electronics

MDPI

Article

Ensemble Learning-Enabled Security Anomaly Identification for IoT Cyber–Physical Power Systems

Hongjun Zhao [1], Changjun Li [1], Xin Yin [1], Xiujun Li [1], Rui Zhou [2] and Rong Fu [2,*]

[1] State Grid Xinjiang Electric Power Corporation Limited, Urumqi 830011, China
[2] College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
* Correspondence: furong@njupt.edu.cn

Abstract: The public network access to smart grids has a great impact on the system's safe operation. With the rapid increase in Internet of Things (IoT) applications, cyber-attacks caused by multiple sources and flexible loads continue to rise, which results in equipment maloperation and security hazard problems. In this paper, a novel ensemble learning algorithm (ELA)-enabled security anomaly identification technique is proposed. Firstly, the propagation process of typical cyber-attacks was analyzed to illustrate the impact on message transmission and power operation. Then, a feature matching identification method was designed according to the sequence sets under different situations. The classification rate of these abnormal attack behaviors was acquired thereafter, which could aid in the listing of the ranking of the consequences of abnormal attack behaviors. Moreover, the weights of training samples can be further updated according to the performance of weak learning error rates. Through a joint hardware platform, numerical results show that the proposed technique is effective and performs well in terms of situation anomaly identification.

Keywords: cyber–physical power system (CPPS); Internet of Things (IoT) scenario; ensemble learning algorithm; analytic hierarchy process; abnormal attack behaviors

Citation: Zhao, H.; Li, C.; Yin, X.; Li, X.; Zhou, R.; Fu, R. Ensemble Learning-Enabled Security Anomaly Identification for IoT Cyber–Physical Power Systems. *Electronics* **2022**, *11*, 4043. https://doi.org/10.3390/electronics11234043

Academic Editors: Hamed Habibi, Afef Fekih, Silvio Simani and Amirmehdi Yazdani

Received: 14 October 2022
Accepted: 2 December 2022
Published: 5 December 2022

1. Introduction

The last decade has witnessed the increasing popularity of the Internet of Things (IoT) applications, in which physical devices connect with each other in high speed, low latency, and flexible ways. The IoT aims to achieve the goal of the interconnection of everything in industry, power systems, smart cities, and other fields [1]. In the train of such a tendency, the IoT cyber security of modern power systems has drawn increasing attention in recent years. With the amount connections between physical devices, perceiving cyber-attacks becomes a difficult task. Recent studies have begun to focus on cyberattack perception issues. Lallie et al. [2] proposed an adapted attack graph method to refine cyberattack perception. Liang et al. [3] established optimal attack models for different cyber-topology attack scenarios, and a metaheuristic optimization algorithm was proposed thereafter. Meanwhile, in Ref. [4], Zhang et al. designed a cyber-attack detection system, which was established on the concept of defense-in-depth, to enhance the cyber security of cyber–physical systems (CPSs). The dependence of the IoT cyber–physical system on information systems continues to increase, which highlights the importance of cyber security in the operation of the system.

Cyber-attacks in the context of the IoT has a significant influence on cyber–physical power systems (CPPSs). With the advancement of digital substations, the risks to IoT power systems risks will result in an era dominated by information-based risks. Many cyber-attack have happened in practice: in 2015, the Ukrainian blackout event was caused by a BlackEnergy cyber-attack, which caused the grid energy management system to fail. The Stuxnet cyber-attack also significantly affected the Iranian nuclear power plant SCADA

system, which laid too much burden on relevant security researchers [5]. Choraś et al. [6] proposed a machine learning-based technique to model behaviors and detect cyber-attacks. Meanwhile, Xin et al. [7] concluded that system operations become seriously vulnerable to line outages and failures if cyber–physical attacks occur.

Many fault diagnosis methods have been addressed in system fault detection. Different evaluation methods have been discussed for assessing the security of cyber–physical systems. Li et al. [8] proposed an enhanced IoT CPS security model, via which different cyber-topology attack scenarios can be recognized and classified. In [9], a risk assessment method was presented by Liu et al. to evaluate the cyber security issues with the aid of customized protection module.

On the other hand, Bi et al. [10] described the relationship between the context, the attack, the degree of vulnerability, and the network flow, which was proven to reflect the state of IoT CPS security. In [11], an IoT network security situation awareness model was designed by Xu et al. to enhance the reasoning ability of an ontology model. Xiao et al. [12] realized the online security assessment of the power grid via situation awareness technology. However, this type of method is not suitable for the real-time measurement of power grids [13,14]; when the information obtained by the state estimation is not accurate and comprehensive enough, this will cause a power grid dispatch center response delay and even ultimately damage to the operational stability of the power system [15].

Jinjie et al. [16] analyzed the situation of network topology being tampered with based on graph theory knowledge. In [17,18], two effective cyber-attack schemes for remote control units (RTUs) were proposed by Lallie et al. and Wang et al., respectively. Meanwhile, a method was proposed based on an improved attack graph to evaluate the hazard of cross-space cascading faults in cyber–physical power systems [19,20]. Zhu et al. [21] underlined the comparison requirement in a broader range of settings, which provided promising suggestions for further work.

At present, the research on the identification methods of network attacks on the grid side mainly focuses on the identification of malicious data injection, and the coverage of attack types is insufficient. Amir et al. [22] introduced a method to detect FDIAs targeting the AGC system by developing a stochastic unknown input estimator. The datasets in attack detection rely on the collection of abnormal traffic or vulnerability data [23] and multi-source network data information fusion [24]. The collection of the measured data considers single factors to analyze the overall operational state of the system, while the information fusion has the associated problems of the weak correlation between various elements and difficult data fusion, so it lacks mature and effective methods [25]. Recently, machine learning-based technologies were studied by Benisha et al. [26]., and Tahir et al. [27] proposed a false data injection attack detection with an adaptive distributed sampling sequence, which can improve detection efficiency while ensuring robustness; however, data fusion was processed to train a cyber-attack detector more effectively. To achieve better efficiency, ensemble learning techniques are proposed in attack detection.

Ensemble learning techniques can achieve higher accuracy and efficiency by improving the machine learning performances of diverse base learners. Ensemble learning methods train multiple base learners to obtain improved performance and become better at transferring information than each diverse base learner [28]. Meanwhile, the base learner algorithms result in high variance, high bias, and low accuracy. The ensemble learning algorithm usually chooses three main ensemble methods: bagging, boosting, and stacking. Several studies have shown that ensemble models often achieve higher accuracy than single machine learning models [29]. The fundamental idea behind ensemble learning is the recognition that machine learning models have certain limitations and can make errors. Hence, ensemble learning aims to improve abnormal identification performance by adjusting the weight coefficient of multiple base models. Ensemble methods can limit the variance and bias errors associated with single machine learning models; for example, bagging reduces variance without increasing the bias, while boosting reduces bias [30,31]. Overall, ensemble classifiers are more robust and perform better than the individual ensem-

ble learner. These methods have resulted in the promotion of security studies of CPPSs. Notwithstanding, the existing methods above mainly focuse on single-source data scenarios in IoT CPSs because it is difficult to trace the abnormality from physical faults and cyber-attack accurately. Meanwhile, research on multi-source heterogeneous data scenarios, which is an inevitable trend for IoT CPSs, is still lacking. Motivated by such a challenge, the main contributions of this paper are concentrated upon three aspects: (1) A hierarchical structure model of consequences under typical attack types was built and the effects of the abnormal behaviors were quantified. (2) Message data collected under different cyber-attacks was fitted by a novel ensemble learning algorithm (ELA) to identify anomalies, including power side data and information side status. (3) The security value and its probability under different cyber-attacks were determined by the combination of the feature matching method and ensemble learning algorithm. Aiming at the abnormal behaviors of cyber–physical power systems, a method to identify different attack types and physical faults is proposed to provide system situations through the proposed algorithm.

2. System Abnormal Behaviors in the IoT CPPS Hierarchical Structure

Figure 1 shows the typical architecture of an IoT CPPS [32,33], which includes the physical layer, sensor/actuator layer, network layer, control layer, and information layer.

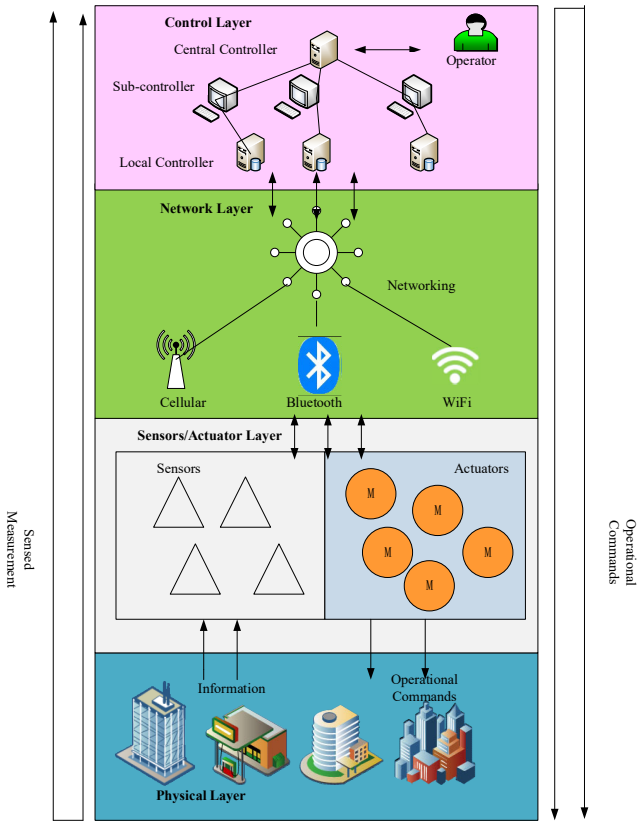

Figure 1. Architecture of an IoT CPPS.

