international journal of critical infrastructure protection 45 (2024) 100674

Contents lists available at ScienceDirect

International Journal of Critical Infrastructure Protection

journal homepage: www.elsevier.com/locate/ijcip

ResNet50-1D-CNN: A new lightweight resNet50-One-dimensional convolution neural network transfer learning-based approach for improved intrusion detection in cyber-physical systems

Yakub Kayode Saheed [a,*], Oluwadamilare Harazeem Abdulganiyu [b], Kaloma Usman Majikumna [c], Musa Mustapha [d], Abebaw Degu Workneh [e]

[a] School of IT & Computing, American University of Nigeria, Nigeria
[b] University of Maiduguri, Nigeria
[c] Computer Engineering Department, University of Maiduguri, Nigeria
[d] Department of Computer Science, Yobe State University, Nigeria
[e] Department of Software Engineering, Debre Berhan University, Ethiopia

ARTICLE INFO

Keywords:
Cyber-physical systems (CPS)
Intrusion detection systems
Transfer learning
One-dimensional CNN (1D-CNN)
Resnet-50
Adaptive Gradient
Control System HAI dataset
Feature Importance

ABSTRACT

The cyber-physical system (CPS) plays a crucial role in supporting critical infrastructure like water treatment facilities, gas stations, air conditioning components, and smart grids, which are essential to society. However, these systems are facing a growing susceptibility to a wide range of emerging attacks. Cyber-attacks against CPS have the potential to cause disruptions in the accurate sensing and actuation processes, resulting in significant harm to physical entities and posing concerns for the overall safety of society. Unlike common security measures like firewalls and encryption, which often aren't enough to deal with the unique problems that CPS architectures present, deploying machine learning-based intrusion detection systems (IDS) that are specifically made for CPS has become an important way to make them safer. The application of machine learning algorithms has been suggested as a means of mitigating cyber-attacks on CPS. However, the limited availability of labelled data pertaining to emerging attack techniques poses a significant challenge to the accurate detection of such attacks. In the given scenario, transfer learning emerges as a promising methodology for the detection of cyber-attacks, as it involves the implicit modelling of the system. In this research, we propose a new lightweight transfer learning method via ResNet50-CNN1D for intrusion detection in CPS. The Adaptive Gradient (Adagrad) optimizer was applied in the proposed model to minimize the loss function through the adjustment of network weight. We tested how well the suggested ResNet50-1D-CNN model worked using the UNSW-NB15 dataset and a control system dataset called HAI. The HAI dataset was taken from the testbed and based on a planned physical attack scenario. By calculating the coefficient scores for the top ten (10) features in the HAI and UNSW-NB15 data, it was possible to determine the relevance of a feature. The rationale behind employing transfer learning was to mitigate the complexity associated with the classification of cyber-attacks and runtime. The utilization of transfer learning resulted in notable reductions in both the training and testing times required for the detection of attacks. On the HAI data, the results showed an accuracy of 97.32 %, recall of 98.41 %, F1-score of 96.32 %, and precision of 97.09 %. On the UNSW-NB15 data, the results showed an accuracy of 99.89 %, recall of 99.09 %, F1-score of 98.01 %, and precision of 98.70 %.

1. Introduction

Cyber-Physical Systems (CPS) are complex and interconnected systems that are geographically dispersed, federated, and diverse. These systems play a crucial role in various domains, including sensors, actuators, control mechanisms, and networking components. The combination of sensor technologies, communication systems, and control approaches has been extensively studied by both academia and industry in the field of CPS [1]. Examples of CPS include pervasive healthcare systems, first responder alerting systems, unmanned aircraft systems,

* Corresponding author.
E-mail address: yakubu.saheed@aun.edu.ng (Y.K. Saheed).

https://doi.org/10.1016/j.ijcip.2024.100674
Received 31 December 2023; Received in revised form 18 March 2024; Accepted 27 March 2024
Available online 2 April 2024

and smart grids [2]. The aforementioned systems exhibit a multitude of control loops, stringent timing constraints, predictable patterns of network traffic, components that have been in use for a considerable period of time, and include elements of wireless networks. CPS integrates the cyber realm, which includes commodity servers and network components, with the physical domain, which encompasses sensors and actuators. The attack model for CPS incorporates both long- and short-duration attacks. An enemy with a lack of caution has the ability to infiltrate the network and promptly interrupt the relevant processes, resulting in a catastrophic event [3]. Conversely, a more advanced opponent may exhibit caution in order to avoid disrupting regular system functioning [4], with the intention of orchestrating a coordinated attack launched simultaneously from several points. Existing studies [5] identified the Stuxnet attack methodology. The design of intrusion detection systems (IDS) is a critical challenge in the field of CPS [6]. The primary objective of designing CPS-IDS is to effectively utilize its distinctive characteristics in order to identify and mitigate previously unidentified attacks. The primary objective of designing CPS-IDS is to effectively utilize its distinctive characteristics in order to identify and mitigate previously unidentified attacks. An IDS is essential for ensuring the security of a CPS used in public infrastructure such as gas pipelines, power systems, water stations, rail roads, smart grids, and water treatment. Fig. 1 illustrates the composition of a supervisory control and data acquisition (SCADA)-based CPS consisting of three layers: the physical layer, the cyber-physical system layer, and the operation/corporate layer [7]. The physical layer comprises programmable logic controllers (PLC) and remote terminal units (RTU) that gather data from sensors and carry out commands. The cyber-physical system layer oversees and manages devices in the physical layer using IDS. The operational layer relies on an IT system with IDS to facilitate remote business processes and administration. According to Karnouskos in 2011, the cyber components of CPS, such as SCADA, PLC, and human-machine interface (HMI), are susceptible to various malicious attacks. Fig. 1 illustrates that attackers can carry out attacks by exploiting the absence of access control in CPS and the software vulnerabilities present in PLCs, RTUs, and SCADA systems.

The utilization of network-related services has witnessed a significant rise in recent years, leading to a corresponding growth in the volume of sensitive data present on the internet [8]. Networks are susceptible to intrusions in which individuals, without authorization and with malicious intentions, gain access to a system inside a network and endeavor to acquire sensitive information or disrupt the functioning of the system [9]. Despite the implementation of various network security measures, instances of cyberattacks continue to persist. The prevention of network infiltration is a significant challenge, necessitating the implementation of NIDSs [10]. The analysis of packet data during attacks can provide valuable insights for the detection of similar attacks in subsequent instances [11]. When constructing a NIDS, it is crucial to consider that misclassifying an intrusion as a non-intrusion poses a greater risk than erroneously identifying a non-intrusion as an intrusion [12]. CPS effectively merges computational capabilities with physical processes through the utilization of feedback loops, facilitated by the employment of network services [13]. These systems are comprised of several components, such as actuators, sensors, and other elements that establish communication among themselves via a network. The communication network employs similar protocols to those utilized in computer networks, namely at lower levels such as TCP/IP and wireless protocols. Therefore, cyber-physical systems are susceptible to the same attacks as those encountered in a basic computer network. Nevertheless, it is imperative to acknowledge that cyber-physical systems possess a significant level of importance in terms of safety. The occurrence of abrupt malfunctions, whether resulting from cyber-attacks or other causes, can lead to substantial harm not only to the physical systems under control but also to individuals who rely on these systems. Therefore, it is imperative that measures are taken to prevent such attacks. A CPS integrates computational resources and physical processes, enabling effective control via communication with interconnected devices [14]. Control of systems, remote access, devices, and machines are crucial functionalities that are integral to various industrial environments [15]. However, a number of security flaws that are associated with the widespread use of CPS have the potential to seriously harm both the people who depend on them and the physical objects under control [16]. Therefore, it is imperative to put NIDS on such systems in order to proactively mitigate potential damages caused by these attacks. Network monitoring has been widely employed for purposes such as security and forensics [17]. However, the emergence of new technological developments has presented numerous novel obstacles [2]. Several of the most urgent concerns include;

Diversity: The aforementioned phenomenon arises due to the proliferation of novel protocols and the diverse range of data transmitted via contemporary networks. This poses a challenge in acquiring the knowledge of discerning the distinguishing characteristics between regular and anomalous network traffic.

Accuracy: In order to get optimal performance that meets the desired levels of accuracy, it is necessary to incorporate contextual awareness, and higher levels of granularity. This approach allows for a more comprehensive and holistic perspective.

Volume: The exponential expansion of data is a direct consequence

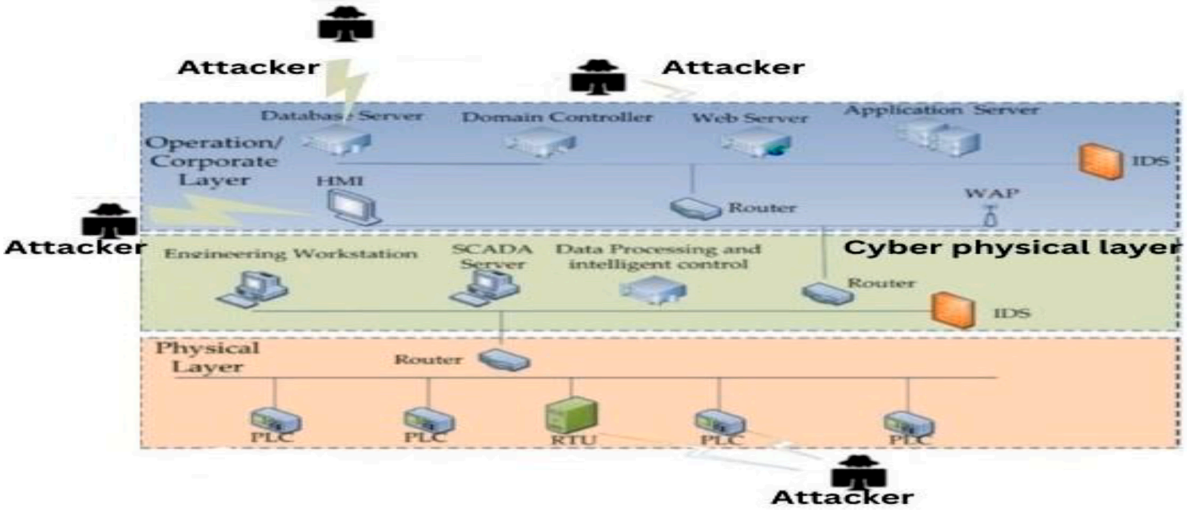

Fig. 1. A structure of SCADA-based CPS.

of the rising prevalence of the cloud-based services, IoT, and other related factors. In order to analyze vast amounts of data in a manner that is both efficient and effective, it is imperative to develop novel methodologies.

Attacks that are low frequency: The occurrence of these attacks results in an uneven distribution within the training dataset for artificial intelligence methodologies, thereby resulting in suboptimal accuracy in the detection process.

