\right) \tag{2}$$

$$\overrightarrow{\mathcal{I}p_t} = \alpha\left(\left(\overrightarrow{\mathcal{W}e_{ip}\mathcal{H}_{t-1}}\right) + \left(\overrightarrow{\mathcal{W}e_{ip}\mathcal{X}_t}\right) + \overrightarrow{\mathcal{B}s_{ip}}\right) \tag{3}$$

$$\overleftarrow{\mathcal{C}_t} = \tanh\left(\left(\overleftarrow{\mathcal{W}e_c\mathcal{H}_{t-1}}\right) * \left(\overleftarrow{\mathcal{W}_c\mathcal{X}_t}\right) + \overleftarrow{\mathcal{B}s_c}\right) \tag{4}$$

$$\overleftarrow{\mathcal{Z}_t} = \left(\overleftarrow{\mathcal{F}g_t\mathcal{H}_{t-1}}\right) + \left(\overleftarrow{\mathcal{I}p_t\mathcal{C}_t}\right) \tag{5}$$

$$\overleftarrow{\mathcal{I}p_t} = \alpha\left(\left(\overleftarrow{\mathcal{W}e_{ip}\mathcal{H}_{t-1}}\right) + \left(\overleftarrow{\mathcal{W}e_{ip}\mathcal{X}_t}\right) + \overleftarrow{\mathcal{B}s_{ip}}\right) \tag{6}$$

The $\mathcal{F}g_t$ takes the current input ($\mathcal{X}_t$) and the previous hidden state ($\mathcal{H}_{t-1}$) as inputs. Further, it uses the sigmoid function (σ) to output a value.

$$\overrightarrow{\mathcal{F}g_t} = \alpha\left(\left(\overrightarrow{\mathcal{W}e_f g\mathcal{H}_{t-1}}\right) + \left(\overrightarrow{\mathcal{W}e_f g\mathcal{X}_t}\right) + \overrightarrow{\mathcal{B}s_{fg}}\right) \tag{7}$$

$$\overleftarrow{\mathcal{F}g_t} = \alpha\left(\left(\overleftarrow{\mathcal{W}e_f g\mathcal{H}_{t-1}}\right) + \left(\overleftarrow{\mathcal{W}e_f g\mathcal{X}_t}\right) + \overleftarrow{\mathcal{B}s_{fg}}\right) \tag{8}$$

The $\mathcal{O}p_t$ determines the next timestep hidden state ($\mathcal{H}_t$) which comprise all the information of the prior inputs, thus it is required to make the predictions. Such a process requires two steps for finding the next timestamp:

$$\overrightarrow{\mathcal{O}p_t} = \alpha\left(\left(\overrightarrow{\mathcal{W}e_{op}\mathcal{H}_{t-1}}\right) + \left(\overrightarrow{\mathcal{W}e_{op}\mathcal{X}_t}\right) + \overrightarrow{\mathcal{B}s_{op}}\right) \tag{9}$$

$$\overrightarrow{\mathcal{H}_t} = \tanh\left(\overrightarrow{\mathcal{Z}_t}\right) * \overrightarrow{\mathcal{O}p_t} \tag{10}$$

$$\overleftarrow{\mathcal{O}p_t} = \alpha\left(\left(\overleftarrow{\mathcal{W}e_{op}\mathcal{H}_{t-1}}\right) + \left(\overleftarrow{\mathcal{W}e_{op}\mathcal{X}_t}\right) + \overleftarrow{\mathcal{B}s_{op}}\right) \tag{11}$$

$$\overleftarrow{\mathcal{H}_t} = \tanh\left(\overleftarrow{\mathcal{Z}_t}\right) \odot \overleftarrow{\mathcal{O}p_t} \tag{12}$$

where $\overrightarrow{\mathcal{W}e_c}$, $\overrightarrow{\mathcal{W}e_{ip}}$, $\overrightarrow{\mathcal{W}e_{fg}}$, $\overrightarrow{\mathcal{W}e_{op}}$ *and* $\overleftarrow{\mathcal{W}e_c}$, $\overleftarrow{\mathcal{W}e_{ip}}$, $\overleftarrow{\mathcal{W}e_{fg}}$, $\overleftarrow{\mathcal{W}e_{op}}$ are the weight matrices, while $\overrightarrow{\mathcal{B}s_c}$, $\overrightarrow{\mathcal{B}s_{ip}}$, $\overrightarrow{\mathcal{B}s_{fg}}$, $\overrightarrow{\mathcal{B}s_{op}}$ and $\overleftarrow{\mathcal{B}s_c}$, $\overleftarrow{\mathcal{B}s_{ip}}$, $\overleftarrow{\mathcal{B}s_{fg}}$, $\overleftarrow{\mathcal{B}s_{op}}$ are its respective biases. The $\mathcal{X}_t$ represents the current input and the Hadmard product is denoted by $\odot$.

2) BiGRU: A BiGRU consists of two GRUs; one processing the information in the forward direction and the other processing it backward. It consists of Update and reset gates($\mathcal{U}p_t$), ($\mathcal{R}e_t$) along with a candidate cell($\mathcal{C}_t$) and a final state($\mathcal{H}_t$). To prevent the RNN gradient disappearance or explosion, the gate structure might opt to save context information. The GRU has a simpler structure than the LSTM and trains more quickly. The following equations compute the BiGRU transition functions for the forward process ($\rightarrow$) [24].