Unlike the other layers, the network layer penetrates through the entire CPPS from bottom to top [34]. Specifically, the measurement data are delivered from sensors to control units. By contrast, the operation commands are delivered from control units to actuators. Cyber-attacks in an IoT CPPS usually interfere with the normal communication between the information layer and the physical layer, which affects the information layer, causing it to send the wrong instructions to the physical layer.

The cyber-attack first occurs in the information layer. As the abnormal attack behaviors mentioned are inconsistent with the characteristics of the data set, the consequences on the information side are difficult to distinguish. For example, network flow, protocol,

and business abnormality will cause the information side probe to detect a sudden change in the network flow. Table 1 describes six typical types of abnormal attack behaviors.

Table 1. Category of cyber attack.

Type	Name	Description
1	Denial of service attack	TCP, UDP
2	Abnormal port scan	ICMP.VNC
3	Violent cracking	Ssh brute force
4	Network flow anomaly	Abnormal direction
5	Protocol abnormal	Message header abnormal
6	Business abnormal	Exceeds the standard

According to the above mentioned six typical CPS cyber-attacks and abnormal message, the type of attack is classified according to "CIA", with "C" representing confidentiality, "I" referring to integrity, and "A" denoting availability.

(a) "A" is for the purpose of destroying the available communication information:

(1) Denial-of-service attack.

A denial-of-service attack is a type of resource exhaustion attack. It uses the defects of network protocol/software or sends many useless requests to exhaust the resources of the attacked object (such as its network bandwidth), so that the server or communication network cannot provide normal services. DoS attacks can be generally divided into the following four categories: ◎ using protocol vulnerabilities to attack (such as in a SYN flood attack); ◎ using software defects to attack (such as in an OOB attack, teardrop attack, land attack, IGMP fragment packet attack, etc.); ◎ sending a large number of useless requests to occupy resources (such as in an ICMP flood attack, connection flood attack, etc.); and ◎ a blocking buffer spoofing attack (such as an IP spoofing DoS attack):

(2) Abnormal port scan.

Abnormal port scans include ICMP scanning, 3389 external scanning, vnc scanning, etc. For example, the Windows system prohibits 135, 137, 138, 139, 445, 3389, and other high-risk ports vulnerable to malicious attacks as service ports.

(b) "C" is to destroy the confidentiality of information.

(3) Violent cracking.

Violent cracking includes SSH brute force cracking, FTP brute force cracking, etc., and refers to the most widely used attack techniques, with hackers using the password dictionary to guess the user's password by employing exhaustive methods.

(c) "I" is to destroy data integrity.

(4) Network flow anomaly.

Network flow anomalies includes abnormal directions, sizes and types of network flow:

(5) Abnormal protocol.

Abnormal protocols include message header format exception, the length of the format exception, and control domain format exceptions, etc.

(6) Abnormal business.

The cumulative number of remote-control trips exceeds the limit, and the interval value in terms of transmission times exceeds the limit; thus, the telemetry information exceeds the limit.

It is difficult to identify abnormal attack behavior on the data of information side. The abnormal information behavior propagates to the physical layer, which affects the stability of the frequency, the voltage values of the buses, and the current values of flexible loads. In this paper, the flexible loads are considered to be interruptible loads and transferable

loads. Considering the constraints of power supply interruption capacity and times, the model of providing power for interruptible load is as follows:

$$\underline{P}_{\mathrm{IL},j} \le P_{\mathrm{IL},j,t} + r^{\mathrm{U}}_{\mathrm{IL},j,t} \le \overline{P}_{\mathrm{IL},j}\ j \in \Omega_{\mathrm{IL}} \tag{1}$$

$$\sum_{t\in\Omega_{\mathrm{T}}} u_{\mathrm{IL},j,t} \le N_{\mathrm{IL},j} \tag{2}$$

where $P_{\mathrm{IL},j,t}$ is the power of interruptible load j at time t, $r^{\mathrm{U}}_{\mathrm{IL},j,t}$ is the upper reserve capacity of interruptible load j at time t, $\underline{P}_{\mathrm{IL},j}$ and $\overline{P}_{\mathrm{IL},j}$ are the minimum and maximum values of interruptible load j, respectively, Ω_{IL} is the collection of interruptible loads, and $u_{\mathrm{IL},j,t}$ is the state variable of interruptible load j at time t. If $u_{\mathrm{IL},j,t} = 1$, this indicates that the standby capacity of load j at time t is interruptible; otherwise, if the $u_{\mathrm{IL},j,t} = 0$, this indicates that it is not interruptible. Ω_{T} is the statistical time set, and $N_{\mathrm{IL},j}$ is the maximum number of interruptions allowed for interruptible load j.

The power consumption of the transferable load can be flexibly adjusted, while the total power consumption in a dispatching cycle is unchanged. The model is as follows:

$$\underline{P}_{\mathrm{SL},k} \le P_{\mathrm{SL},k,t} \le \overline{P}_{\mathrm{SL},k}\ k \in \Omega_{\mathrm{SL}} \tag{3}$$

$$\underline{P}_{\mathrm{SL},k} \le P_{\mathrm{SL},k,t} + u^{\mathrm{U}}_{k,t} r^{\mathrm{U}}_{\mathrm{SL},k,t} \le \overline{P}_{\mathrm{SL},k} \tag{4}$$

where $P_{\mathrm{SL},k,t}$ is the power consumption of the transferable load k at time t, $\underline{P}_{\mathrm{SL},k}$ and $\overline{P}_{\mathrm{SL},k}$ are the minimum and maximum values of transferable load k, respectively, Ω_{SL} is a collection of transferable loads, $r^{\mathrm{U}}_{\mathrm{SL},k,t}$ is the adjusted capacity of the transferable load k at time t, and $u^{\mathrm{U}}_{k,t}$ is the state variables of the transferable load k at time t.

The flexible loads have time variable and stochastic characters. The probability density function of the flexible loads can be described as:

$$f(P_{fl}) = \frac{1}{\sqrt{2\pi}\sigma_{fl}} \exp\left(-\frac{(P_{fl}-\mu_{fl})^2}{2\sigma^2_{fl}}\right) \tag{5}$$

where P_{fl} is the active power of the flexible load, μ_{fl} is the average value of the active power in several dispatching cycles, and σ_{fl} is the standard deviation of active power.

Referring to the aforementioned abnormal attack behaviors, the hierarchy of abnormal consequences is shown in Figure 2.

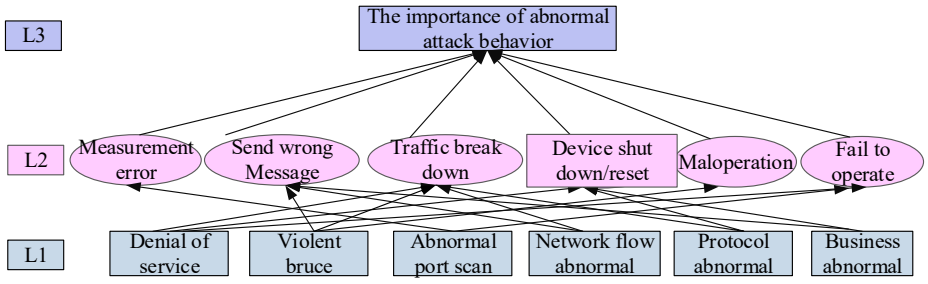

Figure 2. Correlation analysis under different abnormal modes.

The attacker can obtain the port number through the vulnerability scanning device. After obtaining the port number, the attacker can use the operating system and the device memory vulnerability to log in through the Xshell software and connect with the device IP, so that the attacker can invade the editing interface and execute the attack instructions. The attacker can tamper with the instructions of the master station to attack the stability control device in the execution station, thus affecting the power side. Table 2 shows different abnormal behaviors and the caused possible maloperations.

Table 2. Consequence hierarchical analysis of abnormal behavior hazards.

Abnormal Behavior	Possible Performances with Abnormality
Denial of service	Traffic breakdown; device shut down/reset; failure to operate
Violent brute force	Send wrong message; traffic breakdown; maloperation
Abnormal port scan	Measurement error; failure to operate
Network flow abnormal	Send wrong message; traffic breakdown
Protocol abnormal	Traffic breakdown; device shut down/reset
Business abnormal	Send wrong message; device shut down/reset

3. Feature Matching Method Based on Cyber–Physical Cooperation

The cyber-attacks and fault events targeting the power system may cause abnormal phenomena on the user side. When the training data is sufficient, the gradient decision tree is used to detect the abnormality of the power system. Therefore, the current research has the problem of whether the abnormal phenomenon is caused by network attacks or failures. For example, certain common faults in a power system, such as different phase-to-phase short-circuit faults or offline faults, may cause the system to fail to operate. Meanwhile, network attacks such as man in middle attacks and denial of service attacks may cause system maloperation. All failures caused by cyber-attacks and physical faults are classified as abnormal phenomena.