In addition, the majority of IDSs include machine learning techniques for the purpose of attack classification. Consequently, it is imperative to extract effective features for various types of intrusions, which can subsequently be employed in supervised learning algorithms to accurately detect and classify attacks in CPS [18]. However, the availability of enough and relevant traffic data that enables effective feature learning is frequently limited. Moreover, there is a significant disparity in the frequency of intrusions and non-intrusions, which poses challenges in the training process. Previous research and theoretical frameworks have evaluated intrusion detection techniques utilizing obsolete and unrepresentative datasets. The datasets utilized in this study do not encompass a comprehensive representation of recent cyberattacks, and the development of the models did not account for the network of CPS [19]. In this study, we have introduced a deep neural network-based IDS that utilizes CNNs. The IDS is trained using a combined dataset consisting of both control system and UNSW-NB15 datasets. The suggested model exhibits innovation in its ability to effectively identify several hostile cyber-attacks within a network of CPS. Furthermore, its lightweight nature and low complexity render it highly suitable for deployment on edge devices. This paper presents several significant contributions which includes;

- We suggested a lightweight ReNet50-D-CNN in this study. The network is trained using a combination of diverse datasets (control and network) representing various types of intrusions in CPS.
- We design ReNet50 model that has undergone training for the purpose of transfer learning. This approach aims to enhance the detection rate while simultaneously decreasing the complexity associated with model training.
- The proposed lightweight transfer learning model experimental findings were compared with the existing shallow machine learning models trained on non-representative CPS data. The findings in our proposed model outperformed the existing shallow ML models.
- The proposed CNNs empower our model to discern intricate patterns indicative of cyber-attacks, contributing to improved detection accuracy.
- The results of our experiment illustrate the effectiveness of our proposed model in accurately categorizing different types of CPS networking attacks and identifying the characteristic patterns of network data.

The subsequent sections of the paper are organized in the following manner. Section 2 contained a discussion of the related research. Section 3 outlines the approach, while Section 4 presents the evaluation results derived from experiments and comparative studies. Section 5 encompasses the concluding remarks and prospects for future research.

2. Related work

In recent years, there has been a growing scholarly focus on the study of IDS within the domain of CPS. In a study conducted by Yang et al. [20]. in 2018, a strategy utilizing zone partition was developed to effectively identify and mitigate cyber-attacks in CPS. This approach demonstrated the capability to detect both known and undiscovered attacks, even in scenarios when many zones within the system are compromised simultaneously. The authors [21] introduced a deep learning approach utilizing a stacked auto-encoder to identify two-stage sparse cyber-attacks targeting the AC state estimation in smart grid CPS. The authors [22] introduced a deep autoencoder model that is both generic and domain-specific to detect intrusions in cyber-physical systems. The generic model is designed to learn the common features present in all network intrusions, while the specialized models are tailored to learn just the features that are specific to a certain domain. The evaluation of the model was conducted using the CICIDS 2017 dataset. The findings of the study indicate favorable outcomes across the majority of attack categories. Nevertheless, the speed is compromised in certain classes, such as SQL Injection and Cross-Site Scripting (XSS), resulting in the misclassification of the Web Attack domain. One potential explanation for this phenomenon could be featured to the commonality of web-based attacks across all of these instances. In the context of web attacks, it has been discovered that the features associated with these attacks exhibit minimal variation compared to other forms of attacks. Consequently, detecting the underlying pattern may pose challenges. In their study, the researchers [23] introduced a Bayesian-based search and scoring method known as ARTINALI. This strategy is designed to effectively identify the crucial areas for instrumentation in a CPS. The ARTINALI# system was implemented to develop an IDS for two CPS: a smart meter and a smart artificial pancreas. The suggested method showcases an average decrease of 64 % in the number of security monitors, resulting in reductions of 52 % and 69 % in memory and runtime overhead respectively. Despite these reductions, the system maintains an average detection rate of over 98 % for mimicked assaults. The utilization of ARTINALI# allows for the adaptability of IDS to be extended to a diverse array of CPS that possess varying levels of available resources. Furthermore, it expedites the process of detecting attacks, a crucial aspect for systems that prioritize safety. Nevertheless, in security-critical systems where more precise real-time attack detection and prevention is required, the IDS initiates data analysis only after the initial iteration of CPS execution has concluded, which may be deemed unacceptable. The authors [24] suggested an intrusion prediction approach to cyber-physical communication networks. The proposed method uses a genetic algorithm technique to search for the best hyperparameter to be configured and used for the proposed Improved Deep Feed Forward Neural Network (IDFNN) approach. The network traffic data is transformed into time series data, and the preprocessing technique is applied. Before starting the training, the hyperparameter search is performed using a genetic algorithm to get the optimal Deep Feed Forward Neural Network configuration. The training method is tasked with predicting the attack and performing analysis on the prediction result. The model was evaluated on UNSW-NB15 and CICIDS2017 data. The result of the study shows that it achieves high accuracy on the two benchmark datasets, i.e., 99.99 % and 99.91 % for CICID2017 and UNSW-NB15, respectively. However, this work is not evaluated on adversarial attacks and real-time settings. Table 1 presents a comprehensive overview of the relevant literature, focusing on the employed methodology, the utilized dataset, and the inherent constraints associated with each study.

2.1. Motivation for CNN integration in this research

In the context of CPS security, the choice to integrate CNNs into our proposed model stems from the unique characteristics and challenges posed by CPS datasets. Unlike traditional machine learning models, CPS data often exhibits spatial dependencies, temporal patterns, and hierarchical structures that demand specialized processing [28]. CNNs excel at capturing intricate spatial features within data, making them particularly well-suited for tasks that involve image-based information, temporal sequences, or sensor data with inherent spatial correlations. In the realm of intrusion detection for CPS, where attack patterns may manifest in complex and spatially distributed ways, CNNs offer a compelling solution to enhance feature extraction and pattern recognition. Furthermore, the ability of CNNs to automatically learn hierarchical representations aligns with the intricate nature of cyber threats in CPS environments. By hierarchically extracting features from raw data,

Table 1
Review of the existing studies in terms of the methodology, dataset used and limitations.

Authors	Methodology	Dataset	Limitations
[24]	IDFNN	CICIDS2017, UNSW-NB15	The efficacy in identifying emerging attack modalities is suboptimal.
[25]	CUSUM, ABDA	Simulation of price manipulation attack	This study is constrained by a restricted number of attacks, and it exhibits a significant percentage of false positives.
[22]	Generic and Domain Specific Autoencoder, Random Forest	CICIDS2017	The suboptimal execution of online attacks has resulted in the misclassification of data.
[26]	Gravitational Search Algorithm (GSA)	IEEECEC2013, and IoT_bonet	The dataset utilized in the models is unsuitable for CPS and does not accurately represent the protocol network of CPS.
[27]	Gaussian Mixture Model Siamese Convolutional Neural Network with Kalman Filter	Power System and UNSW-NB15	The performance of the system is being hindered by the Preprocessing stage.

CNNs empower our model to discern intricate patterns indicative of cyber-attacks, contributing to improved detection accuracy. The integration of CNNs is motivated by their unparalleled efficacy in handling spatially correlated and hierarchically structured data, inherent to the challenges posed by cyber-physical systems.

3. Methodology

The proposed study is comprised of two distinct phases. During the first stage, the work was evaluated using benchmark datasets. Furthermore, during the second stage of our research, we conducted testing on the functionality of the work within a real-time environment specifically designed for an industrial CPS. The system architecture for detecting harmful events on a CPS network is depicted in Fig. 2. Initially, it is important to establish that all devices are interconnected within the local network in order to facilitate effective communication and the seamless exchange of information among them. The entirety of the data traverses through a router, while the Raspberry Pi module intercepts and captures all the packets for the purpose of analyzing the packet transmission within the network. The features utilized in the trained model are derived from the packets that have been captured. Subsequently, the retrieved features are transmitted to the suggested IDS framework, wherein the trained model undertakes the task of categorizing the data into either general or specialized attack classifications.

3.1. Data collection

The data used in this research was collected from the repository provided in the following links. All of these data were preprocessed to establish network-level patterns for the varied types of traffic generated by devices and to use these similarities to spot attack behavior in the CPS architecture. https://research.unsw.edu.au/projects/unsw-nb15-dataset, and https://www.kaggle.com/icsdataset/hai-security-dataset.

3.2. Data preprocessing

An IDS cannot function effectively without the preparation of data for analysis [29]. Therefore, data pre-processing is crucial [30]. Two

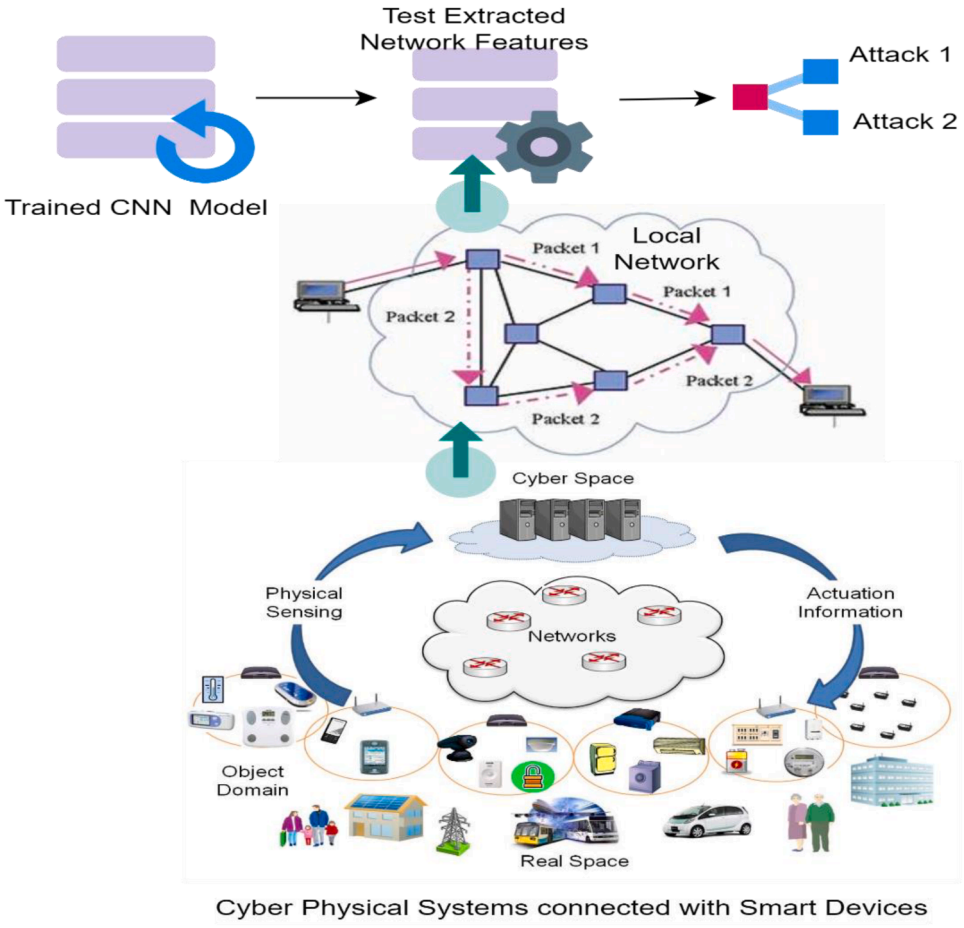

Fig. 2. Proposed System Architecture.

units comprise the pre-processing phase: one-hot encoding and normalization.

3.2.1. One hot encoding. One of the most commonly used methods for converting categorical features into numerical values is the one-hot encoding technique [31]. We used this method to convert each feature of a categorical type into a binary vector, with a value of 1 assigned to the relevant category and 0 assigned to all other categories.

3.2.2. Normalization method. Using maximum-minimum normalization methods after the transformation of categorical variables prevented a possible overlap in the training process resulting from the manipulation of large datasets [32]. In the normalization procedure, we used a scaling range of 0 to 1 to scale the dataset to the same range. Eq. (1) illustrates the fundamental formula utilized in Min-Max normalization.

$$J_{new} = \frac{j - \min(j)}{\max(j) - \min(j)} \tag{1}$$

In the context of this study, the variable j_i denotes a specific feature, while j min and j max represent the minimum and maximum values of said feature, respectively.