$$\overrightarrow{\mathcal{U}p_t} = \sigma\left(\overrightarrow{\mathcal{W}e_{\mathcal{N}up}\mathcal{N}_t} + \overrightarrow{\mathcal{W}e_{hu}(\mathcal{H}_{t-1})} + \overrightarrow{\mathcal{B}s_{up}}\right) \tag{13}$$

$$\overrightarrow{\mathcal{R}e_t} = \sigma\left(\overrightarrow{\mathcal{W}e_{\mathcal{N}re}\mathcal{N}_t} + \overrightarrow{\mathcal{W}e_{hr}(\mathcal{H}_{t-1})} + \overrightarrow{\mathcal{B}s_{re}}\right) \tag{14}$$

$$\overrightarrow{\mathcal{C}_t} = \tanh\left(\overrightarrow{\mathcal{W}e_{\mathcal{N}c}\mathcal{N}_t} + \overrightarrow{\mathcal{R}e_t} \odot \overrightarrow{\mathcal{W}e_{hc}(\mathcal{H}_{t-1})} + \overrightarrow{\mathcal{B}s_c}\right) \tag{15}$$

$$\overrightarrow{\mathcal{H}_t} = \overrightarrow{\mathcal{U}p_t} \odot \overrightarrow{(\mathcal{H}_{t-1})} + \overrightarrow{\left(1 - \mathcal{U}p_t\right)} \odot \overrightarrow{\mathcal{C}_t} \tag{16}$$

where σ is the sigmoid operator, $\overrightarrow{\mathcal{H}_{t-1}}$, $\overleftarrow{\mathcal{H}_{t-1}}$ are the prior block hidden states, while $\overrightarrow{\mathcal{W}e_{\mathcal{N}up}}$, $\overrightarrow{\mathcal{W}e_{\mathcal{N}re}}$, $\overrightarrow{\mathcal{W}e_{\mathcal{N}c}}$ and $\overleftarrow{\mathcal{W}e_{\mathcal{N}up}}$, $\overleftarrow{\mathcal{W}e_{\mathcal{N}re}}$, $\overleftarrow{\mathcal{W}e_{\mathcal{N}c}}$ are the weight matrices for the current input $\overrightarrow{\mathcal{N}_t}$ and $\overleftarrow{\mathcal{N}_t}$. Further, $\overrightarrow{\mathcal{B}s_{up}}$, $\overrightarrow{\mathcal{B}s_{re}}$, $\overrightarrow{\mathcal{B}s_c}$ and $\overleftarrow{\mathcal{B}s_{up}}$, $\overleftarrow{\mathcal{B}s_{re}}$, $\overleftarrow{\mathcal{B}s_c}$ are its respective biases. Moreover, $\odot$ is the element-wise multiplication between the two vectors, and tanh represents the non-linear point-wise activation function. The following equations computes the transition functions for the backward process ($\leftarrow$):

$$\overleftarrow{\mathcal{U}p_t} = \sigma\left(\overleftarrow{\mathcal{W}e_{\mathcal{N}up}\mathcal{N}_t} + \overleftarrow{\mathcal{W}e_{hu}(\mathcal{H}_{t-1})} + \overleftarrow{\mathcal{B}s_{up}}\right) \tag{17}$$

$$\overleftarrow{\mathcal{R}e_t} = \sigma\left(\overleftarrow{\mathcal{W}e_{\mathcal{N}re}\mathcal{N}_t} + \overleftarrow{\mathcal{W}e_{hr}(\mathcal{H}_{t-1})} + \overleftarrow{\mathcal{B}s_{re}}\right) \tag{18}$$

$$\overleftarrow{\mathcal{C}_t} = \tanh\left(\overleftarrow{\mathcal{W}e_{\mathcal{N}c}\mathcal{N}_t} + \overleftarrow{\mathcal{R}e_t} \odot \overleftarrow{\mathcal{W}e_{hc}(\mathcal{H}_{t-1})} + \overleftarrow{\mathcal{B}s_c}\right) \tag{19}$$

$$\overleftarrow{\mathcal{H}_t} = \overleftarrow{\mathcal{U}p_t} \odot \overleftarrow{(\mathcal{H}_{t-1})} + \overleftarrow{\left(1 - \mathcal{U}p_t\right)} \odot \overleftarrow{\mathcal{C}_t} \tag{20}$$

Finally, the concatenation ($\oplus$) of the $\rightarrow$ and $\leftarrow$ is done by the following equation:

$$\mathcal{H}_t = \overrightarrow{\mathcal{H}_t} \oplus \overleftarrow{\mathcal{H}_t} \tag{21}$$

B. Connected Layers and Classifier for Threat Detection

The proposed threat detection module comprises two layers of BiLSTM having 200 and 100 neurons with 0.2% dropout rate to avoid overfitting followed by a dense layer of 30

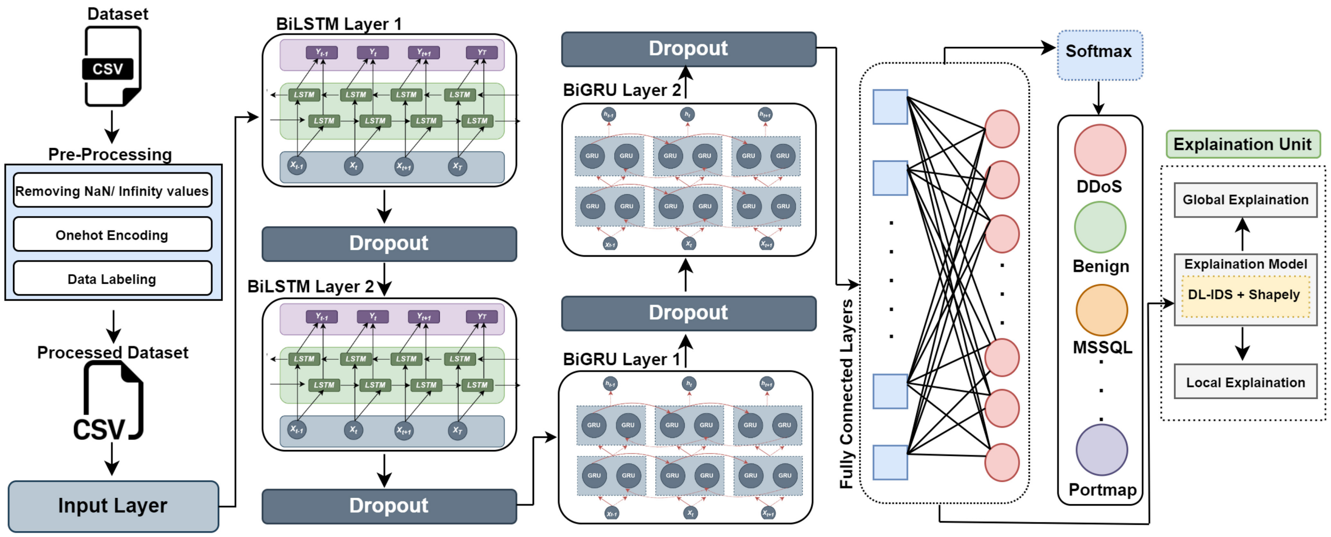


Fig. 1. Proposed Intrusion Detection System's Architecture.