To analyze whether the abnormal phenomenon is caused by a network attack or a fault, a feature sequence matching method is proposed based on cyber–physical cooperation. The causes of abnormal phenomena cannot be analyzed from the section data alone; the state transition process should be also considered. Taking denial of service (DoS) attacks as an example, the consequences caused by a DoS attack are similar to the communication delay in the system under normal conditions. How to extract features from the process of state transition to trace the causes of abnormal phenomena is the subject to be analyzed. Therefore, a feature sequence matching method was used for different abnormal phenomena to trace the cause.

3.1. Establish a State Transition Matrix

The physical elements are collected from the SCADA (Supervisory Control and Data Acquisition) system, which can reflect the operating process on the physical side. The information data mainly includes network traffic changes, log records of commands, modification records of the master station, etc. The information data mainly represents the network, protocol, and system, etc., so as to reflect the abnormal phenomena. To establish the state transition matrix, various physical elements and information data are handled to cooperate with the cyber–physical data.

Firstly, to discretize the physical analog data, the specific method is to discretize the continuous value into a value that can be used to represent the anomaly states. The advantage is that the data fluctuation caused by a small disturbance can be ignored, and only the abnormal change in the system state needs to be focused on. Considering that large disturbances, such as a frequency stability problem after line failures, the numbers impact little on the simulating results. Therefore, our specific aim was to discretize the continuous value into the a value that can be used to represent the voltage state, current state, and frequency state. The advantage of this process is that the data fluctuation caused by small disturbances can be ignored, and only the change in the system state needs to be focused on when large disturbances occur, which can be easily handled in the training and learning process. Thus, the physical measurement data includes the bus voltage, line current, system frequency, and control signals such as relay protection device action records and security control records.

3.1.1. Voltage Value Processing Method

The setting of the voltage rating generally involves certain standards according to different equipment. The equipment generally operates under the rated voltage value U_n, so it is discretized as a judgment of voltage level. The measured voltage value is converted into a standard unit value. If the measured voltage value $U \leq 0.9U_n$, the discretized voltage state S_U is described as 0. If $0.9U_n < U \leq 1.1U_n$ and the discretized voltage state S_U is described as 1. Otherwise, if $U > 1.1U_n$, the discretized voltage state S_U is described as 2.

3.1.2. Current Value Processing Method

The line current processing method is when the measured current values of multiple loads are compared with the current values in the steady state I_s. The sharp fluctuation in the measured current can be regarded as abnormal. The measured current value is converted into a standard unit. If the measured current value $I \leq 0.9I_s$, the discretized current state S_I is described as 0. If $0.9I_s < I \leq 1.5I_s$ and the discretized current state S_I is described as 1. Otherwise, if $I > 1.5I_s$, the discretized current state S_I is described as 2.

3.1.3. Frequency Processing Method

A fluctuation in the frequency between 0.996 and 1.004 of the rated frequency value f_n is considered normal, while a fluctuation below $0.996f_n$ or above $1.004f_n$ is abnormal. Therefore, the frequency can be divided into three levels, as in the case of the voltage value processing method. If the measured frequency value $f \leq 0.996f_n$, the discretized frequency state S_f is described as 0. If $0.996f_n < f \leq 1.004f_n$ and the discretized frequency state S_f is described as 1. Otherwise, if $f > 1.004f_n$, the discretized frequency state S_f is described as 2.

3.1.4. Information Processing Method

Since the information values are originally discrete quantities, it is necessary to convert the event contents into numerical values.

The development process of each event can be represented by the system state transition process with a subscript. When an event occurs, an event state transition table can be obtained and each state can be numbered. After discretization, the states of each physical measurement or information value are limited, so the total number of the transition table is also limited. In the state transition table, the features can be extracted from the state transition process to represent the anomaly of this event.

Therefore, the state transition table E_1 can be expressed $E_1 = [S_1, S_2, \ldots, S_m]^T$, which can be shown as follows:

	t	$freq$uency	v_1	$\cdots$	v_n	i_1	$\cdots$	i_n	$switch_1$	$\cdots$	$switch_n$	$comm_1$	$\cdots$	$comm_n$
S_1	t_1	f_1	$v_{1,1}$	$\cdots$	$v_{1,n}$	$i_{1,1}$	$\cdots$	$i_{1,n}$	$switch_{1,1}$	$\cdots$	$switch_{1,n}$	$comm_{1,1}$	$\cdots$	$comm_{1,n}$
S_2	t_2	f_2	$v_{2,1}$	$\cdots$	$v_{2,n}$	$i_{2,1}$	$\cdots$	$i_{2,n}$	$switch_{2,1}$	$\cdots$	$switch_{2,n}$	$comm_{2,1}$	$\cdots$	$comm_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
S_m	t_m	f_m	$v_{m,1}$	$\cdots$	$v_{m,n}$	$i_{m,1}$	$\cdots$	$i_{m,n}$	$switch_{m,1}$	$\cdots$	$switch_{m,n}$	$comm_{m,1}$	$\cdots$	$comm_{m,n}$

In the horizontal direction of E_1, t is the record time, v_n is the measured voltage value in node n, i_n is the measured voltage value in node n, $switch_n$ is the switch state in node n, and $comm_n$ is the measured information value in the order of n.

Since the record discrete information value is inconsistent in the time dimension with the physical values, it is necessary to make up the values in the same time dimension. Thus, the event state transition table can be used to facilitate the analysis of change during the state transition process.

3.2. Feature Extraction during the State Transition Process

When an abnormal phenomenon occurs, there is no operational difference on the physical side, and it is difficult to distinguish between physical faults and cyber-attacks. Due to the various processes of the system state transition among different events, the extracted feature sequence will also be different. Thus, based on cyber–physical cooperative

detection, the real cause of the abnormal phenomenon can be traced and the abnormal phenomenon can be classified.

Through the generation of the event state transition sequence, the characteristic events on the information side can help distinguish the types of cyber-attack and physical faults. For the preset event simulations, the feature sequences that meet the preset possibility degree are extracted from the event sequence set, and the possibility degree related to the feature sequences is formed from the preset event. Among them, the feature sequences corresponding to different possibility degrees will be different for different events. Through the idea of sequence matching, i.e., by comparing whether the sequence of unknown events contains the same sequence as the feature events extracted from a certain type of event, the unknown events are classified.

Abnormal identification can be used to analyze an abnormal behavior through the preset feature sequences. If the abnormal identification method traverses the database to detect the difference, the efficiency is slow. Therefore, to improve the accuracy and speed of identification, the ensemble learning algorithm is proposed for abnormality identification.

4. Abnormality Identification Based on the Ensemble Learning Algorithm

The machine learning solution process can be regarded as looking for a learning model with a good generalization ability and robustness in in terms of classification, but it is not easy to find an appropriate model in different situations. Therefore, as a combinatorial learning method, the ensemble learning algorithm can not only form an excellent combinatorial model by combining multiple single models, but it can also design flexible strategies for specific machine learning problems to obtain more useful solutions.

4.1. Ensemble Learning Algorithm for Anomaly Label Classification

Ensemble learning is the combination of several base machine learning learners to form a model with a smaller variance, smaller deviation, or better classification prediction effect. The ensemble learning algorithm can summarize the selection of base learners, model training, and model combination in integrated learning. In a CPPS, abnormality identification accuracy can be improved by an ensemble learning algorithm when considering the training of multiple sub-learners.

4.2. Selection of Base Learners

To guarantee classification accuracy in ensemble learning, certain algorithms are introduced as base learner candidates: the decision tree (DT), radial basis function (RBF), backward propagation (BP), least square support vector machine (LSSVM), and extreme learning machine (ELM) algorithms, as shown in Table 3.

Table 3. Comparison of several regression algorithms in ensemble learning.

Algorithm	Advantage	Disadvantage
Decision tree (DT)	Easy to understand and implement	Discrete output
Least square support vector machine (LSSVM)	High accuracy	Parameter optimization
Radial basis function (RBF)	Strong generalization ability	Slow training rate
Backward propagation (BP)	Parameter optimization	Parameter optimization
Extreme learning machine (ELM)	Quickness	Simple network structure

Since ensemble learning has a diversity property, the characteristics of its different algorithms can be different. Theoretically, to improve the diversity of the ensemble learning model, the number of base learners should be large. Such a preference, however, is opposed to computational situations in practice. Thus, a tradeoff should be considered according to practical requirements. Generally, the error of the ensemble model is related to both the accuracy and the ambiguity of the base learners. In other words, the performance

of the ensemble model has a strong correlation with the accuracy or ambiguity of the base learners. Therefore, DT, RBT, and ELM base learners were selected from five candidate datasets in the ensemble learning algorithm.

4.2.1. Select Decision Tree (DT) as a Base Learner

In the DT method, a classification tree and a regression tree are involved. The Gini index can be defined as:

$$gini(D^k) = \sum_{i=1}^{n} \sum_{i' \neq i} p_i p_{i'} = 1 - \sum_{i=1}^{n} p_i^2 \tag{6}$$

which is used to determine the order of the internal nodes. In the above equation, D is the dataset, $gini(D^k)$ is the Gini index of the k-th class in D, and p_i is the proportion of the i-th class. To feature k in D, the *Gini* index can be expressed as:

$$Gini(D,k) = \sum_{k=1}^{K} \frac{\left|D^k\right|}{|D|} gini(D^k) \tag{7}$$

where K is the class amount, $|D^k|$ is the number of the k-th class, and $|D|$ is the sample amount.