3.2.3. Bias handling via synthetic minority oversampling technique (SMOTE). To address imbalances in the dataset, particularly in scenarios where instances of cyber-attacks might be underrepresented, we employed SMOTE. This oversampling technique generates synthetic instances of the minority class, enhancing the model's ability to detect and generalize rare events.

3.2.4. Outlier detection. The interquartile range (IQR) is a statistical measure commonly employed for the purpose of identifying outliers. The IQR is a commonly used statistical measure that quantifies the spread of data by calculating the difference between the first quartile (Q1) and the third quartile (Q3). It is widely recognized as a reliable indicator of central tendency and offers valuable information regarding the distribution of the HAI and UNSW-NB15 dataset. The IQR serves as a valuable tool in spotting probable outliers by establishing the range that spans the central 50 % of the HAI and UNSW-NB15 dataset. The procedures (i-iv) were implemented to identify and handle anomalies.

i. The process of identifying the 1st quartile
ii. The process of identifying the 3rd quartile
iii. The IQR can be determined by subtracting the Q1 from the Q3.
iv. The higher and bottom boundaries of the normal range of data were determined using Eqs. (2) and (3) correspondingly.

$$Whiskerthatisinthe Lowerbound = Q1 - -(IQRx1.5) \tag{2}$$

$$Whiskerthatisinthe Upperbound = Q2 - -(IQRx1.5) \tag{3}$$

Outliers were identified and removed using the IQR method. This statistical measure helps in detecting and mitigating the impact of extreme values, ensuring that the model is not unduly influenced by outliers.

3.3. Proposed ResNet50-one-dimensional (ResNet50-1D-CNN) model

This research focuses on the development of a ResNet50-1D-CNN for the purpose of detecting anomalies in CPS. Fig. 2 depicts the overall structure of the proposed model, while Fig. 3 provides a detailed representation of the CNN block along with its layer description. This model is constructed with an input layer, three blocks of convolution layers, a flatten layer, two fully connected dense layers, and an output layer. Every block consists of two convolutional layers, a normalization layer, a pooling layer, and a dropout layer. Furthermore, it is worth noting that

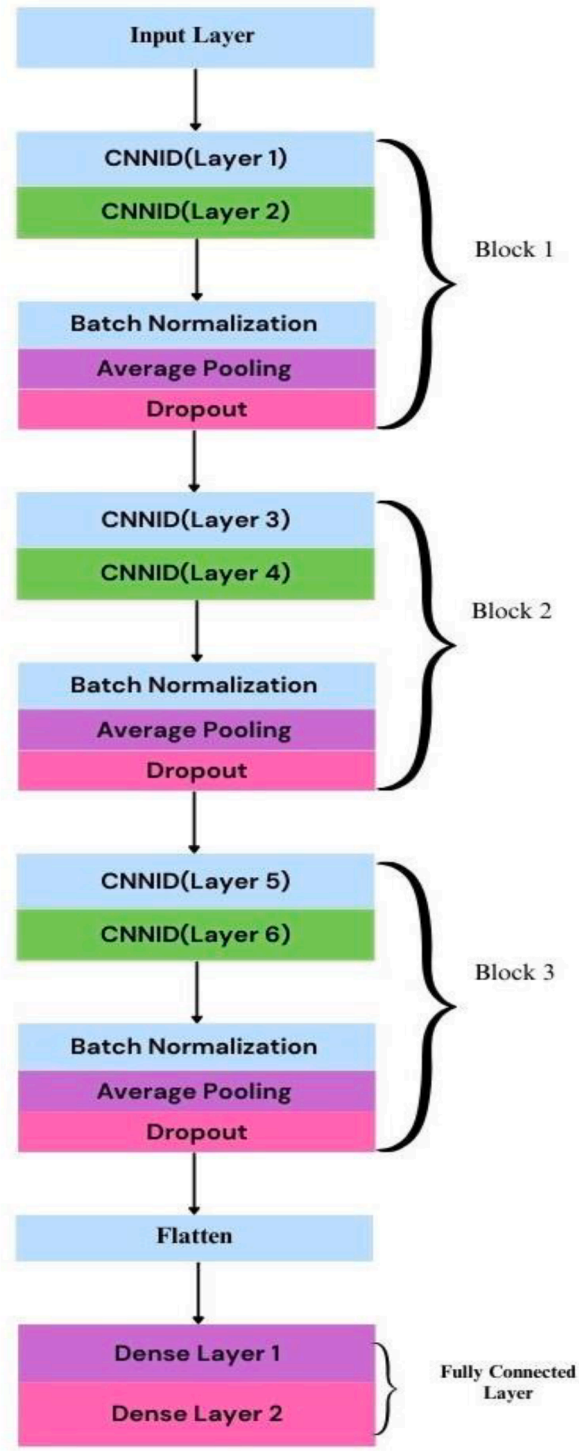

Fig. 3. Proposed 1D-CNN model.

each block includes a pooling layer. The reshaping methodology has the potential to offer inputs to the input layer. The data conversion process within the reshaping system is responsible for transforming the network data into a format that is compatible with the CNN for further analysis and interpretation.

Convolutional layer blocks serve as the fundamental components utilized in the creation of a CNN. The primary function of the convolutional layer is to extract features from the input data and acquire knowledge from these features [33]. The normalization layer of the neural network ensures that all the information is standardized to a

consistent level. The role of the normalization layer is to standardize the output of the convolution layer, enabling it to be utilized by the average pooling layer. The pooling layer enhances the quality of information by consolidating features into sub-maps that exhibit prominent characteristics. The average pooling layer provides an approximation of the cumulative number of updates that have taken place throughout the function maThis enables the system to compute the aggregate count of characteristics that are present within each patch. Overfitting poses a potential concern in the context of CNNs, necessitating careful adjustment of the parameters pertaining to the test dataset. The inclusion of a dropout layer in the neural network architecture mitigates the risk of overfitting by selectively deactivating a subset of neurons during the training process.

3.4. Setting of the hyper-parameter

3.4.1. Loss function. The loss function was computed using the sparse categorical cross-entropy method [34]. During the training process, the objective is to minimize the discrepancy among the expected and actual outputs. This discrepancy is quantified by a loss function that evaluates the disparity between the detected and actual output values. When the number of classes exceeds two, the categorical cross entropy loss function, as specified in Eq. (4) [35], is employed.

$$K = \frac{1}{q}\sum_{y=1}^{q}\sum_{z=1}^{d} v_{y,z}\log m_{y,z} \tag{4}$$

In this context, the variable q denotes the total number of samples that have been processed. The variable v is used to represent the label, which can take on values ranging from 0 to D-1, where d represents the total number of classes. Lastly, the variable m denotes the probability associated with a certain category. The likelihood of the given input being classified into each predetermined category is determined, and the class with the highest probability is chosen as the ultimate outcome.

3.4.2. Optimizer. Deep learning algorithms frequently include many intrinsic optimizers, including but not limited to Gradient Descent (GD), Stochastic Gradient Descent (SGD), AdaDelta, and Adam [36]. The Adaptive Gradient (Adagrad) optimizer [37] has been applied in our model to minimize the loss function by the adjustment of network weights. Eq. (5) elucidates the mathematical formulation that underlies the Adagrad optimizer, a widely employed algorithm for determining the optimal weights of neurons.

$$m_j = m_{j-1} - \eta j \times \frac{\delta R}{\delta mj} \tag{5}$$

The variable m_j denotes the mass of a neuron in the j^{th} repetition, whereas the variable η reflects the learning rate. The symbol δR represents the partial inverse of the loss function with respect to the variable δm_j.

3.5. Transfer learning

Transfer learning (TL) refers to the process of applying knowledge acquired from one domain to another domain for the goals of classification and feature extraction [38]. From the perspective of deep learning, TL is implemented by employing a previously established deep convolutional neural network model that has been trained on a substantial dataset. The pre-existing CNN model undergoes additional training, specifically fine-tuning, using a novel dataset that contains a lower quantity of training data, which is comparable in size to the datasets used in prior training iterations. In recent times, TL has gained significant traction in numerous deep learning applications because to its expedited and simplified process of fine-tuning pre-trained CNN models, as opposed to training CNN models with randomly initialized weights from the beginning [39]. In CNN models, it is commonly observed that the earliest layers are responsible for learning features such as edges, curves, corners, and color blobs. On the other hand, the last layers of these models are known to capture more abstract and specialized properties [40]. TL is an approach in ML where a pre-trained model, originally developed for a specific task, is used as a starting point for a different task. As depicted in Fig. 2, a pre-trained CNN model has been implemented through the utilization of transfer learning for the purpose of binary classification. The Power dataset and UNSW-NB15 datasets were subsequently subjected to classification using an identical pre-trained model. TL concepts were utilized in the setting of binary classification for the aforementioned datasets. In the context of TL, the output layer of the pre-trained CNN model is often eliminated. The model incorporates a novel output layer that includes neurons corresponding to the attack categories included in the dataset. During the training phase of the model, only the dense and output layers remain active. The activation of all additional layers of the pre-trained model is disabled. During the development of the classification model, the convolution, normalization, dropout, pooling, and flatten layers were all maintained in a fixed state, as depicted in Fig. 2. During the training phase, the acquisition of new knowledge was limited to the dense and output layers exclusively. The proposed 1D-CNN can be mathematically represented as follows.

1. Given an input sequence x with N elements, the convolution operation can be defined as;

$$Input\, x = [x_1, x_2, x_3, ..., x_N] \tag{6}$$

$$W = [w_1, w_2, w_3, ..., w_M] \tag{7}$$

Where M is the size of the filter. Let K be the number of filters

2. The convolution operation at a certain point i is computed by taking the product of dots between the filter w and the appropriate input subsequence.

$$Z_i = \sum_{j=1}^{M} x_i + j - 1.wj \tag{8}$$

3. Add a non-linear activation function σ(.) to each element of the convolution result.

$$a_i = \sigma(Z_i) \tag{9}$$

4. Perform a pooling operation (max pooling), to decrease the spatial dimensions.
5. Connect the output to one or more layers that are completely linked for additional processing.

This procedure is implemented across the full input sequence, producing a sequence of activation values. The model parameters encompass the filter weights w, biases, and any parameters within the fully linked layers.

$$a_i = \sigma\sum_{j=1}^{M} x_i + j - 1.wj \tag{10}$$

3.5.1. Transfer learning ResNet-50 structures. The limited number of labelled samples in CPS data, compared to the eleven million annotated images in ImageNet, makes it difficult to train deep CNN models for cyber-attacks detection in CPS. This limitation hinders the accuracy of CNN models in predicting brain tumor detection. Nevertheless, with the utilization of transfer learning technique [41], the deep CNN models that have been trained on ImageNet can exhibit satisfactory

performance when applied to limited data in various domains [42], such as the field of CPS attack detection. This research involves the application of a pre-trained ResNet-50 model, originally built on the ImageNet dataset, to the field of CPS attacks detection. ResNet-50 can achieve higher accuracy by increasing its depth significantly. Given the excellent image classification performance and ability to extract high-quality image features of ResNet-50, we hypothesize that the feature extraction layers of ResNet-50 will likewise exhibit strong performance in the field of CPS attack detection. Table 2 provide a depiction of the process involved in preparing the ResNet-50 architecture layer.

a. ResNet50-1D-CNN Model Architecture: Our ResNet50-1D-CNN model is designed based on the ResNet50 architecture, adapted for one-dimensional data. The architecture comprises residual blocks that facilitate the training of deep networks.

b. Convolutional Layers and Filter Weights: The model incorporates multiple convolutional layers responsible for extracting hierarchical features from the input data. Each convolutional layer is characterized by a set of filter weights, representing learnable parameters. These weights are tuned during the training process to capture distinctive features at various levels of abstraction.

c. Biases: Biases are introduced to each convolutional layer to provide flexibility in fitting the model to the data [43]. The biases contribute to the translation of the activation function and aid in the adaptability of the model to the inherent complexity of the input.

d. Fully Connected Layers: The FC layers, play a crucial role in combining extracted features for the final classification. Each FC layer consists of weights and biases. The weights determine the strength of connections between neurons, while biases contribute to the overall adaptability of the layer.

e. Batch Normalization and Activation Functions: Batch Normalization was applied after convolutional layers; batch normalization enhances the stability and convergence of the model during training. The model employs appropriate activation functions, Rectified Linear Unit, to introduce non-linearity and enable the model to learn complex patterns.