Algorithm 1 Proposed DL-Based Intrusion Detection System

```
procedure INPUT:(Dataset (CICDDoS2019))
    OUTPUT: Attack classes
    Pre-process Dataset Removing NaN values Imputation of Infinity values Convert categorical features into numeric data Data Normalization (0 and 1)
    Divide CICDDoS2019 into $CDS19_{Train}$, $CDS19_{Val}$ and $CDS19_{Test}$.
    Perform Encoding and decoding
        Add BiLSTM layers and perform encoding using Equations (1) to (12)
    Build model using BiGRU and Softmax classifier using Equations (13) to (21)
        Add softmax layer
        $\sigma(\overrightarrow{\mathcal{D}})_i = \frac{e^{\mathcal{D}_i}}{\sum_{Z=1}^{\mathcal{K}} e^{\mathcal{D}_z}}$
        Calculate categorical cross-entropy loss
        $L(\hat{y}_c, y_c) = -\sum_{i=1}^{n}\sum_{c=1}^{C} y_c^{x_i} \cdot \log(p(\hat{y}_{ic} = y_{ic} \mid x_i))$
    Perform Testing using $CDS19_{Test}$
    Evaluate performance using various metrics
    Use SHAP library to analyze the features
end procedure
```

neurons. We further employed 2 layers of BiGRU with 100 and 50 neurons respectively. We adopt ADAM as an optimizer, while CC-E and RELU as activation and loss functions. A complete architecture of the proposed scheme is shown in Fig. 1 and the complete procedure of the proposed IDS is explained in Algorithm 1. Finally, in the output layer, we use Softmax classifier for attack classification. The following equations compute such operations:

$$\sigma\left(\overrightarrow{\mathcal{D}}\right)_i = \frac{e^{\mathcal{D}_i}}{\sum_{Z=1}^{\mathcal{K}} e^{\mathcal{D}_z}} \tag{22}$$

where σ is the softmax, $\mathcal{D}_i$ is input vector, $e^{\mathcal{D}_i}$ is the standard exponential function for the $\mathcal{D}_i$. Further, $\mathcal{K}$ represents the number of classes and $e^{\mathcal{D}_z}$ is the standard exponential function for the output vector respectively. Finally, we calculate the loss with categorical cross-entropy loss:

$$L(\hat{y}_c, y_c) = -\sum_{i=1}^{n}\sum_{c=1}^{C} y_c^{x_i} \cdot \log\left(p\left(\hat{y}_{ic} = y_{ic} \mid x_i\right)\right) \tag{23}$$

where y_c is actual and $\hat{y}_c$ is predicted output, x is the pattern of input sequence, n is the number of observations, and p belongs to a specific threat type y.

C. Explainable AI

DL-based models are getting popularity in safety-critical IoT applications and the demand for justifications for their predictions is rising [25]. The XAI provides methodical and comprehensible justifications for its behavior that human users can follow. Many ML-based models, i.e., NB, LR, and DT are fundamentally understandable on a modular level [26]. Unlike ML-based models, the DL models provide superior performance but these models are unable to interpret their predictions. Understanding the rationale behind a model's decision for users and stakeholders helps build trust and confirms that the model is solving an issue securely and robustly. One of the reasons for the "black-box" DL model's hesitant acceptance in many safety-critical sectors is their lack of transparency. Thus, scholars have been looking into numerous explainability methods to aid users in interpreting the decisions of black-box models. Some of them are as follows: 1) Text Explanations: By computing a relevance score for the model's controlled variables, this method is utilized to explain the intricate internal workings of the model. 2) Local Explanations: it is used for measuring a model's reaction to small modifications for building explanations. 3) Explanations using representative examples: The training data and its effect on a model's decision are better understood using this method. 4) Visual Explanations: The model's behavior is visualized using the visual explanation technique. It is used to provide captions for images that explain why they belong to a certain class in image classification tasks. SHAP is one of the approaches

TABLE II
SYSTEM SPECIFICATIONS FOR EXPERIMENTATION

System Aspect	**Specification**
Central Processing Unit (CPU)	Corei7(2.260GHz)
Graphics Processing Unit (GPU)	Geforce RTX 2060
Operating system (OS)	Windows 10
Random Access Memory (RAM)	32GB
language	Python
Libraries	sklearn, Pandas, numpy

TABLE III
DETAILS OF $\mathcal{TPR}$, $\mathcal{TNR}$, $\mathcal{FPR}$ & $\mathcal{FNR}$

Terms	**Description**
$\mathcal{TPR}$	*Positive occurrences that have been properly determined by the model.*
$\mathcal{TNR}$	*Negative instances that the model adequately characterised.*
$\mathcal{FPR}$	*The model maps negative data as positive.*
$\mathcal{FNR}$	*Positive instances are designated as negative instances by the model.*

that has been proposed for relevance explanations [27]. In this paper, we explain the importance of features in the decision of proposed DL-based IDS by employing the SHAP framework.

IV. EXPERIMENTAL SETUP

This section presents the experimental design followed by the dataset details, pre-processing, and evaluation metrics.

1) Experimental Design: The experiments are performed using a Legion PC with a 2.60 GHz Hexacore Coffee Lake CPU, 32 GB RAM, and a Geforce RTX 2060 Max-Q 8GB GPU. The proposed threat detection scheme is developed through the Keras library of TensorFlow. Further, we have employed Python to run the implementation scripts. Complete details are provided in Table II.