When all the features have been sorted by the *Gini* index, a complete classification tree can be easily established via arranging all the features in sequence.

4.2.2. Select Radial Basis Function (RBF) as a Base Learner

The RBF neural network is designed using radial basis functions, and there are no local minimum points and slow learning rates, as shown in Figure 3. The Gaussian function is used as the kernel function in the hidden layer nodes, which can be written as:

$$R_i(Z_k) = G\|Z_k - C_i\| = \exp\left[-\frac{\|Z_k - C_i\|}{2{\sigma_i}^2}\right] \tag{8}$$

where Z_k is the k-th sample vector; C_i is the center of the i-th hidden layer neuron; σ_i is the variable of the hidden layer node i; and $\|Z_k - C_i\|$ is the norm.

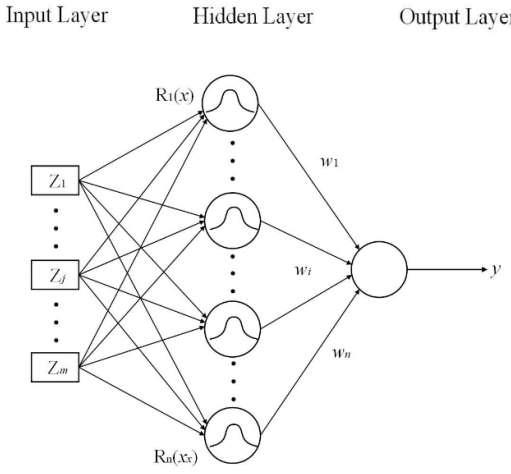

Figure 3. RBF neural network structure.

The output value of the network is:

$$y = \sum_{i=1}^{n} \omega_i R_i(Z_k) \tag{9}$$

where y is the output value, and ω_i is the network weight between the hidden layer nodes and the output layer nodes.

4.2.3. Select Extreme Learning Machine (ELM) as a Base Learner

Compared to the RBF algorithm, which features an iterative parameter generation process, ELM, which instead features a random parameter generation process, has a much faster training speed. The structure of ELM is shown in Figure 4.

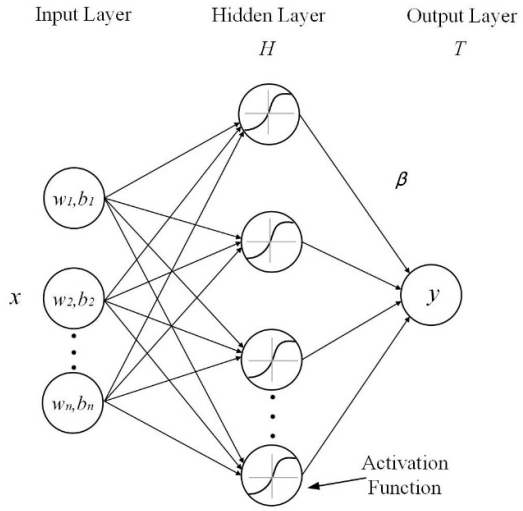

Figure 4. ELM neural network structure.

The mathematical relation can be expressed as:

$$H\beta = T \tag{10}$$

where H denotes the hidden layer output matrix, β represents the output weight vector connecting hidden layer and output layer, and T is the target output matrix. In practice, the output weight vector β can be solved by $\hat{\beta} = H^{+}T$, where H^{+} is the Moor–Penrose generalized inverse of H.

4.3. Basic Learner Combination under Ensemble Learning

Ensemble learning for abnormality identification is achieved via the following steps. (1) The training subset is generated by randomly sampling the dataset, which can effectively reduce the training time consumption; (2) each subset is then input into each algorithm (DT, RBT, ELM) correspondingly, which ensures that the algorithm can comprehensively learn data characteristics; (3) the input of the test set into each model will then provide the identification results; (4) lastly, the results are aggregated to obtain the final abnormal identification results. The entire processes above can be visualized as Figure 5 illustrates.

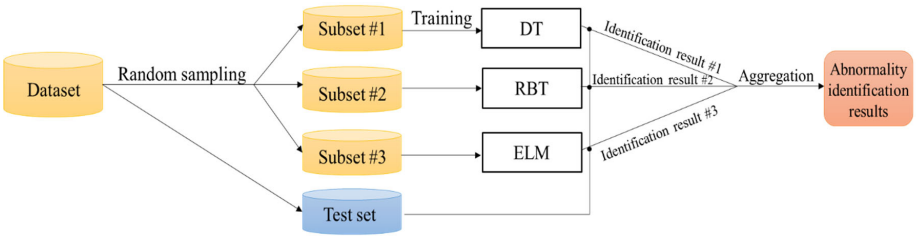

Figure 5. Basic learner combination process under ensemble learning.

4.4. Abnormality Identification under Ensemble Learning

The ensemble learning process to identify attack behaviors can be described as follows.

Firstly, the weight distribution of training data is initialized. The initial weight w of the data set D_1 under the machine learning algorithm is divided by the average value:

$$D_1 = \{w_{1,1}, w_{1,2}, \ldots, w_{1,i}, \ldots, w_{1,N}\}, i = 1, 2, \ldots\ldots, N \tag{11}$$

where N is the size of the data set D_1, $w_{1,i} = \frac{1}{N}$.

In the experiment, the characteristic values of the network sample data are time stamp, message length, source, and destination address, source and destination port, data offset, check code, message version number with the length, identifier, serial number, confirmation number, check sum, emergency pointer, voltage values, current values, frequency value, switch states, and so on. As these alarm data only contain information layer data, the power change, frequency oscillation amplitude, voltage drop value, and operation condition of the related measuring equipment caused by the information layer fault in the dataset are combined.

Secondly, the training data set are used to obtain the basic classifier $G_m(x)$ with the weight distribution D_m. $D_m = \{w_{m,1}, w_{m,2}, \cdots, w_{m,i}, \cdots, w_{m,N}\}$, m = 1, 2, ... , M. In the ELA machine learner, DT, RBF, and ELM were chosen as the three different base learners.

Then, the classification error rate e_m of $G_m(x)$ on the training data set is calculated, with y_i referring to the output, which is defined in the training date.

$$e_m = Prob(G_m(x_i) \neq y_i) = \sum_{i=1}^{N} w_{m,i} I(G_m(x_i) \neq y_i) \tag{12}$$

where $G_m(x_i)$ represents the type of abnormal behavior under m-th base learner, while y_i is the actual type of the abnormality in the dataset.

According to the e_m obtained by each base learner algorithm, a_m is used to determine whether this part of the data set should increase the weight distribution in the next iteration.

$$a_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \tag{13}$$

When the weight distribution of the training data set is updated, the weight $w_{m+1,i}$ can be written as:

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m} \exp(-a_m y_i G_m(x_i)) \tag{14}$$

where $w_{m+1,i}$ refers to the weight updated in D_{m+1}, $D_{m+1} = \{w_{m+1,1}, w_{m+1,2}, \cdots, w_{m+1,i}, \cdots, w_{m+1,N}\}$. Z_m is the normalization factor, which makes D_{m+1} become a probability distribution.

$$Z_m = \sum_{i=1}^{N} w_{m,i} \exp(a_m y_i G_m(x_i)) \tag{15}$$

Finally, a linear combination of basic classifiers is constructed:

$$f(x) = \sum_{m=1}^{M} a_m G_m(x) \tag{16}$$

This shows that the smaller the classification error rate, the greater the role of the classifier in the final classifier. After that, the weight distribution of training data can be updated for the next round preparation. Equation (14) can then be written as:

$$w_{m+1,i} = \begin{cases} \frac{w_{m,i}}{Z_m} e^{-am}, G_m(x_i) = y_i \\ \frac{w_{m,i}}{Z_m} e^{am}, G_m(x_i) \neq y_i \end{cases} \tag{17}$$

In feature selection, cross-validation is used to divide the training set features, which are divided into 10 sub-training sets. Each sub-training set has been trained. It is understood that the time stamp, message length, source and destination address, source and destination port, message version, and other features in 30 sub-training sets have a significant influence on the feature identification result. Therefore, the data eigenvalues obtained from the above cross-validation are used as the dataset for ensemble learning.

It can be seen that the weight of samples misclassified by the basic classifier is increased, and the weight of samples correctly classified is reduced. Thus, the weight of the misclassified samples is enlarged by $e^{2am} = \frac{e_m}{1-e_m}$ multiples, so that the misclassified samples play a vital role.

Through the combination of these measures, the bias and variance of the result can be used to identify different abnormal behaviors. An indicator RSME is used to express the quality of the established model when ensemble learning is used to solve regression problems,

$$RSME = \frac{1}{m}\sum_{i=1}^{m}\Delta psw\_m \tag{18}$$

$$\Delta p_{sw\_m} = |psw\_m(actual) - psw\_m(prediction)| \tag{19}$$

where Δp_{sw_m} represents the absolute difference between the real situation value and the predicted situation value under the training data of group m, $psw\_m(actual)$ refers to the actual situation value of the group m, and $\Delta psw\_m(prediction)$ means the prediction value of group m.

The model of machine learning is very dependent on data. $Bias^2$ means the square of deviation and *variance* refers to the complexity of the training set. Figure 6 shows the optimum model complexity under ensemble learning. The errors in ELA are computed as $(Bias^2 + variance)$. Generally, the more complex the model, the lower the deviation, which leads to the reduction in the total error. Therefore, the optimal model should balance the deviation and variance.