3.5.2. Residual building block. The residual building block (RBB) is the crucial component in the ResNet-50 architecture. The concept of RBB involves the utilization of shortcut connections to bypass convolutional layers, resulting in the skipping of blocks [44]. These shortcuts are beneficial for optimizing the trainable parameters in error back-propagation to prevent the issue of vanishing gradients. This, in turn, aids in constructing deeper CNN structures to enhance the final performance for CPS attack detection.

The RBB architecture has multiple convolutional layers (Conv), batch normalizations (BN), a Relu activation function, and a single shortcut. In this research, there are two distinct RBB structures referred to as RBB-1 and RBB-2, as illustrated in Fig. 4. Both RBB-1 and RBB-2 consist of three Convolutional and batch normalization layers [44]. However, the shortcut in RBB-1 corresponds to the identity x, as depicted in Fig. 4a. Let F represent the non-linear function for the convolutional path in RBB-1. The output of RBB-1 can be expressed using Eq. (11). The structure of RBB-2 is depicted in Fig. 4b. The shortcut consists of alternative Convolutional (Conv) and Batch Normalization (BN) layers. Let H represent the abbreviated route, and the result of RBB-2 can be expressed as Eq. (12).

$$y = F(x) + x \tag{11}$$

$$y = F(x) + H(x) \tag{12}$$

Following the initial convolutional layer in ResNet-50, many RBB-1 and RBB-2 blocks are arranged in a stacked formation. The ResNet-50 model, which is documented in reference [45], is utilized in this research.

3.6. Pooling Operation Type

In our ResNet50-1D-CNN model, we employ MaxPooling1D as the pooling operation. MaxPooling1D is a one-dimensional pooling operation that extracts the maximum value from a set of consecutive values within a given window. This operation helps to retain the most salient features while reducing the spatial dimensions of the input.

3.7. Impact on Spatial Dimensions

The MaxPooling1D operation has the following impacts on spatial dimensions:

3.7.1. Spatial Reduction. MaxPooling1D systematically reduces the size of the input data along the temporal axis. By selecting the maximum value within each window, it retains essential information about dominant features while discarding less critical details.

3.7.2. Translation Invariance. MaxPooling1D introduces a degree of translation invariance, allowing the model to recognize essential patterns regardless of their precise temporal location. This is particularly beneficial in capturing hierarchical features relevant to intrusion detection.

3.7.3. Computational Efficiency. By down-sampling the spatial dimensions, MaxPooling1D enhances computational efficiency during subsequent layers. This reduction in dimensionality also aids in mitigating overfitting.

4.1. Description of the datasets

The performance of our model has been evaluated on the control system HAI and UNSW-NB15 datasets.

4.2. Datasets

In order to assess the efficacy of the proposed transfer learning approach for intrusion detection in CPS, two existing datasets specifically designed for CPS are employed. The datasets are partitioned into two separate subsets, known as the training set and the testing set, with a distribution ratio of 75 % and 25 % respectively. One of the primary

Table 2
Parameters of the propose ResNet50–1D-CNN.

Layers	Type	Output Shape	Parameters	Connection
Input_1	InputLayer	None, 256,256,3	0	[]
Conv1_pad	ZeroPadding2D	None, 262, 262, 3	0	input_1[0][0]
Conv1_conv	Conv2D	None, 128,128,64	9472	conv1_pad[0][0]
Conv1_bn	BatchNormalization	None, 128,128,64	256	conv1_conv[0][0]
Conv1_relu	Activation	None, 128,128,64	0	conv1_bn[0][0]
Pool1_pad	Zeropadding2D	None, 130,130,64	0	conv1_relu[0][0]
Pool1_pool	MaxPooling2D	None, 64,64,64	0	pool1_pad[0][0]
Conv2_block1_1_conv2D	Conv2D	None, 64,64,64	4160	pool1_pool[0][0]
Conv2_block1_1_bn	batchNormalization	None, 64,64,64	256	Conv2_block1_1_conv[0][0]
Conv2_block1_1_relu	Activation	None, 64,64,64	0	Conv2_block1_1_bn[0][0]

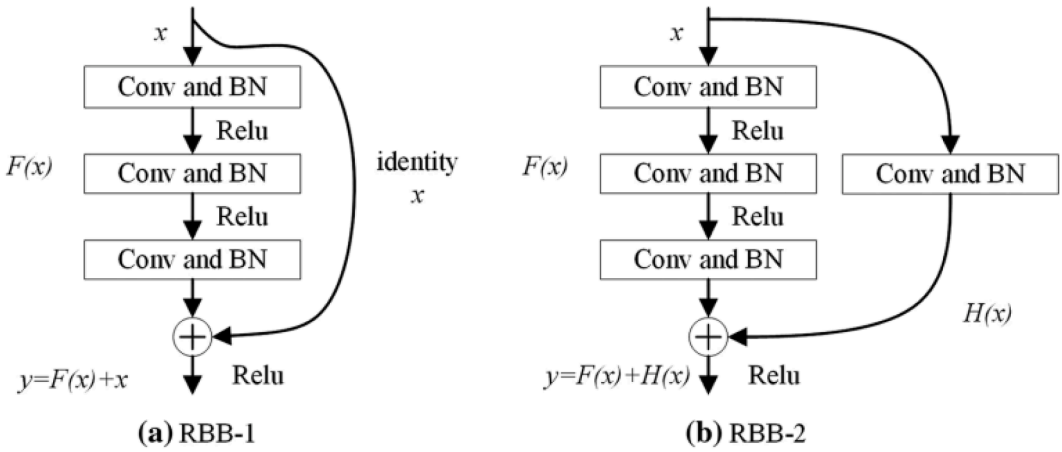

Fig. 4. Residual building block.

challenges encountered in the domain of anomaly detection research revolves around obtaining or generating a suitable dataset for experimental endeavors. In this study, we conducted an analysis of pre-existing datasets in order to identify the dataset that is most appropriate for further exploration. The authors delineated the dataset prerequisites by the research objective of identifying anomalies in CPS:

R1: the acquisition of the dataset ought to be conducted from the CPS.

R2: the dataset ought to comprise records of events.

R3: It is recommended that the dataset includes anomalies.

R4: The dataset must be appropriately labeled to distinguish between normal and abnormal data.

R5: It is recommended that the dataset utilized in the study closely approximates real-world data, specifically data derived from authentic or partially authentic systems.

The datasets that meet the specified criteria, namely those that comprise labeled sensors and network data, include the recently developed HAI dataset [46] and the UNSW-NB15[47] dataset. These datasets were subjected to a comprehensive analysis by the authors. The particulars of each dataset are delineated as follows;

4.2.1. HAI dataset

The provided data presents a thorough depiction of the parameters linked to a testbed specifically developed for an industrial control system, which include an integrated simulator. The experimental setup has four distinct components, namely a boiler, turbine, water treatment unit, and a hardware-in-the-loop simulator [48]. The HIL simulation integrates a model of pumped-storage hydropower plants with thermal electricity for the purpose of simulation. The analysis of the findings was conducted based on the criteria that were provided. The aforementioned conditions, namely R1, R2, R3, and R4, have been successfully fulfilled. The accomplishment of requirement R5 has been attained. However, it is imperative to recognize that the total integrity of the dataset is contingent upon the caliber of the simulated element within the experimental framework. The preliminary empirical findings align with the results documented in other research publications. Hence, it can be inferred that the aforementioned dataset demonstrates coherence and is considered suitable for use in the context of anomaly detection for CPS. The distribution of the attacks is given in Table 3.

Fig. 5 depicts the features that exhibit the strongest correlation in the HAI data. Features that have a correlation greater than 0.95 with the target variable are combined.

4.2.2. UNSW-NB15 dataset

The dataset known as UNSW-NB15 consists of a collection of both regular network traffic logs and records of malicious attacks. The dataset has a data volume of around 100 gigabytes and consists of a total of 2540,044 events [49]. The inclusion of 48 features, including the class label, accounts for the high-dimensional big data characteristic of each observation. The dataset demonstrates a velocity of approximately 5–10 megabytes per second throughout its transmission between sources and destinations, thus simulating a realistic network environment [46]. It comprises 10 distinct classes, with one being normal and the remaining nine about various security events. The aforementioned requirements, namely R1, R2, R3, R4, and R5, have been successfully satisfied and achieved. In order to input the dataset into the proposed model, we have identified the top 10 features from both the power and UNSW-NB15 data features. The distribution of the UNSW-NB15 data is shown in Table 4.

Fig. 6 illustrates the features with the most significant correlation in the UNSW-NB15 data. Features with a correlation exceeding 0.95 with the target variable are merged.

4.3. Feature selection process via random forest's feature relevance technique: ResNet50-1D-CNN for CPS intrusion detection

To identify the most relevant features for intrusion detection, we employed the random forest algorithm. Random Forest is a robust ensemble learning technique capable of assessing feature importance. This technique was chosen for its ability to handle both classification tasks and gauge the significance of individual features. The gini

Table 3
Summary of vulnerability attacks.

Category	Service	Target of control systems			Severity			
		Boiler	Turbine	Water	Critical	High	Moderate	Total
Web	HTTP	√		√	148	0	0	148
Remote	DCERPC	√	√		29	39	2	70
OS	SMB	√	√	√	24	54	3	81
CPS	Modbus			√	0	2	0	2
Time	NTP	√	√		1	1	0	2
Historian	Database	√			16	12	0	28

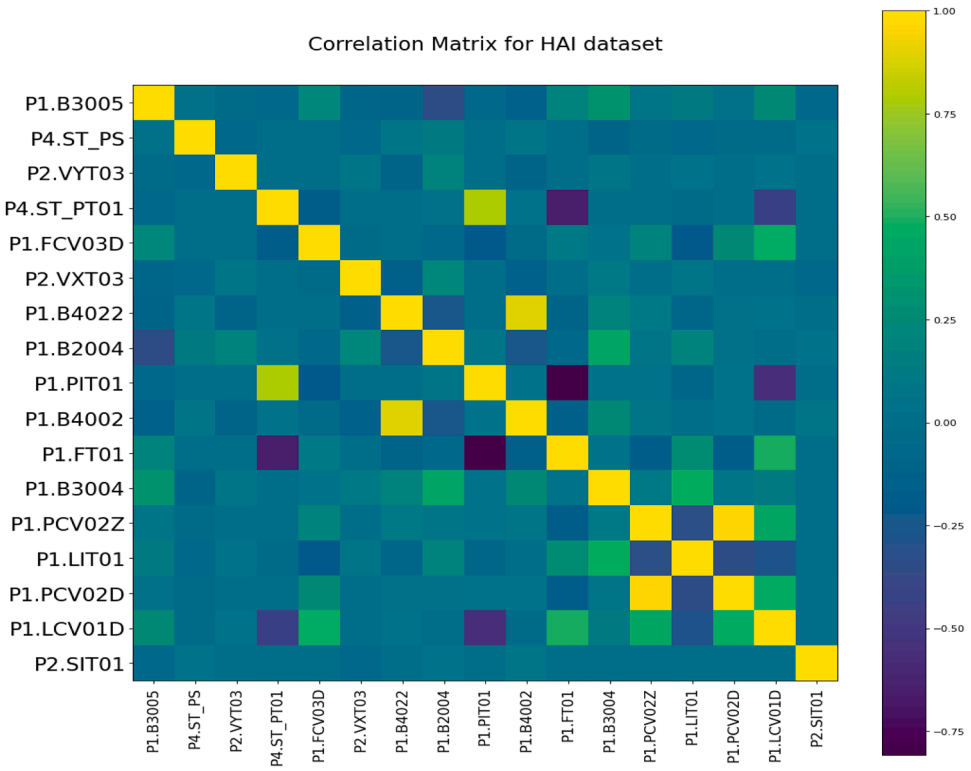

Fig. 5. Correlated features in the HAI dataset.