2) Dataset: The proposed work used the CICIDDoS2019 [28] for experimentation purposes. The dataset contains modern reflective DDoS attack types, i.e., MSSQL, SSDP, UDP, SYN, UDP-Lag, Portmap, and WebDDoS attacks. This work divides the dataset into training, validation and testing sets, i.e., 60%, 10%, and 30%.

3) Dataset Pre-Processing: First, we removed all rows with NaN and Infinity values because they could affect the model's performance. We further used Sklearn label encoder to convert all non-numerical values to numerical values. The only non-numerical feature in the dataset is the *'Label'*, which we converted to binary using the Scikit-learn label encoder. Moreover, MinMax scalar function is employed for data normalization [29].

4) Evaluation Metrics: This work evaluates the proposed threat detection scheme by employing the standard evaluation metrics, such that Receiver Operating Characteristic ($\mathcal{ROC}$) curve, Confusion Matrix ($\mathcal{CM}$), Accuracy ($\mathcal{ACC}$), Recall ($\mathcal{RE}$), Precision ($\mathcal{PR}$), F1-score ($\mathcal{F}1$), and extended evaluation metrics, i.e., $\mathcal{TPR}$, $\mathcal{NPV}$, $\mathcal{TNR}$, $\mathcal{FPR}$, $\mathcal{FNR}$, $\mathcal{FDR}$, and $\mathcal{FOR}$. The extended evaluation metrics are defined in Table III. While the following equations compute the value of $\mathcal{ACC}$, $\mathcal{PR}$, $\mathcal{RE}$, and $\mathcal{F}1$.

1) $\mathcal{ACC}$: The effectively predicted instances over the complete number of instances.

$$\mathcal{ACC} = \frac{\mathcal{TP} + \mathcal{TN}}{\mathcal{TP} + \mathcal{TN} + \mathcal{FP} + \mathcal{FN}} \tag{24}$$

2) $\mathcal{PR}$: It is the extent of positives that are genuine positives.

$$\mathcal{PR} = \frac{\mathcal{TP}}{\mathcal{TP} + \mathcal{FP}} \tag{25}$$

3) $\mathcal{RE}$: The ratio of $\mathcal{TP}$ to the sum of $\mathcal{TP}$ and $\mathcal{FN}$.

$$\mathcal{RE} = \frac{\mathcal{TP}}{\mathcal{TP} + \mathcal{FN}} \tag{26}$$

4) $\mathcal{F}1$: The harmonic mean of the $\mathcal{RE}$ and $\mathcal{PR}$. The $\mathcal{F}1$ is determined utilizing the underneath numerical condition.

$$\mathcal{F}1 = 2 * \frac{\mathcal{PR} * \mathcal{RE}}{\mathcal{PR} + \mathcal{RE}} \tag{27}$$

V. RESULT ANALYSIS

In this section, we discuss the simulation results and performance analysis of the proposed intrusion detection scheme.

A. Performance Analysis of Proposed Threat Detection Framework

In this subsection, we discuss the efficiency of the proposed IDS. The proposed DL-based threat detection model has efficiently learned from the dataset as proven by the accuracy vs loss in Fig. 2. The model achieved Validation $\mathcal{ACC}$ of 99.77% with a validation loss of 0.0055% with 10 epochs. We also measure the performance of the proposed IDS class-wise in terms of $\mathcal{ACC}$, $\mathcal{PR}$, $\mathcal{TPR}$, and $\mathcal{FNR}$. The model has significantly learned the normal and attack signatures and achieved $\mathcal{ACC}$, $\mathcal{PR}$ and $\mathcal{TPR}$ values between 92% to 100% except for the DDoS class, where the model achieved 78.63% $\mathcal{TPR}$ as depicted in Table IV. Further, the model achieved $\mathcal{FNR}$ of 0.00012% to 0.0213% for the respective classes accordingly. We further provide the $\mathcal{CM}$ and $\mathcal{ROC}$ of the proposed model to prove its efficacy. Table V depicts the $\mathcal{CM}$ of the model where it demonstrates its efficiency by categorizing all instances of the datasets into their respective classes. Similarly, the $\mathcal{ROC}$ curve values derived for various attack classes is illustrated in Fig. 3. It shows that the scores for all the classes are almost equal to one. Moreover, the proposed model achieved macro and micro-average of 0.99 and 1.00 respectively.

B. XAI Interpretation for CICDDoS2019 Dataset

In this subsection, we discuss the XAI interpretation of the dataset. The demonstration of decision-making by complex models is illustrated via the SHAP decision graphs. SHAP provides a number of plots, i.e., Decision Plot ($\mathcal{DP}$), Waterfall Plot ($\mathcal{WP}$) and Summary Plot ($\mathcal{SP}$). The $\mathcal{DP}$ plot is given in Fig. 4. The explainer's expected value is used to center the

TABLE IV
CLASS-WISE RESULT ANALYSIS OF THE PROPOSED IDS

Parameters	*Benign*	*MSSQL*	*SSDP*	*DDoS*	*SYN*	*WebDDoS*	*Portmap*	*UDP*	*UDP-Lag*
$\mathcal{ACC}$	99.98%	100%	99.03%	99.03%	99.98%	99.99%	100%	99.99%	100%
$\mathcal{PR}$	99.98%	100%	92.30%	98.31%	99.70%	99.90%	100%	100%	100%
$\mathcal{TPR}$	99.98%	100%	98.45%	78.63%	100%	99.90%	100%	99.89%	100%
$\mathcal{FNR}$	0.000013%	0.000125%	0.001541%	0.021369%	0.000110%	0.000956%	0.00014%	0.000102%	0.00025%

TABLE V
CONFUSION MATRIX OF PROPOSED IDS

Actual Predicted	Benign	MSSQL	SSDP	DrDoS	SYN	WebDDoS	Portmap	UDP	UDP_Lag
Benign	**48266**	0	0	35	0	10	0	0	11
MSSQL	0	**2701**	0	0	2	5	0	0	8
SSDP	0	0	**9367**	0	0	0	015	0	0
DDoS	0	0	0	**1691**	0	6	0	2	0
SYN	0	1	0	0	**1673**	0	0	0	4
WebDDoS	9	0	0	0	0	**2309**	0	0	0
Portmap	0	0	3	0	20	0	**2984**	0	0
UDP	0	0	11	0	6	1	0	**1316**	0
UDP_Lag	0	2	0	0	4	0	0	0	**1708**