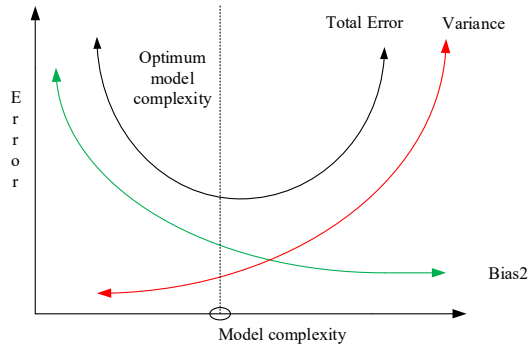

Figure 6. Optimum model complexity under ensemble learning.

5. Case Verification

5.1. Simulation Platform

From the above analysis, we can see that network attacks and failures have many similar features in the process of feature transfer. Both the information side and the physical side increase the difficulty of identification, and the existence of a communication delay further increases the difficulty in terms of identification. Therefore, the hardware in the loop-based joint simulation platform is used to conduct batch data simulation and feature variable screening of a large number of original data, providing the original basis for event sequence generation and feature sequence extraction based on the ensemble learning algorithm.

According to the collected information side and the physical side state, the discretized cyber–physical sequence training set D is formed, the base learners are initialized from Equation (11), and the number of training rounds is set to 30 in the ensemble learning al-

gorithm. Then, the classification error rates are calculated from Equation (12). The weight of the classifier and the weight distribution of the training data set are calculated using Equation (14). During the training process, the ensemble learning size is tested to divide different training subsets. Finally, Equation (16) is applied to combine the base classifiers to obtain the final strong classifier.

Batch data simulations are carried out based on the hardware joint simulation platform, which provides original data for event sequence generation. The simulated CPPS architecture is shown in Figure 7.

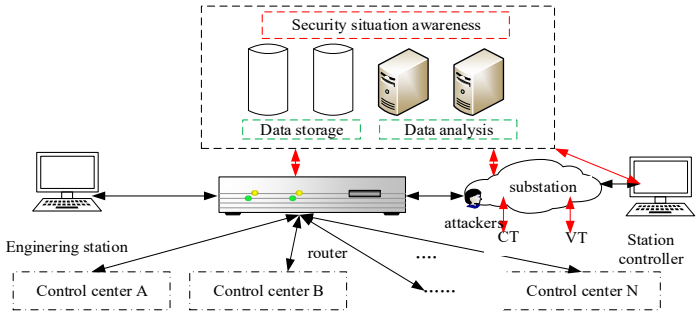

Figure 7. Abnormality identification scene test in a CPPS.

To verify the effectiveness of the proposed machine learning algorithm, the dataset was analyzed based on a variety of test events in a CPPS. The packets were captured through the Wireshark packet capture tool and analyzed by the communication protocol to obtain partial datasets.

The software simulation platform mainly consisted of RT-LAB and OPNET, as shown in Figure 8. The main power system network was modeled in RT-LAB to form an equivalent physical network to a power system. RT-LAB uses Ethernet to communicate with the outside, and the main communication protocols were TCP/IP and UDP protocols. The network interface included communication control, packet acceptance, and sending modules. The sampling frequency of measured data was 1000 Hz. OPNET allows for the detailed modeling of communication devices and the simulation of a digital informative system.

Figure 8. Hardware joint simulation platform in CPPS.

5.1.1. Abnormal Behavior Identification in a 4-Gens CPPS

A 4-Gens CPPS was built in the hardware joint simulation platform, as shown in Figure 9. The fault protection workflow can be described as follows. When a fault process is detected in this area, the corresponding control action is to disconnect the circuit breakers, and the connection between switch R1 and switch R2 is disconnected. At this time, the voltage between bus conductor B3 and bus conductor B8 is overloaded. According to the protection action, bus conductor B3′s load will be immediately cut off, i.e. the linkage L1 becomes disconnected, so as to prevent the accident from causing more serious phenomena and maintain the stability of the power grid.

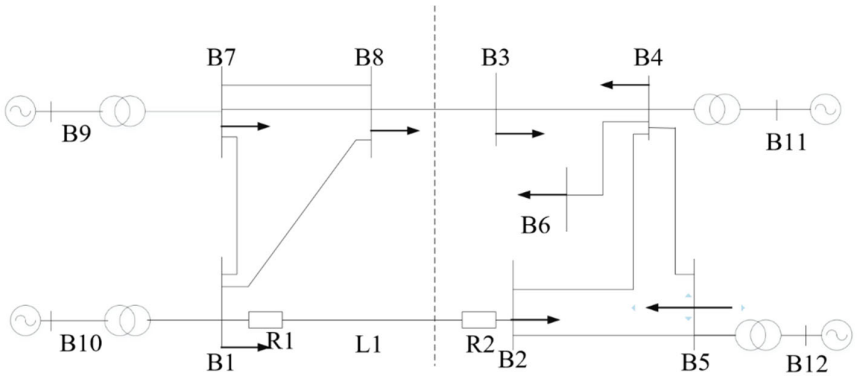

Figure 9. Fault protection workflow of test system in a 4-Gens CPPS.

The communication network was built in OPNET to simulate different communication environments. Wireshark was used to acquire the data in the information layer with different abnormal behaviors. Based on the fault protection workflow in Figure 9 above, our following simulations were carried out to show the influence of network attacks or physical failures. Without loss of generality, the basic configuration of a two-area four-generations power system were adopted, as in the widely acknowledged method used in [35]. Figure 10 showcases the tendencies of the communication transmission delays between devices and the control center. Considering the information signals with a transmission delay, the IoT CPP is a multi-delay system. In this experiment, communication environments 1, 2, and 3 were set with different locations. Specifically, environment 1 (nodes 1–4) and environment 3 (nodes 9–12) were set for no fault process scenarios, while environment 2 (nodes 5–8) was set for fault detected scenarios. To fully represent input delay in each communication environment, we introduce a single time varying delay $d(t)$ satisfying $0 \leq d(t) \leq d_m$. In communication environments 1, 2, and 3, the d_m was 0.02, 0.05, 0.1 s, respectively. Different delay messages were obtained in the different communication environments and applied in the operation schemes.

As we can observe in Figure 10, in average delay performance in practical situations is marginally worse than that in test situations. This is due to the situations designed for test situations being limited, while the practical negative occurrences cannot be traversed via testing. In practice, a small number of uncontrollable factors (e.g., processing request interruptions, communication distortions, etc.) will affect the actual delay performance. The second phenomenon to observe relates to the comparison of different communication environments. Specifically, nodes belonging to a same kind of communication environment have similar delay performances. Environment 3 (nodes 9–12) has a higher average delay performance than that of environment 1 (nodes 1–4), which is mainly caused by the the average divergence (about 0.8 s) between their time-varying delays. While for environment 2 (nodes 5–8), the delay performance is distinctly degenerated by the influence of fault detection and the protection process, as the above Figure 9 shows. Fortunately, as the practical and steady operation of a power system is tolerable for second class delay oscillations [36], such a gap (highest about 0.6 s) among different communication environments is acceptable.

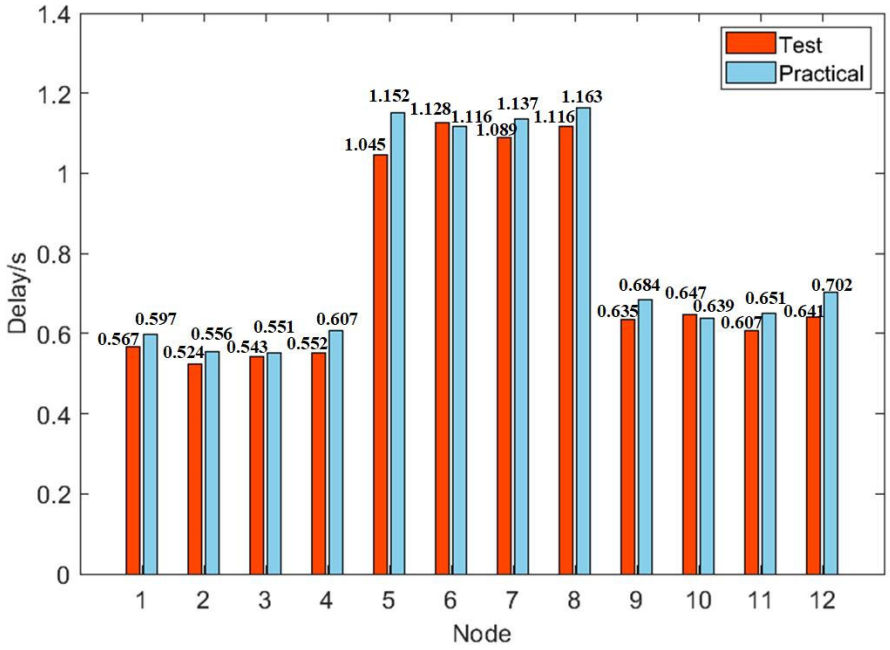

Figure 10. Communication transmission delays between devices and the control center.

5.1.2. Feature Sequence Extraction

Based on the measured data from the informative and physical side, collaborative datasets can be used to form the state transition matrix. The feature sequences are extracted from the state transition matrix. In the test system platform, more than 2000 events were simulated including network attacks and physical failures. Similar test scenarios of the consequences were classified as caused by simulated faults. The sampling scenarios of the test cases are shown in Table 4.

Table 4. The sampling scenarios of simulations.