Table 4
The spread of classes within the UNSW-NB15 data.

Category	Train set	Test set
Benign	56,000	37,000
Backdoor	1746	583
Exploits	33,393	11,132
Generic	40,000	18,871
Shellcode	1133	378
Analysis	2000	677
DoS	12,264	4089
Fuzzers	18,184	6062
Worms	130	44
Reconnaissance	10,491	3496
Total	175,341	82,332

importance, a widely used metric in random forest models, was utilized to calculate the significance of each feature [50]. The gini importance score quantifies the contribution of each feature to the homogeneity of the decision trees within the ensemble. Features with higher gini importance are considered more relevant to the classification task. Our objective was to identify the top ten features with the highest gini importance scores in each dataset. This selection process ensures a focus on the most influential features while maintaining a manageable feature subset for model training. The decision to select features was based on their individual gini importance scores [51]. Features demonstrating the highest gini importance were prioritized for inclusion in the top ten, ensuring a comprehensive yet concise set of features that significantly contribute to the intrusion detection process. This selection was made using the random forest's feature relevance technique. The gini importance is utilized to compute the significance of a feature. The features that were chosen of the datasets and additional relevant information pertaining to the dataset are shown in Table 5.

4.4. Feature importance

The feature importance metric provides a numerical score that indicates the degree of usefulness of each feature in the model's construction [52]. In the realm of machine learning, feature importance and its presentation are significant and popular analysis techniques [53]. Due to the ease and readability of feature ranking, it is particularly employed in fields like biology and the social sciences [54]. The coefficient value of each feature is used to determine feature significance and ranking.

4.4.1. The feature importance of coefficient score of the proposed model

The ResNet-1D-CNN classification model assesses the coefficient score and feature importance for each feature. We estimated the feature importance with their coefficient scores for the top ten features for each of the models in this section. Figs. 7 and 8 displays the graphical representation of the top ten features with their respective coefficient score.

4.5. The metrics employed for evaluating the performance of the propose resnet50–1D-CNN

The utilization of ML performance evaluation metrics enables the quantification of the performance of an ML model. The utilization of these metrics enables the assessment of the efficacy of the ML model. The ratios used to evaluate the efficacy of our model were derived from the confusion matrix.

i. Accuracy: The metric that evaluates the effectiveness of a classifier in predicting classes is accuracy. The aforementioned ratio is derived through the division of the aggregate count of precise predictions, encompassing true positives and true negatives, by the overall count of predictions made, which includes true positives, true negatives, false positives, and false negatives. We included accuracy to gauge the model's overall performance in correctly classifying both normal and anomalous instances.

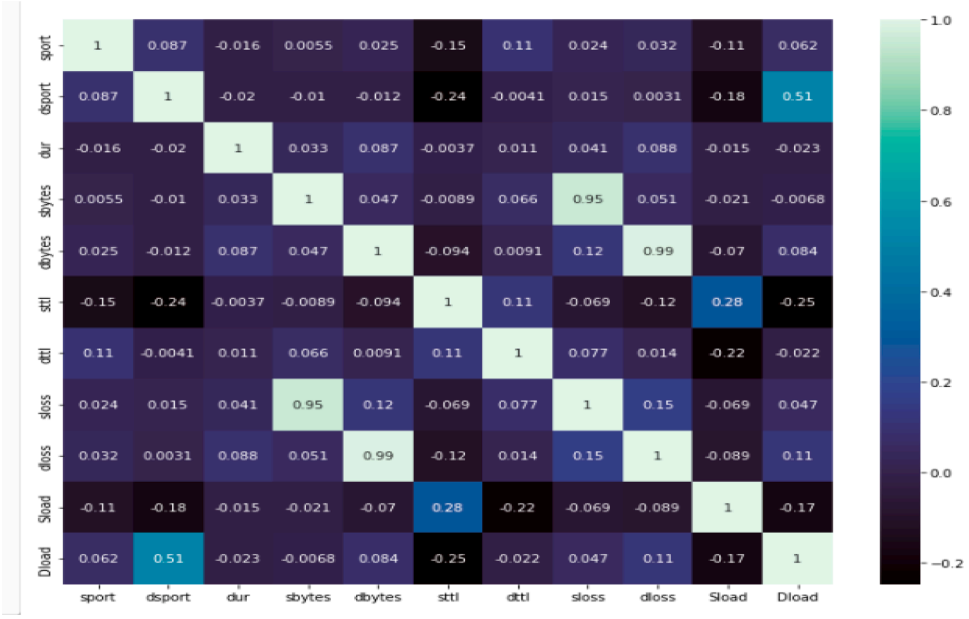

Fig. 6. Features correlation in the UNSW-NB15 data.

Table 5
Features selected in the HAI and UNSW-NB15 Testbeds.

Datasets	No of classes	No of features selected
HAI	38	AP-P1LC-SPRP, AP-P1LC—CO, AP-P1LC—CORP, AP-P1FC—CORP, AP-P1LC-SP, AP-P1PC-SP, AP-P1PC-SPRP, AP-P1PC—CO, AP-P1PC—CORP, AP-P2SC-SPRP (10)
UNSW-NB15	48	Dur, Src_ip, dstip, dloss, spkts, sload, dtcpb, dmeansz, stime, dintpkt (10)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

ii. The concept of precision pertains to the ratio of true positive predictions to the total number of positive predictions made. The calculation involves the division of the count of accurately predicted positive classes (TP) by the overall count of predicted positive classes (TP + FP). It helps evaluate the model's ability to avoid false positives, which is critical for minimizing the impact of false alarms in real-world deployment scenarios.

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

iii. The recall is calculated by dividing the number of TP by the number of TP and FN. The recall is regarded a measure of a classifier's

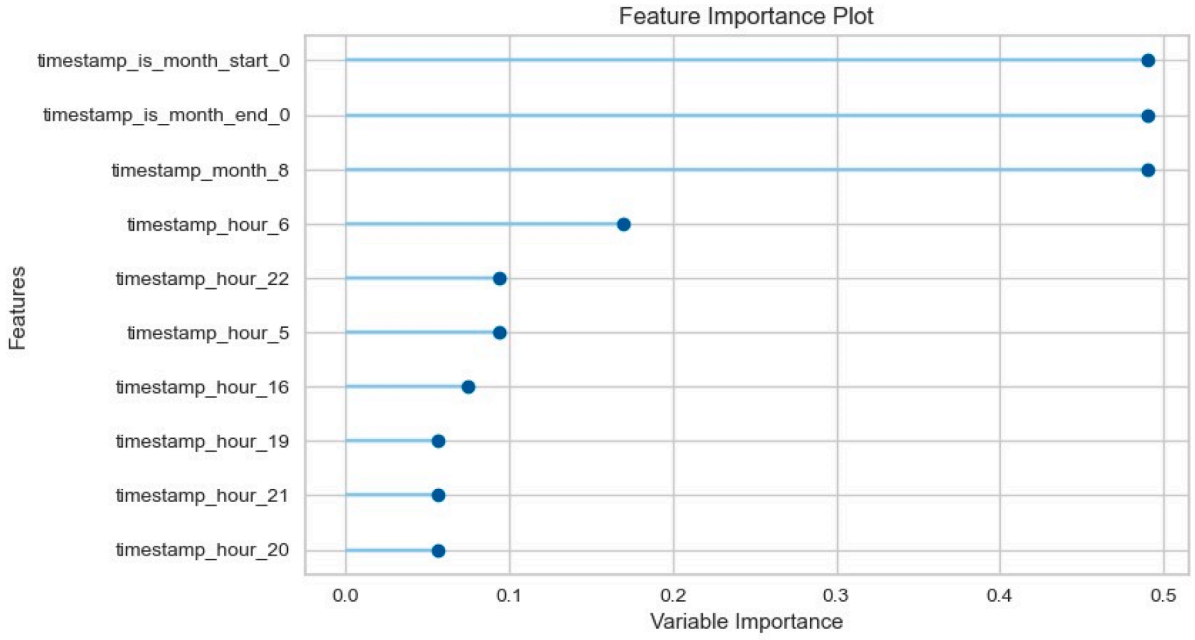

Fig. 7. ResNet50–1D-CNN top ten features with the coefficient score on HAI.

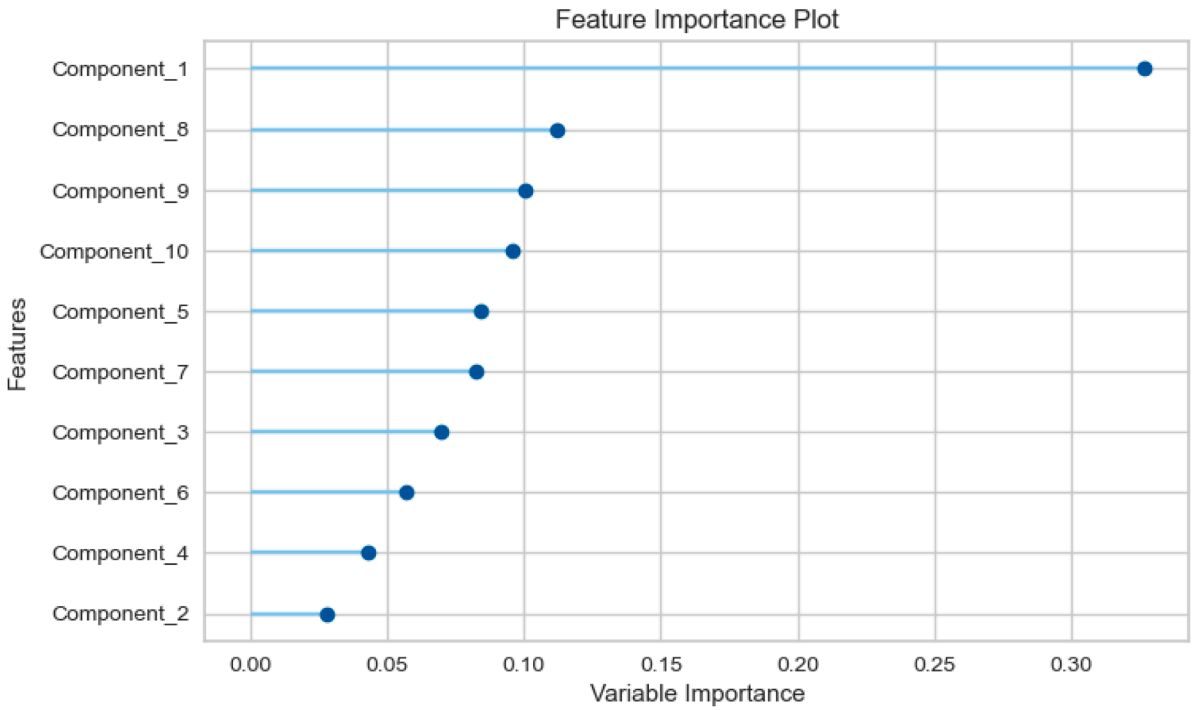

Fig. 8. ResNet50–1D-CNN top ten features with the coefficient score on UNSW-NB15.

completeness, with a low recall value resulting in a large number of FN. Using Eq. (15), recall is estimate. It is particularly crucial in intrusion detection as it indicates the model's ability to detect cyber-attacks without missing any. We chose recall to ensure high sensitivity in detecting malicious activities in CPS environments.