TABLE VI
CLASS-WISE DETECTION ACCURACY COMPARISON OF THE PROPOSED IDS AGAINST BASELINE DETECTION SCHEMES

Schemes	*Benign*	*MSSQL*	*SSDP*	*DDoS*	*SYN*	*WebDDoS*	*Portmap*	*UDP*	*UDP-Lag*
RF	100%	98.59%	98.50%	97.68%	95.56%	96.59%	99.15%	99.26%	99.17%
DT	99.93%	98.83%	98.87%	98.14%	97.89%	98.67%	99.23%	99.14%	99.59%
Proposed IDS	100%	100%	99.03%	99.11%	99.98%	99.99%	100%	99.99%	100%

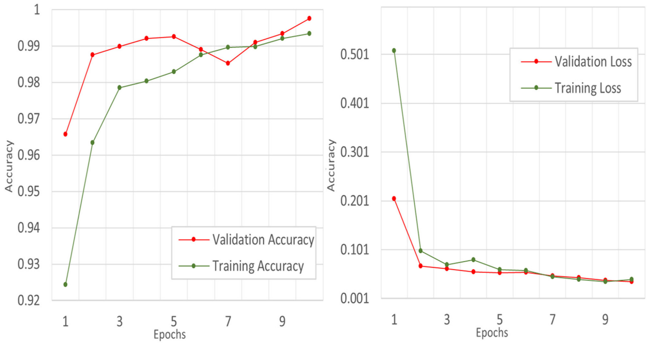


Fig. 2. Accuracy vs Loss of Proposed IDS.

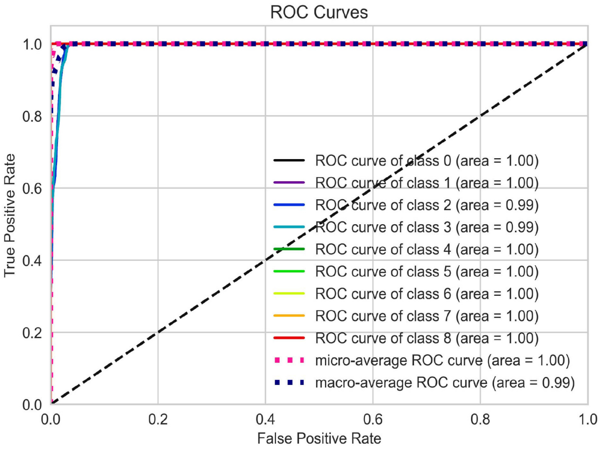


Fig. 3. ROC of Proposed IDS.

plot on the X-axis. Similar to how the effects of the linear model are relative to the intercept, every SHAP value ($\mathcal{SV}$) is related to the estimated value ($\mathcal{EV}$) of the model. The Y-axis represents the model features. Fig. 5 depicts the ($\mathcal{WP}$) of the dataset. It demonstrates how this kind of representation behaves for $\mathcal{TP}$, $\mathcal{FP}$, $\mathcal{FN}$ and $\mathcal{TN}$. The blue and red bars of the $\mathcal{WP}$ plot depicts the features and contributes to the overall classification score. Further, it can decrease or increase the classification score. Finally, Fig. 6 represents the $\mathcal{SP}$ plot of the dataset. With the most essential features at the top and the least important at the bottom, the $\mathcal{SP}$ plots are a Visualization Plot ($\mathcal{VP}$) for global interpretation that conveys the value of global features across the entire model. Since they have a greater average influence on the output of the model, the features in this plot with big absolute $\mathcal{SV}$ are classified as significant.

C. Comparison With Baseline Detection Schemes

In this subsection, we have conducted the performance comparison of the proposed model against the baseline detection schemes. We have used the results obtained from the proposed threat detection scheme for comparison with GRU and LSTM. Fig. 7 depicts the values of $\mathcal{ACC}$, $\mathcal{PR}$, $\mathcal{RE}$, and $\mathcal{F}1$ achieved by the Proposed IDS as 99.77%, 98.79%, 98.42%, and 99.41% respectively.

Table VI depicts the class-wise Detection $\mathcal{ACC}$ comparison. The Proposed IDS achieved a detection $\mathcal{ACC}$ of 97% to 100% for MSSQL, SSDP, DDoS, Portmap, UDP, UDP-Lag, and Benign. However, it obtained 95.56% detection $\mathcal{ACC}$ for SYN and 96.59% for WebDDoS attack classes. Fig. 8 depicts the values of $\mathcal{TPR}$, $\mathcal{TNR}$, and $\mathcal{NPV}$, where the proposed IDS achieved $\mathcal{TPR}$ of 97.42% and 99.88% $\mathcal{TNR}$ and

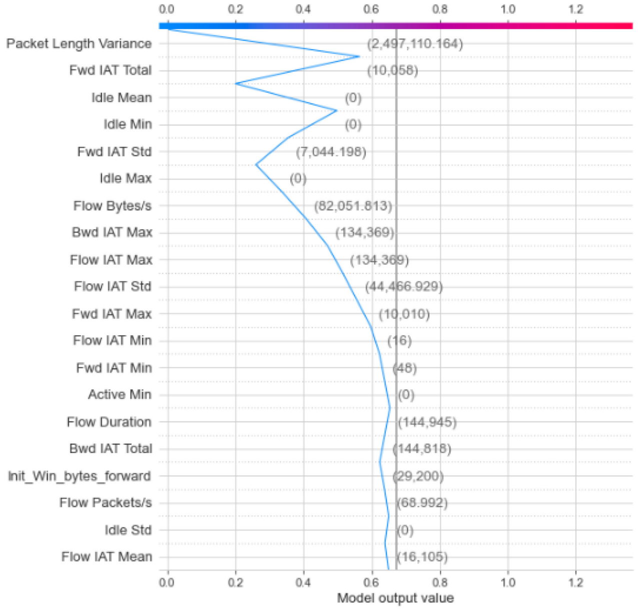


Fig. 4. SHAP values for CICDDoS2019 dataset using decision plot.