Classification Number.	Scenarios	Scenario Description
No. 1	Q1–Q3	Short-circuit grounding faults on line L1 with different communication delays
No. 2	Q4–Q6	Remote control of the relays at R1 and R2 of line L1 by the dispatcher
No. 3	Q7–Q12	Tripping request forgery attack at R1 or R2 of line L1 with different communication delays by the attacker
No. 4	Q13–Q15	Tripping request forgery attack on tripping protection equipment of line L1
No. 5	Q16–Q19	Tripping request forgery attack on the changing protection equipment thresholds of line L1
No. 6	Q20–Q23	The physical failure of the protection equipment to operate flexible loads
No. 7	Q24	Normal scenario

To take scenario Q1 after the three-phase short-circuit grounding fault and scenario Q4 after the tripping command forgery attack as examples of a comparison, these events lead to the action of the protection device and physical changes.

Table 5 shows a portion of the data sequence in a short circuit grounding fault. In this Table, state refers to system state order number, where S1–S10 means 10 states corresponding to 10 moments; t(s) refers to the time parameter; VB1–VB4 and IB1–IB4 refer to the discretized current state corresponding to the measured voltage and circuit values in node 1–4; f refers to the discretized frequency state corresponding to the system frequency; R1 and R2 refer to the state of switches R1 and R2. If the switch is on, the state value is 1, otherwise if the switch is off and the value is 0. During the state transition process, VB3 changes in S1–S2, circuit breaker R1 acts in S2 to S3, seven measure quantities change in S3–S4, three measure quantities change in S4–S5, and the remaining number of state changes is one or two. The feature sequence set is listed with the largest state change in characteristic

quantity. The matching probability is calculated as the percentage of the data matching sequence length to the common data sequence length. Therefore, the extracted feature sequence is S3-S4-S5, and the matching probability is 30%.

Table 5. Data sequence in a short circuit grounding fault.

State	t(s)	VB1	VB2	VB3	VB4	IB1	IB2	IB3	IB4	f	R1	R2
S1	0	1	1	1	1	1	1	1	1	1	1	1
S2	25.064	1	1	0	1	1	1	1	1	1	1	1
S3	25.083	1	1	0	1	1	1	1	1	1	0	1
S4	25.084	0	0	0	0	1	1	2	2	2	0	0
S5	25.124	0	0	0	1	1	2	2	2	1	0	0
S6	25.125	1	0	1	1	1	2	2	2	1	0	0
S7	25.126	1	0	1	1	1	2	2	1	1	0	0
S8	25.131	1	0	1	1	1	2	1	1	1	0	0
S9	25.133	1	0	2	1	1	2	1	1	1	0	0
S10	50.314	1	0	1	1	1	1	1	1	1	0	0

In scenario Q4, a tripping request forgery attack leads to the disconnection of R1 and R2. In the scenarios in classification No. 1, if the line between node B3 and B8 is overloaded, the protection device acts to cut off B3 load. However, under the cyber-attack in the scenarios in classification No. 2, the attacker would control the protection device to refuse to act when the forgery command is injected. Therefore, the state transition process under the tripping request forgery attack can be obtained in Table 6.

Table 6. Data sequence in a tripping request forgery attack.

State	t(s)	VB1	VB2	VB3	VB4	IB1	IB2	IB3	IB4	f	R1	R2
S_{W-1}	17.751	1	1	1	1	1	1	1	1	1	1	1
S_{W-2}	18.811	1	1	1	1	1	1	1	1	1	1	0
S_{W-3}	18.777	1	1	1	1	1	1	1	1	1	1	0

Since the main change in the forgery attack state is the information side state during the transition, the extracted feature sequence is S_{W_1}-S_{W-2}.

5.1.3. Abnormal Event Matching Process

When matching the events in the feature matching method, the data sequence S1–S10 is compared with the feature database. If the features can be fully matched, it will be regarded as a three-phase short-circuit grounding fault and the matching probability is set to 100%. If the matching value is not successful, the data sequence S3-S4-S5 is tried for feature matching and the matching possibility is reset as 30%. If the matching value is successful, the event is regarded as a fault similar to the short-circuit event.

When matching tripping request forgery attacks and other phenomena occur, the data sequence S_{W_1}-S_{W-2}-S_{W-3} is used to match in the feature database. If the matching is successful, it is considered a tripping request forgery attack and the probability is set to 100%. If the matching is unsuccessful, the data sequence S_{W_1}-S_{W-2} is used to match again. If the matching is successful, the event is regarded as a forgery attack but not conclusively a tripping request attack, and the matching probability is set to 67%.

5.1.4. Selection of Ensemble Learning Size

To showcase the influence of the ensemble learning size, we simulated and obtained the accuracy of ensemble learning for different ensemble learning sizes, as Figure 11 illustrates. As we can see in Figure 11a, the initial classification accuracy $p = 0.5$ is a watershed, at which the final ensemble learning accuracy will be kept constant, regardless of the learning size. When the initial classification accuracy $p < 0.5$, with the increase in learning size, the final ensemble learning accuracy will continue to decline towards zero. On the contrary,

Figure 11b reflects that when the initial classification accuracy $p > 0.5$, the final ensemble learning accuracy will approach 100%.

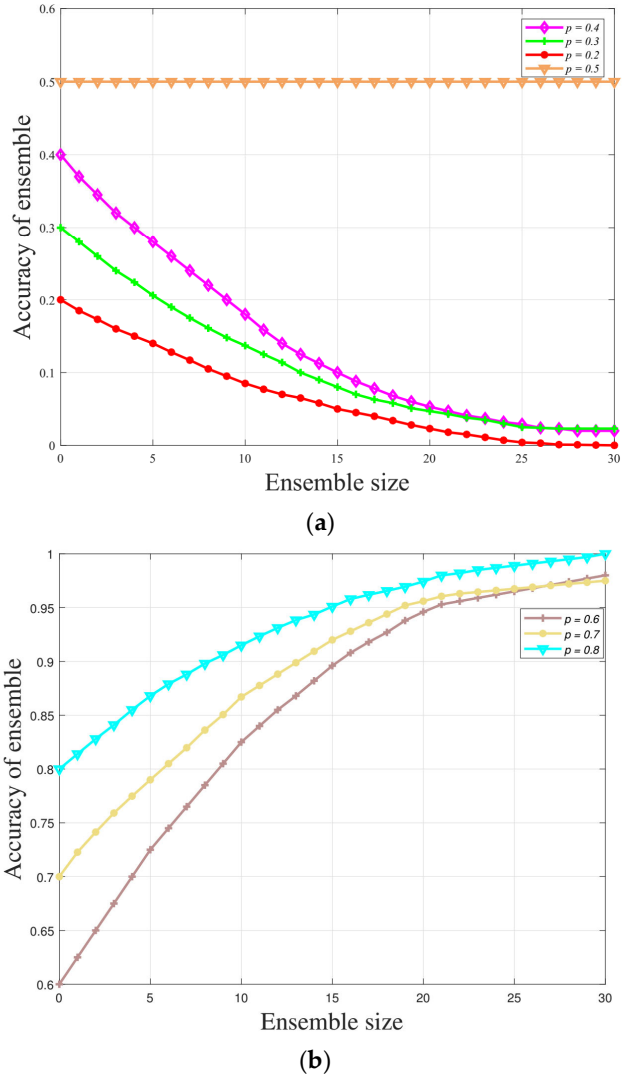

Figure 11. The accuracy of ensemble learning for different ensemble learning sizes.

5.1.5. Accuracy Comparison of Different Methods

The proposed ELA was used to identify abnormal behavior with the extracted feature matching method. The data sequences were selected as training import values, and the abnormal behavior types and identification probability were chosen as training output values.

The sequence matching results for the sampling scenarios are shown in Figure 12. There was no error in the classification of Q1–Q3, Q7–Q15. There was misjudgment between Q7 and Q13 in the classification of Q4, and certain misjudgments among Q5, Q9, Q10, and Q14. Similarly, certain scenarios in Q16, Q17, Q20, Q22, and Q23 were classified incorrectly.

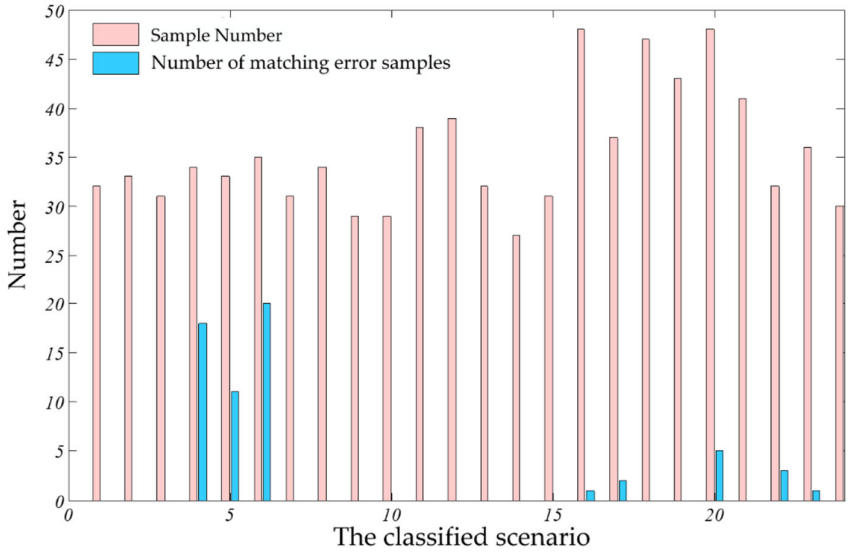

Figure 12. Abnormal phenomenon feature matching results.