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

4.6. Justification for evaluation metrics

In our study on intrusion detection in CPS using the proposed ResNet50-1D-CNN model, we meticulously selected evaluation metrics that accurately reflect the performance of our model in real-world scenarios. Our choice of evaluation metrics is based on their ability to provide a comprehensive assessment of the model's effectiveness in detecting cyber-attacks while minimizing false positives and false negatives. The Section 4.5 shows the breakdown of the evaluation metrics we chose and the rationale behind each.

5.1. Experimental hardware and software requirements

The simulations are executed on a laptop with an Intel Core (TM) i5–8250 U processor clocked at 1.60 GHz and 8GB of RAM. The algorithms are used to classify and identify threats and anomalies across all the HAI and UNSW-NB15 datasets. Scikit learn, PyTorch, TensorFlow, Keras libraries in Python was utilized in the implementation of the models. We utilized TensorFlow and Keras frameworks for model development and training. The specific versions used were TensorFlow 2.5 and Keras 2.4.3.

5.1.1. Simulation Platforms

Our research leverages a combination of versatile and widely adopted libraries to simulate and implement the intrusion detection models. The primary platforms utilized in our simulation work include:

a. Scikit-learn: Scikit-learn is employed for various machine learning tasks, including preprocessing, feature selection, and the implementation of classical machine learning algorithms [55]. Its user-friendly interface, extensive documentation, and compatibility with Python make it an excellent choice for prototyping and experimenting with traditional machine learning techniques.

b. PyTorch: PyTorch is employed for the implementation and training of deep learning models, specifically for the ResNet50-1D-CNN architecture proposed in our research. PyTorch is renowned for its dynamic computational graph, facilitating more flexible model construction and dynamic execution [56]. It has gained popularity in the deep learning community for its intuitive design and seamless integration into the Python ecosystem.

c. Number of Runs: To ensure the robustness of our findings, we conducted each experiment multiple times. For the control system HAI dataset, we performed five runs, and for the UNSW-NB15 dataset, we conducted ten runs. This approach allowed us to account for potential variations and obtain reliable performance metrics.

d. Statistical Tests: We employed rigorous statistical tests to evaluate the significance of our results. Specifically, we utilized paired t-tests to compare the performance metrics (accuracy, recall, F1-score, and precision) between our proposed ResNet50-1D-CNN model and baseline methods. The significance level was set at $\alpha = 0.05$.

5.2. Model hyperparameters and optimization techniques of ResNet50-1D-CNN for intrusion detection in CPS

The proposed model combines the power of ResNet50 with a 1D Convolutional Neural Network to effectively capture spatial and temporal features in the context of Cyber-Physical Systems. To construct optimal models, the major hyper-parameters of all the ResNet50-1D-CNN were optimized using the Adaptive gradient. The hyper-parameters of ResNet50 and 1D-CNN are given in Tables 6 and 7.

5.3. Optimization Techniques

The optimization techniques employed in this research are Adaptive gradient and ResNet50.

Table 6
Hyper-parameters of ResNet50.

Hyper-parameters	Values
Learning rate	0.001
Batch size	32
Optimizer	Adagrad
Loss function	Categorical Cross entropy
Epochs	50
Activation function	ReLU

Table 7
Hyper-parameters of 1D-CNN.

Hyper-parameters	Values
Kernel Size	3
Number of Filters	64
Activation Function	ReLU
Pooling	MaxPooling1D
Dropout Rate	0.5

5.3.1. Adaptive Gradient (Adagrad)

Adagrad is employed as the optimizer, adapting the learning rates individually for each parameter. This helps in effective convergence, particularly in scenarios with sparse data.

5.3.2. Transfer Learning (ResNet50)

Transfer learning from ResNet50 is utilized to leverage pre-trained weights on large datasets. This accelerates training and enhances the model's ability to generalize features relevant to intrusion detection in CPS.

5.4. Justification for selecting the utilized hyper-parameters

5.4.1. Learning Rate

A small learning rate (0.001) is chosen to ensure stable convergence, especially when fine-tuning the model.

5.4.2. Batch Size

A moderate batch size of 32 is selected to balance between computational efficiency and model convergence.

5.4.3. Adaptive Gradient

Adagrad is chosen due to its adaptive learning rate capability, making it suitable for scenarios where features have varying importance.

5.4.4. Transfer Learning

The use of ResNet50 as a pre-trained model helps in capturing complex hierarchical features without the need for extensive training data.

5.5. Results and discussion

The performance of our ResNet50-1D-CNN-based model was assessed in the context of binary classification. The model we employed for training was trained on a GPU and utilized 8GB of RAM on the Google Collaboratory platform. Table 8 and Fig. 9 presents the results of the binary classifications conducted on the HAI control system and UNSW-NB15.

Table 8
Performance evaluation of the proposed ResNet50-1D-CNN model.

Models	Metrics	Control system HAI data	UNSW-NB15
ResNet50–1D-CNN	Accuracy	97.32	99.89
	Recall	98.41	99.09
	F1-score	96.23	98.01
	Precision	97.09	98.70

5.6. Comparative analysis with the existing system

This section presents a comparative analysis of the outcomes obtained from our suggested ResNet50-1D-CNN model in relation to other existing research initiatives. The models proposed in this research exhibited a significant improvement in detecting anomalies in CPS. This study examined the potential application of a ResNet-50-1DCNN in addressing the challenge of anomaly detection within networks of CPS. This was achieved through the implementation of binary classification using transfer learning techniques. The models we proposed exhibited a notable improvement in detecting irregularities within CPS contexts. This study examined the potential application of a CNN in addressing the challenge of anomaly detection inside CPS networks. In this study, we investigated the efficacy of Transfer CNNs in the detection and classification of abnormalities. The evaluation focused on assessing the efficacy of utilizing binary classification through transfer learning in order to find and categorize abnormalities in CPS networks using CNN. The utilization of transfer learning in binary classification is facilitated by the pre-trained CNN paradigm. Based on current understanding, no previous research has utilized transfer learning methodology for the purpose of abnormality (cyber-attacks) identification in CPS networks. This approach involves the reuse of a pre-trained ResNet50 model for binary anomaly detection. TL is commonly employed in order to decrease the complexity of categorization and runtime. Furthermore, our proposed paradigm offers the additional benefit of transfer learning. The utilization of TL methodology leads to a notable reduction in the required durations for training, validation, and testing in the context of classification tasks. The outcomes of our planned binary class categorization effort are presented in Table 8. One class is associated with the anomalous behavior of the network, whereas the other class is associated with the regular behavior of the network. Table 9 and Fig. 10 shows a comparative analysis of our suggested methodology with other relevant studies. The findings indicate that the use of TL achieves a high level of accuracy, with a recorded value of 99.89 %.

Additionally, the recall rate reaches 99.90 %, while the F1-score attains a value of 98.01. Moreover, the precision rate is observed to be 98.70 % for the UNSW-NB15 dataset. The comparison results indicate that the performance of our proposed ResNet50-1D-CNN model surpasses that of the current models in the literature.

5.7. Comparison baseline with other methods

To establish a baseline for comparison with other methods, we carefully selected representative baseline models in Table 10 commonly used in intrusion detection research. These baseline models include traditional machine learning algorithms such as KNN, Random Forest, K-means, NB, and Support Vector Machine (SVM). By comparing our proposed ResNet50-1D-CNN model against these baseline methods, we can assess its superiority in terms of detection accuracy, robustness, and efficiency. Moreover, benchmarking against well-established baseline models provides valuable insights into the practical effectiveness of our approach and highlights its potential for real-world deployment in CPS environments.

5.8. Aligning HAI and UNSW-NB15 datasets with real-world CPS scenarios

5.8.1. HAI Dataset

The HAI dataset is specifically curated to reflect the challenges and intricacies of real-world CPS environments. It encompasses diverse cyber-physical aspects, capturing anomalies and intrusions in complex systems. The dataset is generated with a focus on high-dimensional data, mirroring the intricacies of modern CPS scenarios. Features within the HAI dataset simulate various components of a CPS, providing a rich and

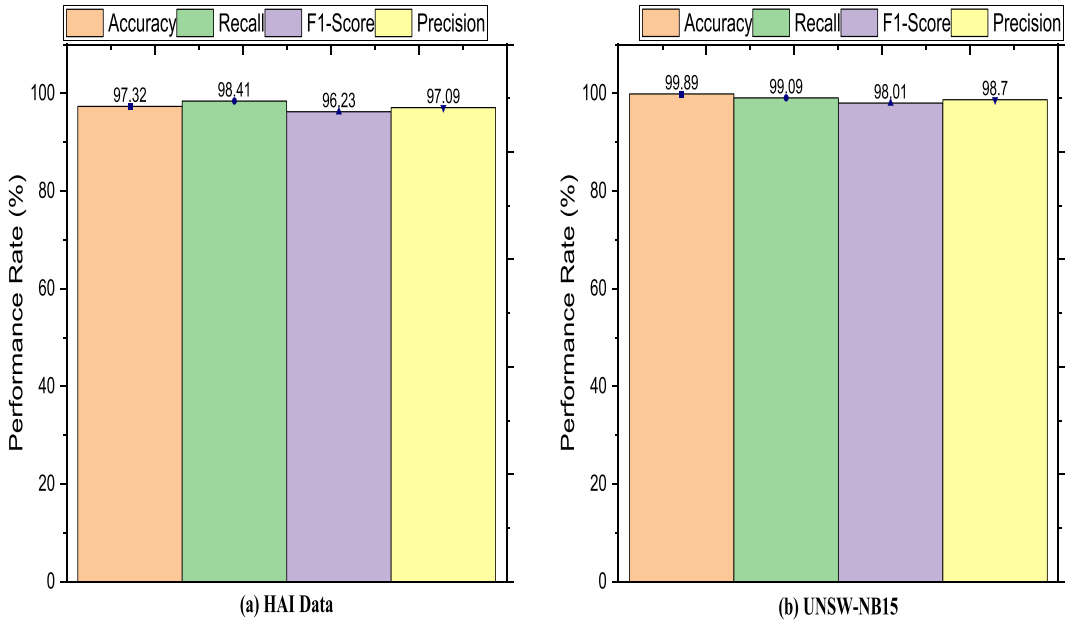

Fig. 9. Performance of the proposed ResNet50–1D-CNN.

Table 9
Comparison of the proposed ResNet50-1D-CNN with the existing systems.

Authors	accuracy	recall	F1-score	precision
[24]	96.89	99.56	99.72	99.55
[25]	98.20	95.00	96	99.80
[22]	99.70	84.93	98.02	79.06
[26]	84.44	–	91.01	–
[27]	98.90	99.64	–	–
Proposed ResNet50-1D-CNN	99.89	99.09	98.01	98.70

Table 10
Comparison with the baseline methods.