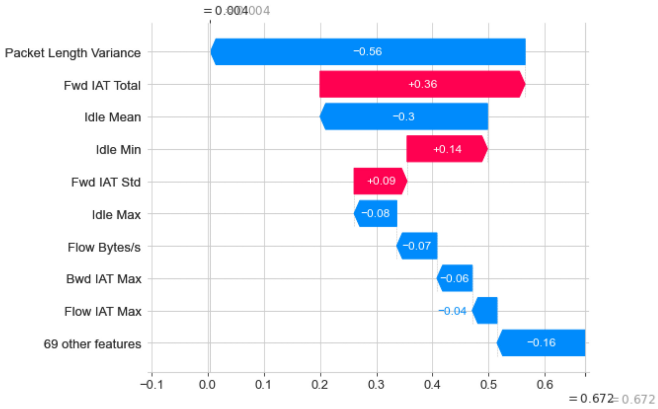


Fig. 5. SHAP values for CICDDoS2019 dataset using waterfall plot.

$\mathcal{NPV}$. However, the GRU and LSTM show less significant performance, which proves the superiority of the Proposed IDS against the baseline detection schemes. Finally, Fig. 9 depicts the comparison in terms of $\mathcal{FPR}$, $\mathcal{FNR}$, $\mathcal{FDR}$, and $\mathcal{FOR}$. It is shown that the Proposed IDS achieved $\mathcal{FPR}$, $\mathcal{FDR}$, $\mathcal{FOR}$ of 0.0041%, 0.0025%, 0.0004% with $\mathcal{FNR}$ of 0.0485% respectively. The values of the Proposed IDS are considerably lower than the other detection models. The lower rates of such metrics prove the efficiency of the Proposed IDS.

D. Comparison With Recent State-of-the-Art Detection Frameworks

Lastly, we compare the Proposed IDS's performance with recent threat detection schemes from the current literature, i.e., [10], [11], [12], [19] and [17] to further validate its efficacy. Table VII depicts the comparison in terms of $\mathcal{ACC}$. Some of the recent works have either used old datasets, i.e., Power system and KDD-CUP99, which have less practical values for IoT or they achieved less significant outcomes. We adopt CICDDoS2019, which contains network flow-based instances and is an IoT-based dataset. The Proposed IDS outperformed the recent detection frameworks by achieving a higher $\mathcal{ACC}$ of more than 2%.

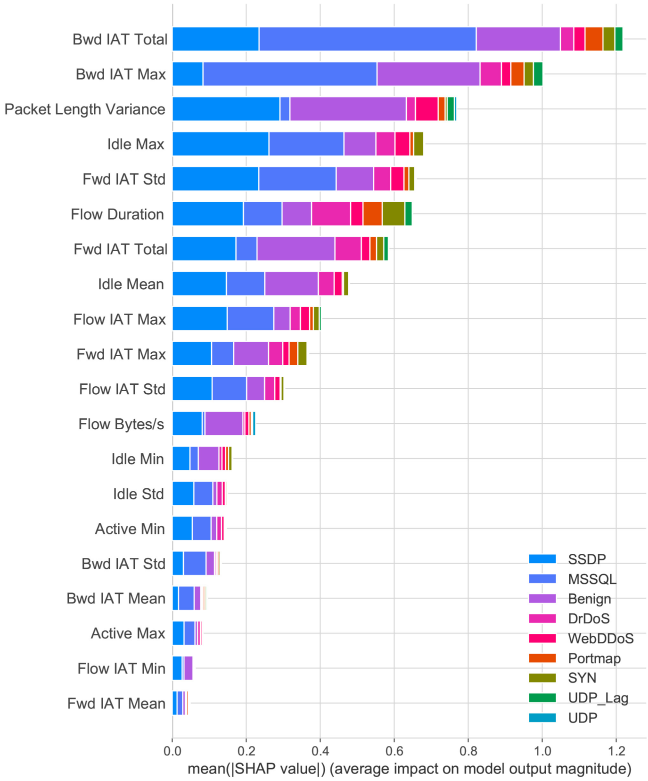


Fig. 6. SHAP values for CICDDoS2019 dataset using summary plot.

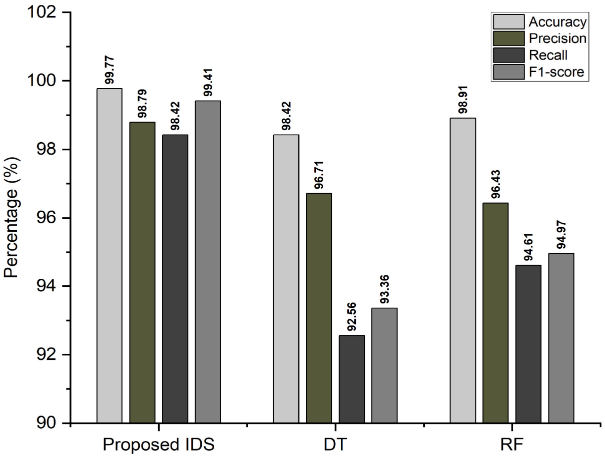


Fig. 7. Overall comparison of Proposed IDS against baseline detection schemes.