The total classification accuracy rate was 91.5% (88/1040). The cyber-attack classification accuracy rate was 85.0% (63/420), and the physical faults classification accuracy was 96.0% (25/620). The classification running time was within the range of 30 s to several minutes.

If the error rate which deviates from the normal distribution is regarded as an error value, the box chart can intuitively show those data deviate from the normal data distribution. As shown in Figure 13, certain test results were selected to show the iterative updating process of the weights in ELA. The red "+" indicates the degree of error rate deviation from the normal data distribution during the learning process. This means that ELA can increase the weight of this data set in the next test to obtain a better fitting effect if the error rate deviation is large.

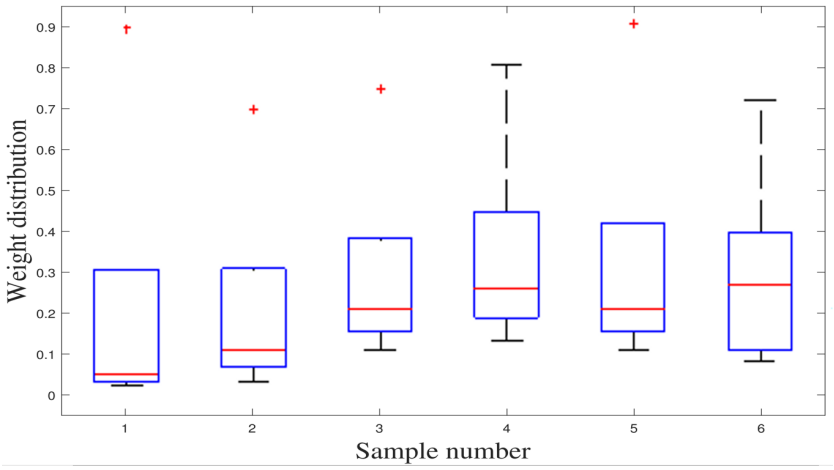

Figure 13. The box diagram of the sample weight distribution.

Figure 14 shows the accuracy of the proposed ELA with RBF and BP algorithms in attack type identifications. Compared with traditional machine learning results, ELA effectively greatly improves the identification accuracy. In this case, the accuracy in terms of BP is not satisfactory. Such a phenomenon results from the network falling into a local extreme value, which should be induced by the selection of the initial connection weight as well as the threshold value of the neural network.

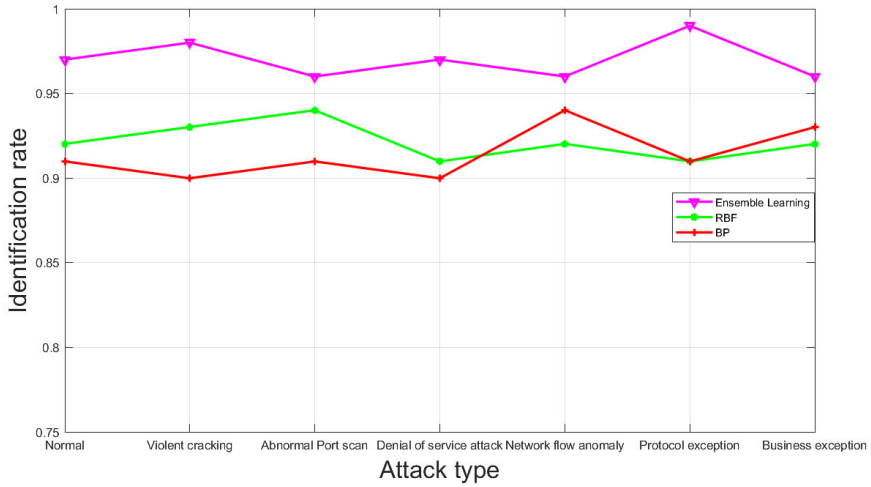

Figure 14. Comparison between traditional machine learning and ensemble learning.

It can be seen clearly that the prediction rate has been greatly improved. The identification accuracy is greatly improved by ELA. The selected 2018-line message data generated in the attacks was verified, and the 404-line training data ensured a 100% recognition rate. The accuracy of the test data ratio was 1533/1614, and the average recognition rate rose to 97%. ELA can gradually eliminate the possibility of large errors in training the unprocessed data.

In order to reduce the phenomenon of scenario misjudgment, the training effectiveness of ELA was compared, and the weight coefficients of the misjudgment data were increased in the integrated machine learning classification. Without loss of generality, we adopted the widely adopted ELM technique [37] as the benchmark to showcase the performance of the proposed ELA technique. In order to reduce errors, more feature paths were adopted in the feature library after specific analysis. After 10 rounds of verification, the results shown in Figure 15 were obtained.

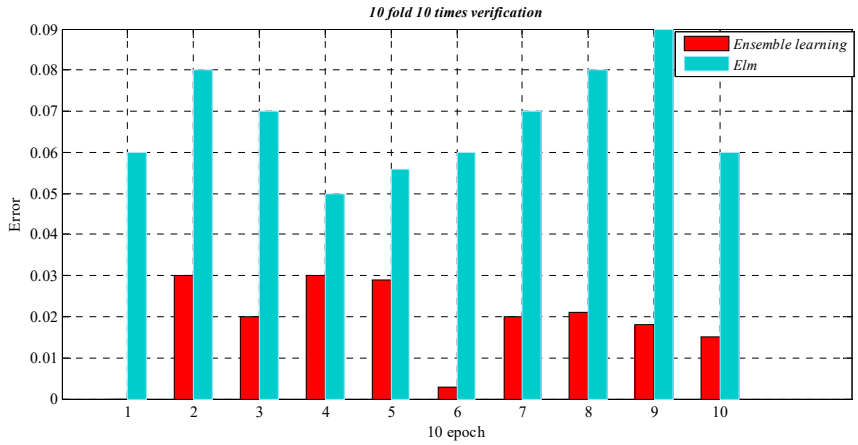

Figure 15. The errors for a cyber–physical power system security situation.

The results show that the classification effects of certain samples affected by delay was significantly improved. This means that the majority of the abnormal features in the sequence matching results were information side state abnormalities, so the integrated machine learning algorithm can identify whether the same abnormal phenomenon is caused by cyber-attacks or physical faults. ELA can improve the generalization ability of the model through the feature verification of the training set and the test set, thus improving the accuracy in terms of identification.

Another interesting phenomenon in Figure 15 is the contrast between the ELM and ELA techniques. As we know, ELM is a promising learning algorithm developed for training single-hidden layer feedforward neural networks which has a fast learning speed and an excellent generalization capability [38]. However, in Figure 15, ELM shows an unsatisfactory error performance in the context of CPPS security when compared to the proposed ELA technique. Such a result is reasonable because the randomness of input weights and bias may result in unstable and diverse results [39]. The randomness involved in ELM parameters results in individual ELMs suffering from degradation in terms of consistency and robustness [40]. Confronted with these factors, ELM cannot perfectly address security anomaly identification applications for IoT CPPSs. By contrast, ELA provides a good opportunity for designing ensemble learning models since the randomness involved can inherently increase the diversity of an ensemble learner and significantly improve the classification accuracy.

6. Conclusions

This paper studied the security anomaly identification issue for IoT CPPSs. The research work jointly evaluated the consequences of abnormal behaviors as well as the processing of information transmissions in physical faults. Multiple abnormal behaviors were identified by an ensemble learning algorithm, and the consequences of abnormal attack behaviors were quantified by employing a feature matching method. Specifically, the abnormal behaviors in different scenarios were first analyzed in the abnormal spread process from the information layer to the physical layer. Then, using an ensemble learning based feature matching method, the system behavior identification rate was largely improved. The key factors in terms of the physical model extracted from the state data were combined through three base learners in an ELA. The DT, BP, and ELM base learners were ensembled after verifying the relationship between classification accuracy and training subset size. The numerical results showed that the proposed technique distinctly improved the identification ratios of the CPPS with multiple flexible loads, which is promising for the practical deployments of IoT CPPSs.