Authors	Baseline models	Accuracy	Recall	Precision
[57]	KNN	81.3	–	–
[57]	RF	88.4	–	–
[58]	K-means	17.31	18.49	53.38
[58]	NB	77.23	81.02	100
[59]	SVM	92.53	93.6	78.2
Our Proposed model	ResNet50-1D-CNN	99.89	99.09	98.70

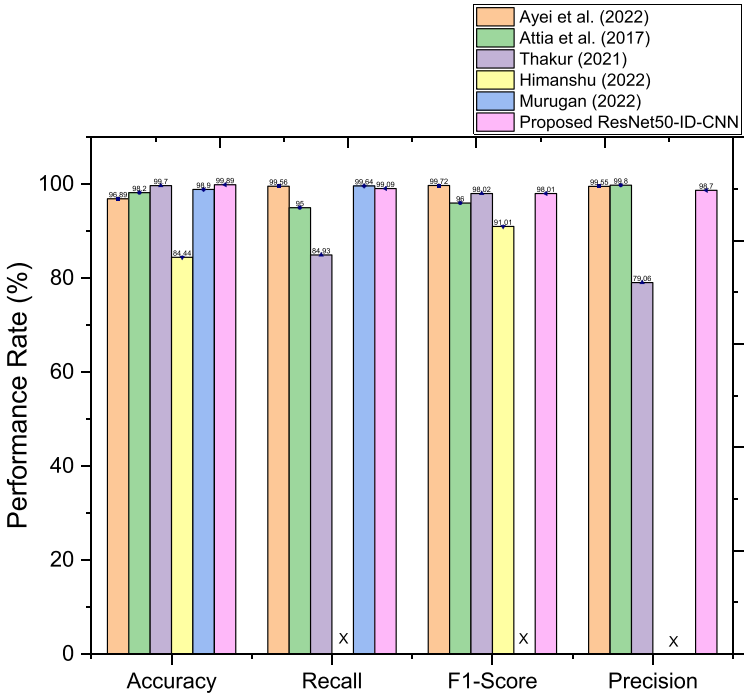

Fig. 10. Performance Comparison of the Proposed ResNet-1D-CNN with the existing systems.

realistic foundation for intrusion detection model training.

5.8.2. UNSW-NB15 Dataset

The UNSW-NB15 dataset, derived from the University of New South Wales, is a well-established benchmark dataset widely used for evaluating intrusion detection systems. It incorporates a variety of cyber threats and attack scenarios, making it relevant to real-world CPS challenges. The dataset's diversity in attack types, network traffic, and system behaviors aligns with the complexities encountered in modern cyber-physical settings.

5.9. Relevance to real-world cyber physical systems

5.9.1. Scenario Diversity

Both datasets encompass a wide range of scenarios, including different types of attacks, network configurations, and system behaviors, reflecting the multifaceted nature of real-world CPS.

5.9.2. Multi-Dimensional Features

The datasets include multi-dimensional features, such as network traffic patterns, system logs, and other relevant parameters, providing a holistic representation of CPS environments.

5.9.3. Imbalance and Anomalies

The datasets incorporate class imbalances and anomalies, replicating the challenges faced in actual CPS settings where detecting rare events is crucial for security.

6.1. Threat to validity

6.1.1. External validity

External validity pertains to the extent to which our experimental findings can be generalized to other initiatives. We do tests using two authentic CPS datasets in order to mitigate the risk associated with this type of validity. Furthermore, the features of our benchmark datasets are well-suited and illustrative of attacks in a CPS environment. We cannot assert that our experimental findings may be extrapolated to datasets that are not representative of the CPS environment. External validity concerns the generalizability of our findings beyond the specific datasets and experimental conditions.

6.1.2. Mitigations

To enhance external validity, we selected diverse datasets, including HAI and UNSW-NB15, covering various cyber-physical system scenarios.

6.2. Internal validity

Internal validity pertains to the accuracy and reliability of causal relationships within the study. To mitigate the potential impact of erroneous implementations on our experimental outcomes, we employ the scikit-learn, TensorFlow, and PyTorch libraries to execute the algorithms. Furthermore, we optimize the parameter values, such as the width of the parameter in the One-dimensional CNN model, and employ the Adaptive gradient to minimize the loss function.

6.2.1. Mitigations

We ensured controlled experimentation by meticulously designing and executing our model training and testing processes. To strengthen internal validity, robust statistical methods were employed during the evaluation of the proposed ResNet50-1D-CNN model.

6.3. Construct validity

Construct validity refers to the extent to which our chosen constructs (features, models) accurately represent the underlying phenomena. We utilize four widely-used metrics to assess the effectiveness of 1D-CNN for IDS in CPS. These metrics do not consider the expense of inspection. In future work, we will assess the efficacy of our strategy using the Mathew correlation coefficient, effect-aware indicators, and AUC.

6.3.1. Mitigations

We conducted a detailed analysis of feature relevance, employing random forest's gini importance to select the most influential features. We provide a comprehensive description of the ResNet50–1D-CNN model, including hyperparameters and optimization techniques, to ensure a clear understanding of our chosen constructs.

6.4. Limitations and future research directions

6.4.1. Dataset Specificity. Our study relies on specific datasets (HAI and UNSW-NB15), which may limit the generalizability of our findings to other cyber-physical system scenarios. Future research should explore diverse datasets from various domains within cyber-physical systems to enhance the robustness and applicability of intrusion detection models.

6.4.2. Model Complexity. While the ResNet50-1D-CNN model significantly reduced complexity, further investigations into model interpretability are warranted. Future work could focus on developing explainable AI techniques to enhance the transparency of complex models, aiding practitioners in understanding model decisions.

6.4.3. Hyperparameter Sensitivity. Although hyperparameters were tuned, the sensitivity of the proposed model to different parameter configurations remains a consideration. Future studies could conduct extensive sensitivity analyses to identify optimal hyperparameter settings for improved model performance across various datasets.

6.5. Areas for future work

6.5.1. Transfer Learning Variants. Investigate different transfer learning variants and architectures to assess their impact on the efficiency and effectiveness of intrusion detection in diverse environments.

6.5.2. Explainable AI Techniques. Develop and integrate explainable AI techniques to enhance the interpretability of intrusion detection models, providing insights into the decision-making process.

6.5.3. Ensemble Approaches. Evaluate the potential benefits of ensemble approaches, combining multiple intrusion detection models to improve overall accuracy and resilience to diverse attack scenarios.

7.1. Conclusion and future work

This paper introduces a transfer learning model for detecting anomalies in CPS networks. Our approach involves utilizing authentic control systems HAI and UNSW-NB15 information to detect anomalous behavior in CPS networks. We employ ResNet50-1D-CNN models to classify different anomalies. The model undergoes training using a substantial dataset, incorporating transfer learning techniques to enhance the dataset's performance and minimize the training complexity of the model. The model's performance is also evaluated in comparison to the current work. The results demonstrate an accuracy of 97.32 % and 99.89 % when using the Control System HAI and UNSWNB-15 datasets, respectively. These accuracy rates are significantly higher than those of the previous model. The suggested model will be compared with existing methods such as GRU and LSTM in future research.

CRediT authorship contribution statement

Yakub Kayode Saheed: Writing – review & editing, Writing –

original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Oluwadamilare Harazeem Abdulganiyu:** Writing – original draft, Validation, Resources, Data curation. **Kaloma Usman Majikumna:** Writing – original draft, Software, Resources, Project administration, Data curation. **Musa Mustapha:** Resources, Methodology, Investigation, Data curation. **Abebaw Degu Workneh:** Writing – original draft, Visualization, Software, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this manuscript is available in the following link https://research.unsw.edu.au/projects/unsw-nb15-dataset. The HAI data used in this study is openly available on kaggle website at https://www.kaggle.com/icsdataset/hai-security-dataset entitled HAI Security Dataset V. 4, created by Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and HyoungChun Kim.

Acknowledgements

We appreciate the reviewers for their great effort, time, and valuable feedback.