TABLE VII
COMPARISON OF PROPOSED IDS WITH RECENT FRAMEWORKS

Ref	*Year*	*Scheme*	*Dataset*	*ACC*
Proposed work	2023	Proposed IDS	CICDDoS2019	99.77%
[12]	2021	DNN	CICDDoS2019	94.57%
[19]	2021	DLAMLP	CICIDDoS2019	98.34%
[10]	2020	SDAE SVM	KDD-CUP99	97.83%
[11]	2019	MLP	NA	79.40%
[17]	2018	MHMMs	Power system	96.32%

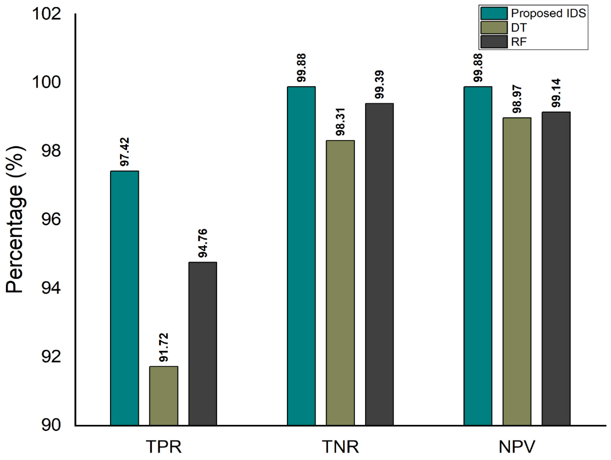


Fig. 8. $\mathcal{TPR}$, $\mathcal{TNR}$ and $\mathcal{NPV}$ comparison of Proposed IDS against baseline detection schemes.

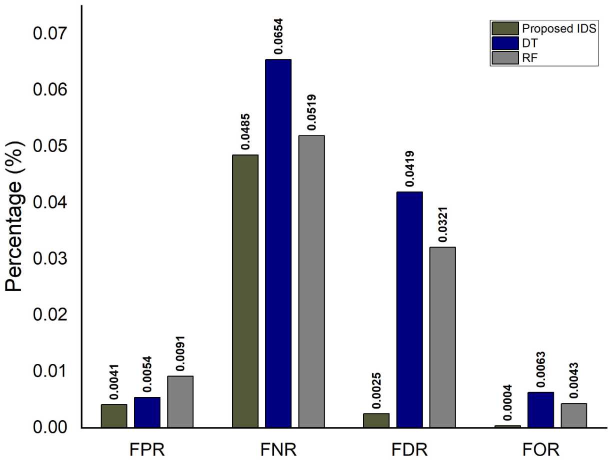


Fig. 9. Comparison of the Proposed IDS against baseline detection schemes in terms of $\mathcal{FPR}$, $\mathcal{FNR}$, $\mathcal{FDR}$ and $\mathcal{FOR}$.

VI. Conclusion

An intrusion detection system is one of the most important security tool for industrial networks. However, most of the existing approaches based on ML and DL techniques are treated as a black box by the security analysts and developers. In this article, we have designed a new explainable and resilient intrusion detection system in Industry 5.0 that combines bidirectional long short-term memory networks, a bidirectional-gated recurrent unit, a fully connected layer, and a softmax classifier for attack detection. Furthermore, the proposed framework adopts SHAP technique to understand the importance of the features that contributed the most to attack detection using the CICDDoS2019 dataset. The experimental results confirm the superiority of the proposed approach over some existing state-of-the-art schemes. However, the proposed IDS has some limitations, such that, it is vulnerable to insider threats where intruders can disrupt the network without interfering with the flow between the industrial network and the Internet. Future research will include integrating blockchain with the proposed framework to enhance decentralization in Industry 5.0.

References

[1] T. Guo, K. Yu, X. Cheng, and A. K. Bashir, "Robust electronic nose in industrial cyber physical systems based on domain adaptive subspace transfer model," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2021, pp. 1–6.

[2] F. Ding, G. Zhu, M. Alazab, X. Li, and K. Yu, "Deep-learning-empowered digital forensics for edge consumer electronics in 5G HetNets," *IEEE Consum. Electron. Mag.*, vol. 11, no. 2, pp. 42–50, Mar. 2022.

[3] M. Wazid, A. K. Das, and S. Shetty, "BSFR-SH: Blockchain-enabled security framework against Ransomware attacks for smart healthcare," *IEEE Trans. Consum. Electron.*, vol. 69, no. 1, pp. 18–28, Feb. 2023.

[4] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for IoT security based on learning techniques," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2671–2701, 3rd Quart., 2019.

[5] L. N. Tidjon, M. Frappier, and A. Mammar, "Intrusion detection systems: A cross-domain overview," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3639–3681, 4th Quart., 2019.

[6] P. M. Dassanayake, A. Anjum, A. K. Bashir, J. Bacon, R. Saleem, and W. Manning, "A deep learning based explainable control system for reconfigurable networks of edge devices," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 7–19, Jan./Feb. 2022.

[7] A. Nguyen et al., "System design for a data-driven and explainable customer sentiment monitor using IoT and enterprise data," *IEEE Access*, vol. 9, pp. 117140–117152, 2021.

[8] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics," *IEEE Netw.*, vol. 34, no. 4, pp. 126–132, Jul./Aug. 2020.

[9] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, "Anomaly-based intrusion detection system for IoT networks through deep learning model," *Comput. Elect. Eng.*, vol. 99, Apr. 2022, Art. no. 107810.

[10] Z. Lv, L. Qiao, J. Li, and H. Song, "Deep-learning-enabled security issues in the Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9531–9538, Jun. 2021.

[11] M. Kadoguchi, S. Hayashi, M. Hashimoto, and A. Otsuka, "Exploring the dark web for cyber threat intelligence using machine leaning," in *Proc. IEEE Int. Conf. Intell. Secur. Inf. (ISI)*, 2019, pp. 200–202.

[12] A. E. Cil, K. Yildiz, and A. Buldu, "Detection of DDoS attacks with feed forward based deep neural network model," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114520.

[13] M. M. Althobaiti, K. P. M. Kumar, D. Gupta, S. Kumar, and R. F. Mansour, "An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems," *Measurement*, vol. 186, Dec. 2021, Art. no. 110145.

[14] H. Jeong, J. Yu, and W. Lee, "Poster abstract: A semi-supervised approach for network intrusion detection using generative adversarial networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–2.

[15] A. Fatani, A. Dahou, M. A. Al-Qaness, S. Lu, and M. Abd Elaziz, "Advanced feature extraction and selection approach using deep learning and Aquila optimizer for IoT intrusion detection system," *Sensors*, vol. 22, no. 1, p. 140, 2022.