Author Contributions: Conceptualization, H.Z., C.L. and X.Y.; methodology, H.Z. and R.Z.; software, R.F.; validation, R.F.; formal analysis, H.Z.; investigation, H.Z.; resources, H.Z., C.L. and X.Y.; data curation, X.L.; writing—original draft preparation, H.Z. and C.L.; writing—review and editing, H.Z. and X.L.; visualization, C.L.; supervision, R.Z.; project administration, H.Z.; funding acquisition, H.Z. and R.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Technology Project of State Grid Xinjiang Electric Power Co., Ltd. "Research on the key technologies of Xinjiang's new multi-load in power grid operation", No. 5230HQ22000C.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical concerns.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, J.; Xu, T.; Zhou, T.; Chen, X.; Zhang, N.; Hu, H. Feature-based Spectrum Sensing of NOMA System for Cognitive IoT Networks. *IEEE Internet Things J.* **2022**. [CrossRef]
2. Lallie, H.S.; Debattista, K.; Bal, J. An Empirical Evaluation of the Effectiveness of Attack Graphs and Fault Trees in Cyber-Attack Perception. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1110–1122. [CrossRef]
3. Liang, G.; Weller, S.R.; Zhao, J.; Luo, F.; Dong, Z.Y. A Framework for Cyber-Topology Attacks: Line-Switching and New Attack Scenarios. *IEEE Trans. Smart Grid* **2019**, *10*, 1704–1712. [CrossRef]
4. Zhang, F.; Kodituwakku, H.A.D.E.; Hines, J.W.; Coble, J. Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4362–4369. [CrossRef]

5. Zhang, J.; Sankar, L. Physical System Consequences of Unobservable State-and-Topology Cyber-Physical Attacks. *IEEE Trans. Smart Grid* **2015**, *7*, 2016–2025. [CrossRef]
6. Choraś, M.; Kozik, R. Machine Learning Techniques Applied to Detect Cyber Attacks on Web Applications. *Log. J. IGPL* **2015**, *23*, 45–56. [CrossRef]
7. Xin, S.; Guo, Q.; Sun, H. Cyber-Physical Modeling and Cyber-Contingency Assessment of Hierarchical Control Systems. *IEEE Trans. Smart Grid* **2015**, *6*, 2375–2385. [CrossRef]
8. Li, F.; Shi, Y.; Shinde, A.; Ye, J.; Song, W. Enhanced Cyber-Physical Security in Internet of Things Through Energy Auditing. *IEEE Internet Things J.* **2019**, *6*, 5224–5231. [CrossRef]
9. Liu, X.; Shahidehpour, M.; Li, Z.; Liu, X.; Cao, Y.; Li, Z. Power System Risk Assessment in Cyber Attacks Considering the Role of Protection Systems. *IEEE Trans. Smart Grid* **2017**, *8*, 572–580. [CrossRef]
10. Bi, S.; Wang, T.; Wang, L.; Zawodniok, M. Novel Cyber Fault Prognosis and Resilience Control for Cyber–Physical Systems. *IET Cyber-Phys. Syst. Theory Appl.* **2019**, *4*, 304–312. [CrossRef]
11. Xu, G.; Cao, Y.; Ren, Y.; Li, X.; Feng, Z. Network Security Situation Awareness Based on Semantic Ontology and User-Defined Rules for Internet of Things. *IEEE Access* **2017**, *5*, 21046–21056. [CrossRef]
12. Xiao, J.; Zhang, B.; Luo, F. Distribution network security situation awareness method based on security distance. *IEEE Access* **2019**, *7*, 37855–37864. [CrossRef]
13. Ranjbar, M.H.; Kheradmandi, M.; Pirayesh, A. A Linear Game Framework for Defending Power Systems Against Intelligent Physical Attacks. *IEEE Trans. Smart Grid* **2019**, *10*, 6592–6594. [CrossRef]
14. Mathaios, P.; Peter, A.; Daniel, S. Assessing the impact of insufficient situation awareness on power system operation. *IEEE Trans. Power Syst.* **2017**, *28*, 2967–2977.
15. Zhao, J.; Gomez-Exposito, A.; Netto, M.; Mili, L.; Abur, A.; Terzija, V.; Kamwa, I.; Pal, B.C.; Singh, A.K.; Qi, J. Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work. *IEEE Trans. Power Syst.* **2019**, *34*, 3188–3198. [CrossRef]
16. Jinjie, T. Topological Attacks on Smart Grids: Undetectable Attacks and Solutions. *J. Am. Soc. Electr. Electron. Eng. Commun.* **2018**, *31*, 1294–1305.
17. Dai, Q.; Shi, L.; Ni, Y. Risk Assessment for Cyberattack in Active Distribution Systems Considering the Role of Feeder Automation. *IEEE Trans. Power Syst.* **2019**, *34*, 3230–3240. [CrossRef]
18. Pascale, F.; Adinolfi, E.A.; Coppola, S.; Santonicola, E. Cybersecurity in Automotive: An Intrusion Detection System in Connected Vehicles. *Electronics* **2021**, *10*, 1765. [CrossRef]
19. Wang, H.; Chen, Z.; Zhao, J.; Di, X.; Liu, D. A Vulnerability Assessment Method in Industrial Internet of Things Based on Attack Graph and Maximum Flow. *IEEE Access* **2018**, *6*, 8599–8609. [CrossRef]
20. Wang, L.; Xu, P.; Qu, Z.; Bo, X.; Dong, Y.; Zhang, Z.; Li, Y. Coordinated Cyber-Attack Detection Model of Cyber-Physical Power System Based on the Operating State Data Link. *Front. Energy Res.* **2021**, *88*, 1–9. [CrossRef]
21. Zhu, Z.; Wang, Z.; Li, D.; Zhu, Y.; Du, W. Geometric Structural Ensemble Learning for Imbalanced Problems. *IEEE Trans. Cybern.* **2020**, *50*, 1617–1629. [CrossRef] [PubMed]
22. Amir, A.; Hooshyar, A.; Yazdavar, A.H.; El-Saadany, E.F.; Youssef, A. Attack Detection for Load Frequency Control Systems Using Stochastic Unknown Input Estimators. *IEEE Trans. Inf. Forensics Secur.* **2018**, *10*, 2575–2585.
23. Zhang, H.; Shi, J.; Chen, X. A Multi-Level Analysis Framework in Network Security Situation Awareness. *Procedia Comput. Sci.* **2013**, *17*, 530–536. [CrossRef]
24. Mets, K.; Verschueren, T.; Develder, C. Integrated Simulation of Power and Communication Networks for Smart Grid Applications. In Proceedings of the IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, Kyoto, Japan, 10–11 June 2011; pp. 61–65.
25. Kwon, Y.; Lee, S.; King, R.; Lim, J.I.; Kim, H.K. Behavior Analysis and Anomaly Detection for a Digital Substation on Cyber-physical System. *Electronics* **2019**, *3*, 326. [CrossRef]
26. Benisha, R.; Ratna, S.R. Design of Intrusion Detection And Prevention in SCADA System for the Detection of Bias Injection Attacks. *Secur. Commun. Netw.* **2019**, *2019*, 12–16. [CrossRef]
27. Tahir, B.; Jolfaei, A.; Tariq, M. Experience Driven Attack Design and Federated Learning Based Intrusion Detection in Industry 4.0. *IEEE Trans. Ind. Informat.* **2022**, *9*, 6398–6405.
28. Giffen, B.; Herhausen, D.; Fahse, T. Overcoming the Pitfalls and Perils of Algorithms: A Classification of Machine Learning Biases and Mitigation Methods. *J. Bus. Res.* **2022**, *144*, 93–106. [CrossRef]
29. Guo, H.; Zhang, J.; Zhang, J.; Li, Y. Prediction of Highway Blocking Loss Based on Ensemble Learning Fusion Model. *Electronics* **2022**, *11*, 2792. [CrossRef]
30. Doroudi, S. The Bias-variance Tradeoff: How Data Science Can Inform Educational Debates. *AERA Open* **2020**, *4*, 1413–1414. [CrossRef]
31. Alelyani, S. Stable Bagging Feature Selection on Medical Data. *J. Big Data* **2021**, *1*, 11–19. [CrossRef]
32. Saghezchi, F.B.; Mantas, G.; Violas, M.A.; de Oliveira Duarte, A.M.; Rodriguez, J. Machine Learning for DDoS Attack Detection in Industry 4.0 CPPSs. *Electronics* **2022**, *11*, 602. [CrossRef]

33. Andronie, M.; Lăzăroiu, G.; Iatagan, M.; Uță, C.; Ștefănescu, R.; Cocoșatu, M. Artificial Intelligence-Based Decision-Making Algorithms, Internet of Things Sensing Networks, and Deep Learning-Assisted Smart Process Management in Cyber-Physical Production Systems. *Electronics* **2021**, *10*, 2497. [CrossRef]
34. 3GPP Technical Specifcation Group Radio Access Network. *Evolved Universal Terrestrial Radio Access and Evolved Universal Terrestrial Radio Access Network*; Overall Description; Stage 2, Release 15, TS 36.300 V15.3.0; 3GPP Mobile Competence Centre: Valbonne , France, 2018.
35. Cheng, Z.; Yue, D.; Hu, S.; Huang, C.; Dou, C.; Chen, L. Resilient Load Frequency Control Design: DoS Attacks Against Additional Control Loop. *Electr. Power Energy Syst.* **2020**, *115*, 105496–105509. [CrossRef]
36. Wang, S.; Meng, X.; Chen, T. Wide-Area Control of Power Systems Through Delayed Network Communication. *IEEE Trans. Contr. Syst. Technol.* **2012**, *20*, 495–503. [CrossRef]
37. Rath, S.K.; Sahu, M.; Das, S.P.; Bisoy, S.K.; Sain, M. A Comparative Analysis of SVM and ELM Classification on Software Reliability Prediction Model. *Electronics* **2022**, *11*, 2707. [CrossRef]
38. Li, H.; Zhao, H.; Li, H. Neural-Response-Based Extreme Learning Machine for Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 539–552. [CrossRef] [PubMed]
39. Chen, Z.; Jiang, C.; Xie, L. A Novel Ensemble ELM for Human Activity Recognition Using Smartphone Sensors. *IEEE Trans. Ind. Informat.* **2019**, *15*, 2691–2699. [CrossRef]
40. Khamis, A.; Xu, Y.; Dong, Z.Y.; Zhang, R. Faster Detection Of Microgrid Islanding Events Using an Adaptive Ensemble Classifier. *IEEE Trans. Smart Grid.* **2018**, *9*, 1889–1899. [CrossRef]