References

[1] K. Di Lu, Z.G. Wu, T. Huang, Differential evolution-based three stage dynamic cyber-attack of cyber-physical power systems, IEEE ASME Trans. Mechatronics 28 (2) (2023) 1137–1148, https://doi.org/10.1109/TMECH.2022.3214314.
[2] Z. Wang, W. Xie, B. Wang, J. Tao, E. Wang, A survey on recent advanced research of cps security, Appl. Sci. 11 (9) (2021), https://doi.org/10.3390/app11093751.
[3] R. Queiroz, T. Cruz, P. Simões, Testing the limits of general-purpose hypervisors for real-time control systems, Microprocess. Microsyst. 99 (2023) 104848, https://doi.org/10.1016/j.micpro.2023.104848. May.
[4] (eds Y.K. Aliyu, G. Fonkam, M.M. Nsang, A.S. Abdulkarim, M. Rakshit, S. and Saheed, Airwaves Detection and Elimination Using Fast Fourier Transform to Enhance Detection of Hydrocarbon, in: A. Khosla, P. Chatterjee, I. Ali, D Joshi (Eds.), Optimization Techniques in Engineering, Wiley, 2023 (eds.
[5] R. Mitchell, I.R. Chen, A survey of intrusion detection techniques for cyber-physical systems, ACM Comput. Surv. 46 (4) (2014), https://doi.org/10.1145/2542049.
[6] O.H. Abdulganiyu, T.A. Tchakoucht, Y.K. Saheed, Towards an efficient model for network intrusion detection system (IDS): systematic literature review, Wirel. Netw (2023), https://doi.org/10.1007/s11276-023-03495-2.
[7] J.C. Huang, G.Q. Zeng, G.G. Geng, J. Weng, K. Di Lu, Y. Zhang, Differential evolution-based convolutional neural networks: an automatic architecture design method for intrusion detection in industrial control systems, Comput. Secur. 132 (2023) 103310, https://doi.org/10.1016/j.cose.2023.103310.
[8] A. Rehman, K. Haseeb, T. Saba, J. Lloret, U. Tariq, Secured big data analytics for decision-oriented medical system using internet of things, Electron 10 (11) (2021) 1–13, https://doi.org/10.3390/electronics10111273.
[9] S. Kumar, S. Velliangiri, P. Karthikeyan, S. Kumari, S. Kumar, M.K. Khan, A survey on the blockchain techniques for the Internet of Vehicles security, Trans. Emerg. Telecommun. Technol. (2021) 1–23, https://doi.org/10.1002/ett.4317. June.
[10] Y.K. Saheed, S. Misra, S. Chockalingam, Autoencoder via DCNN and LSTM Models for Intrusion Detection in Industrial Control Systems of Critical Infrastructures," *2023 IEEE/ACM 4th Int. Work. Eng. Cybersecurity Crit. Syst. (EnCyCriS)*, Melbourne, Aust. (2023) 9–16, https://doi.org/10.1109/EnCyCriS59249.2023.00006.
[11] Y.K. Saheed, O.H. Abdulganiyu, T.A. Tchakoucht, Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the internet of things networks with edge capabilities, Appl. Soft Comput. 155 (2024) 111434, https://doi.org/10.1016/j.asoc.2024.111434. February.
[12] A. Manderna, S. Kumar, U. Dohare, M. Aljaidi, O. Kaiwartya, J. Lloret, Vehicular network intrusion detection using a cascaded deep learning approach with multi-variant metaheuristic, Sensors 23 (21) (2023) 8772, https://doi.org/10.3390/s23218772.
[13] J. Zeng, L.T. Yang, M. Lin, H. Ning, J. Ma, A survey: cyber-physical-social systems and their system-level design methodology, Futur. Gener. Comput. Syst. 105 (2020) 1028–1042, https://doi.org/10.1016/j.future.2016.06.034.
[14] G.C. Konstantopoulos A.T. Alexandridis, "Towards the Integration of modern power systems into a cyber – physical framework," 1–20, 2020.
[15] M.A. Mabayoje, J.F. Ajao, F.E. Usman-Hamza, Y.K. Saheed, K.A. Adeniran, Enhanced data storage security in cloud based on blowfish algorithm and text steganography, J. Niger. Comput. Soc. (2018).
[16] R.G. Jimoh, M.Y. Ridwan, O.O. Yusuf, Y.K. Saheed, Application of dimensionality reduction on classification of colon cancer using ICA and K-NN algorithm, Anale. Ser. Informatică 6 (10) (2018) 55–59 [Online]. Available, http://anale-informatica.tibiscus.ro/download/lucrari/16-1-06-Olatunde.pdf.
[17] Y.K. Saheed, O.T. Kehinde, M.A. Raji, U.A. Baba, Feature selection in intrusion detection systems: a new hybrid fusion of Bat algorithm and Residue Number System, J. Inf. Telecommun. (2023), https://doi.org/10.1080/24751839.2023.2272484.
[18] G. Karatas, O. Demir, O.K. Sahingoz, Increasing the performance of machine learning-based idss on an imbalanced and up-to-date dataset, IEEE Access. 8 (2020) 32150–32162, https://doi.org/10.1109/ACCESS.2020.2973219.
[19] S. W., Y.X.J. Zhang, L. Pan, Q.-L. Han, C. Chen, Deep Learning based attack detection for cyber-physical system cybersecurity: a survey, IEEE/CAA J. Autom. Sin. 9 (3) (2022) 377–391, https://doi.org/10.1109/JAS.2021.1004261.
[20] C. Systems, J. Yang, C. Zhou, "Anomaly detection based on zone partition for security protection of industrial," 65, 5, 4257–4267, 2018.
[21] H. Wang, Deep learning-based interval state estimation of AC smart grids against sparse cyber attacks, IEEE Trans. Ind. Informatics 14 (11) (2018) 4766–4778, https://doi.org/10.1109/TII.2018.2804669.
[22] S. Thakur, A. Chakraborty, R. De, N. Kumar, R. Sarkar, Intrusion detection in cyber-physical systems using a generic and domain specific deep autoencoder model, Comput. Electr. Eng. 91 (2020) 107044, https://doi.org/10.1016/j.compeleceng.2021.107044, 2021.
[23] M. Raiyat Aliabadi, M. Seltzer, M. Vahidi Asl, R. Ghavamizadeh, ARTINALI#: an efficient intrusion detection technique for resource-constrained cyber-physical systems, Int. J. Crit. Infrastruct. Prot. 33 (2021) 100430, https://doi.org/10.1016/j.ijcip.2021.100430.
[24] E.I. Ayei, B.O. Olusoji, A.A. Florence, A. Oladeji. Khadeejah, Novel hybrid model for intrusion prediction on cyber physical systems' communication networks based on bio-inspired deep neural network structure, J. Inf. Secur. Appl. 65 (103107) (2022), https://doi.org/10.1016/j.jisa.2021.103107.
[25] M. Attia, S.M. Senouci, H. Sedjelmaci, E.H. Aglzim, D. Chrenko, An efficient intrusion detection system against cyber-physical attacks in the smart grid, Comput. Electr. Eng. 68 (2017) 499–512, https://doi.org/10.1016/j.compeleceng.2018.05.006, 2018.
[26] M. Himanshu, K.T. Ashish, C.P. Avinash, S. Mohammad, Dahman Alshehri, Mukesh, P. Raju, A new intrusion detection method for cyber–physical system in emerging industrial IoT, Comput. Commun. 0140–3664 (2022) 24–35, https://doi.org/10.1016/j.comcom.2022.04.004, 190AD.
[27] S. Murugan, G.G. Deverajan, A. Kashif, R.P. Mahapatra, S. Mohammed, IADF-CPS : intelligent anomaly detection framework towards cyber physical systems, Comput. Commun. 188 (0140–3664) (2022) 81–89, https://doi.org/10.1016/j.comcom.2022.02.022.
[28] M.O. Hanafi, Abdulfatai Shola, Yakub Kayode Saheed, Arowolo, An effective intrusion detection in mobile ad-hoc network using deep belief networks and long short-term memory, Int. J. Interact. Mob. Technol. 17 (19) (2023) 123–135.
[29] O.H. Abdulganiyu, T.Ait Tchakoucht, Y.K. Saheed, A systematic literature review for network intrusion detection system (IDS), Int. J. Inf. Secur. (2023), https://doi.org/10.1007/s10207-023-00682-2.
[30] Y. Kayode, S. Sanjay, A voting gray wolf optimizer-based ensemble learning models for intrusion detection in the Internet of Things, Int. J. Inf. Secur. (2024), https://doi.org/10.1007/s10207-023-00803-x.
[31] Y.K. Saheed, B.F. Balogun, B.J. Odunayo, A. Mustapha, Microarray gene expression data classification via Wilcoxon sign rank sum and novel grey wolf optimized ensemble learning models, IEEE/ACM Trans. Comput. Biol. Bioinforma. (2023), https://doi.org/10.1109/TCBB.2023.3305429.
[32] Y.K. Saheed, Performance improvement of intrusion detection system for detecting attacks on internet of things and edge of things, in: S. Misra, T.K.A.V. Piuri, L. Garg (Eds.), Artificial Intelligence for Cloud and Edge Computing. Internet of Things (Technology, Communications and Computing, Springer, Cham, 2022, pp. 321–339. Eds.
[33] Y. Saheed, O. Longe, U.A. Baba, S. Rakshit, N.R. Vajjhala, An ensemble learning approach for software defect prediction in developing quality software product, in: M. Singh, V. Tyagi, P.K. Gupta, J. Flusser, T. Ören, V.R. Sonawane (Eds.), Advances in Computing and Data Sciences, Springer, Cham, 2021. Eds.
[34] S. Jadon, A survey of loss functions for semantic segmentation, 2020 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2020 (2020), https://doi.org/10.1109/CIBCB48159.2020.9277638.
[35] Z. Zhang, M.R. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, Adv. Neural Inf. Process. Syst. 2018-Decem (2018) 8778–8788. NeurIPS.
[36] Y.K. Saheed, Data analytics for intrusion detection system based on recurrent neural network and supervised machine learning methods. Recurrent Neural Networks, CRC Press Taylor & Francis Group, 2022, pp. 167–179.
[37] J.C. Duchi, P.L. Bartlett, M.J. Wainwright, Randomized smoothing for (parallel) stochastic optimization, Proc. IEEE Conf. Decis. Control 12 (2012) 5442–5444, https://doi.org/10.1109/CDC.2012.6426698.
[38] E.C. Orenstein, O. Beijbom, Transfer learning & deep feature extraction for planktonic image data sets, Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV (2017) 1082–1088, https://doi.org/10.1109/WACV.2017.125, 2017.
[39] S. Muhammad, S. Bukhari, S. Kumayl, R. Moosavi, M. Hamza, Federated transfer learning with orchard-optimized Conv-SGRU : a novel approach to secure and

accurate photovoltaic power forecasting, Renew. Energy Focus 48 (2024) (2023) 100520, https://doi.org/10.1016/j.ref.2023.100520.
[40] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, Ü. Budak, Transfer learning based histopathologic image classification for breast cancer detection, Heal. Inf. Sci. Syst. 6 (1) (2018), https://doi.org/10.1007/s13755-018-0057-x.
[41] J. He, K. Zhang, X. Ren, S., & Sun, Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778, https://doi.org/10.1002/chin.200650130.
[42] T.D. Jeff Donahue*, Yangqing Jia*, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, DeCAF: a deep convolutional activation feature for generic visual recognition Jeff, in: International conference on machine learning, 2014, pp. 647–655.
[43] A.K. Oladejo, T.O. Oladele, Y... Saheed, Comparative evaluation of linear support vector machine and K nearest neighbour algorithm using microarray data onleukemia cancer dataset, Afr. J. ComICT 11 (2) (2018) 1–10.
[44] L. Wen, X. Li, L. Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50, Neural Comput. Appl. 32 (10) (2020) 6111–6124, https://doi.org/10.1007/s00521-019-04097-w.
[45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90.
[46] H.K. Shin, W. Lee, J.H. Yun, H.C. Kim, HAI 1.0: HIL-based augmented ICS security dataset, in: CSET 2020 - 13th USENIX Work. CSET '20: 13th USENIX Workshop on Cyber Security Experimentation and Test Co-Located With USENIX Security, 2020, 2020.
[47] Y. Kayode Saheed, O. Harazeem Abdulganiyu, T. Ait Tchakoucht, A novel hybrid ensemble learning for anomaly detection in industrial sensor networks and SCADA systems for smart city infrastructures, J. King Saud Univ. - Comput. Inf. Sci. 35 (5) (2023) 101532, https://doi.org/10.1016/j.jksuci.2023.03.010.
[48] H. Shin, H.K. Lee, W. Yun, J. H., & Kim, "{HAI} 1.0:{HIL-based} Augmented {ICS} Security Dataset," 2020.
[49] N. Moustafa, J. Slay, The evaluation of Network Anomaly Detection Systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set, Inf. Secur. J. 25 (1–3) (2016) 18–31, https://doi.org/10.1080/19393555.2015.1125974.
[50] B.H. Menze, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinformatics 10 (2009) 1–16, https://doi.org/10.1186/1471-2105-10-213.
[51] E.A. Algehyne, M.L. Jibril, N.A. Algehainy, O.A. Alamri, A.K. Alzahrani, Fuzzy neural network expert system with an improved gini index random forest-based feature importance measure algorithm for early diagnosis of breast cancer in Saudi Arabia, Big Data Cogn. Comput. 6 (1) (2022), https://doi.org/10.3390/bdcc6010013.
[52] L. Akter, Ferdib-Al-Islam, M.M. Islam, M.S. Al-Rakhami, M.R. Haque, Prediction of cervical cancer from behavior risk using machine learning techniques, SN Comput. Sci. 2 (3) (2021), https://doi.org/10.1007/s42979-021-00551-6m.
[53] A.Q. Adeyiola, Y.K. Saheed, S. Misra, S. Chockalingam, Metaheuristic firefly and C5 . 0 algorithms based intrusion detection for critical infrastructures, in: 2023 3rd International Conference on Applied Artificial Intelligence (ICAPAI), 2023, pp. 1–7, https://doi.org/10.1109/ICAPAI58366.2023.10193917.
[54] M. Nguyen, N. A. V. B, SVMs With Deep Learning and Random, 2, Springer International Publishing, 2019.
[55] J. Hao, T.K. Ho, Machine learning made easy: a review of scikit-learn package in python programming language, J. Educ. Behav. Stat. 44 (3) (2019) 348–361, https://doi.org/10.3102/1076998619832248.
[56] Y.K. Saheed, U.A. Baba, M.A. Raji, Big data analytics for credit card fraud detection using supervised machine learning models, in: K. Sood, B. Balusamy, S. Grima, P. Marano (Eds.), Big Data Analytics in the Insurance Market (Emerald Studies in Finance, Insurance, and Risk Management, Emerald Publishing Limited, 2022, pp. 31–56. Eds.
[57] M. Wu, Z. Song, Y.B. Moon, Detecting cyber-physical attacks in Cyber Manufacturing systems with machine learning methods, J. Intell. Manuf. 30 (3) (2019) 1111–1123, https://doi.org/10.1007/s10845-017-1315-5.
[58] S.N. Shirazi et al., "Evaluation of anomaly detection techniques for SCADA communication resilience," 140–145, 2016.
[59] D.D. Anton, Anomaly-based intrusion detection in industrial data with SVM and random forests, in: International *Conference Software, Telecommunication Computer Networks*, 2019, pp. 1–6.