[16] H. Qiu, Q. Zheng, T. Zhang, M. Qiu, G. Memmi, and J. Lu, "Toward secure and efficient deep learning inference in dependable IoT systems," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3180–3188, Mar. 2021.

[17] N. Moustafa, E. Adi, B. Turnbull, and J. Hu, "A new threat intelligence scheme for safeguarding industry 4.0 systems," *IEEE Access*, vol. 6, pp. 32910–32924, 2018.

[18] M. Abdel-Basset, V. Chang, H. Hawash, R. K. Chakrabortty, and M. Ryan, "Deep-IFS: Intrusion detection approach for Industrial Internet of Things traffic in fog environment," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7704–7715, Nov. 2021.

[19] Y. Wei, J. Jang-Jaccard, F. Sabrina, A. Singh, W. Xu, and S. Camtepe, "AE-MLP: A hybrid deep learning approach for DDoS detection and classification," *IEEE Access*, vol. 9, pp. 146810–146821, 2021.

[20] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, M. Ghogho, and F. El Moussa, "DeepIDS: Deep learning approach for intrusion detection in software defined networking," *Electronics*, vol. 9, no. 9, p. 1533, 2020.

[21] S. K. Sahu, D. P. Mohapatra, J. K. Rout, K. S. Sahoo, Q.-V. Pham, and N.-N. Dao, "A LSTM-FCNN based multi-class intrusion detection using scalable framework," *Comput. Elect. Eng.*, vol. 99, Apr. 2022, Art. no. 107720.

[22] A.-H. Muna, N. Moustafa, and E. Sitnikova, "Identification of malicious activities in Industrial Internet of Things based on deep learning models," *J. Inf. Secur. Appl.*, vol. 41, pp. 1–11, Aug. 2018.

[23] I. Ullah and Q. H. Mahmoud, "Design and development of RNN anomaly detection model for IoT networks," *IEEE Access*, vol. 10, pp. 62722–62750, 2022.

[24] T. Jinbao, K. Weiwei, C. Yidan, T. Qiaoxin, S. Chenyuan, and L. Long, "Text classification method based on BiGRU-attention and CNN hybrid model," in *Proc. 4th Int. Conf. Artif. Intell. Pattern Recognit.*, 2021, pp. 614–622.

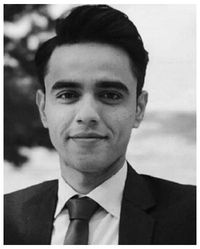
[25] D. Gunning, "Explainable artificial intelligence (xai)," *Defense Adv. Res. Projects Agency*, vol. 2, no. 2, p. 1, 2017.

[26] C. Molnar, *Interpretable Machine Learning*. Morrisville, NC, USA: Lulu Press, 2020.

[27] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4765–4774.

[28] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, 2019, pp. 1–8.

[29] D. Javeed, T. Gao, and M. T. Khan, "SDN-enabled hybrid DL-driven framework for the detection of emerging cyber threats in IoT," *Electronics*, vol. 10, no. 8, p. 918, 2021.



Danish Javeed (Student Member, IEEE) received the M.E. degree in computer applied technology from the Changchun University of Science and Technology, China, under the prestigious fellowship of Ministry of Education funded by the Government of China in 2020. He is currently pursuing the Ph.D. degree in software engineering, specializing in information security, with the Software College, Northeastern University, China, under the prestigious fellowship of Ministry of Education funded by the Government of China. He has authored or coauthored over ten publications in high-ranked journals and conferences. He has many research contributions in the area of deep learning, cybersecurity, intrusion detection and prevention system, Internet of Things, software-defined networking, and edge computing.

Tianhan Gao received the B.E. degree in computer science and technology and the M.E. and Ph.D. degrees in computer application technology from Northeastern University, China, in 1999, 2001, and 2006, respectively. He started as a Lecturer with the Software College, Northeastern University in 2006, and was quickly promoted to an Associate Professor in 2010. From 2011 to 2012, he was a Visiting Scholar with the Department of Computer Science, Purdue University. He has authored or coauthored over 60 research publications. His key research interests are next-generation network security, security and privacy in ubiquitous computing, and virtual/augmented reality. In 2016, he was awarded the Title of Doctoral Tutor.

Prabhat Kumar (Member, IEEE) received the Ph.D. degree in information technology from the National Institute of Technology Raipur, Raipur, India, under the prestigious fellowship of Ministry of Human Resource and Development funded by the Government of India in 2022. He worked with the Indian Institute of Technology Hyderabad, India, as a Postdoctoral Researcher under project "Development of Indian Telecommunication Security Assurance Requirements for IoT Devices." He is currently working as a Postdoctoral Researcher with the Department of Software Engineering, LUT School of Engineering Science, LUT University, Lappeenranta, Finland. He has authored or coauthored over 35 publications in high-ranked journals and conferences, including 13+ IEEE TRANSACTIONS paper. He has many research contributions in the area of machine learning, deep learning, federated learning, big data analytics, cybersecurity, blockchain, cloud computing, Internet of Things, and software-defined networking. One of his Ph.D. publication was recognized as a top-cited article by Wiley in 2020–2021.

Alireza Jolfaei (Senior Member, IEEE) is an Associate Professor of Networking and Cyber Security with the College of Science and Engineering, Flinders University, Adelaide, Australia. Before Flinders University, he was a Faculty Member with Macquarie University, Federation University, and Temple University, Philadelphia, PA, USA. He has published over 100 papers, which appeared in peer-reviewed journals, conference proceedings, and books. His main research interest is in cyber–physical systems security. He received the prestigious IEEE Australian Council Award for his research paper published in the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is the IEEE Consumer Technology Publication Board Member and the Editor-in-Chief of the Consumer Technology Society World Newsletter. He has served as the Regional Chair of the IEEE Technology and Engineering Management Society's Membership Development and Activities for Australia. He has served as a Program Co-Chair, a Track Chair, a Session Chair, and a Technical Program Committee Member for major conferences, including IEEE TrustCom and IEEE ICCCN. He is a Distinguished Speaker of ACM.