IDS Detection Based on Optimization Based on WI-CS and GNN Algorithm in SCADA Network

S. Shitharth, N. Satheesh, B. Praveen Kumar, and K. Sangeetha

Abstract Industry control systems (ICS) are considered as one of the inevitable systems in this contemporary smart world. In that supervisory control and data acquisition (SCADA) is the centralized system that control the entire grid. When a system is considered to be a whole and sole control, obviously an uncompromised security would be the prime. By having that as a major concern, a lot of research is being done on IDS security. In spite of that it has several cons including increased fake positive and fake negative rates, which will invariably lead to a larger chaos. To get rid of these problems, a weighted-intrusion based cuckoo search (WI-CS) and graded neural network (GNN) methods are proposed in this chapter. The key purpose of this chapter is to identify and categorize the anomalies in a SCADA system through data optimization. At initial stage, the collected real-time SCADA dataset is given as input. Then, by using the aforementioned proposed machine learning algorithms, these data are clustered and optimized. Later to find, the type of intrusion will remain as a further challenge and for that we propose HNA-AA algorithm. The investigational results estimate the efficiency of the system by considering sensitivity, false detection rate, precision, recall, Jaccard, accuracy, dice and specificity.

Keywords SCADA · Cukoo search · Neural network · Intrusion detection system · Clustering · Feature optimization

S. Shitharth (✉)
Department of CSE, Vardhaman College of Engineering, Hyderabad 501218, India

N. Satheesh
Department of CSE, St. Martin's Engineering College, Hyderabad 500041, India

B. P. Kumar
Department of EEE, Bharat Institute of Engineering and Technology, Hyderabad 501510, India

K. Sangeetha
Department of CSE, Sri Satya Sai University of Technology and Medical Sciences, Sehore, Madhya Pradesh 466001, India

S. K. Das et al. (eds.), *Architectural Wireless Networks Solutions and Security Issues*, Lecture Notes in Networks and Systems 196,
https://doi.org/10.1007/978-981-16-0386-0_14

1 Introduction

SCADA collects information from the remote systems which is a part of ICS [1, 2]. It is an automation system that is used to receive the data from various sensors and components located at remote sites and transmit the data at a master site for monitoring purposes. SCADA systems are mainly used to both control and monitor various industrial applications. The significant and most essential duty is to detect intrusions in SCADA network. In real life, every country is moving toward a digital and smart environment. Smart grid is one of the prime setups in making a smart city. Marinating the smart grid security is always a vital area to look out and that is been done by SCADA systems. SCADA is primarily used to monitor electric power generation, transmission and distribution in smart grid systems. Along with that water sewage system, traffic signal controlling, mass transit, manufacturing, building facilities and environment are the other prime real-time uses of a SCADA system [3]. Intrusion detection is the process used to detect and monitor such malicious activity on a network. Monitoring the network manually is quite impossible. The foremost target of IDS [4–6] is to defend the availability, integrity and confidentiality of the system by digging out the malicious nodes, intrusions and destructive attacks. The IDS achieves the subsequent processes: Both user and system behavior were observed and analyzed. Then, the network configurations and its vulnerabilities are audited and also assess the reliability of critical organization and data files. The research work by Santosh and Sachin elaborated on [7] methodologies like software defining network (SDN) along with linear programming in ad hoc network are used to overcome uncertainties and imprecisions. These stats again have a larger impact while calculating confusion matrix results in any network security. Based on calculating those results and matching of well-known attacks, activity patterns are analyzed along with the irregular activities of the nodes in the system are also analyzed. The Modus operandi of anomaly detection in SCADA system is shown in Fig. 1.

Intrusion detection systems (IDS) are used to observing network and system activities to detect malicious activities. It can be categories into network-based intrusion detection system (NIDs) and host-based intrusion detection system (HIDs) [8, 9]. For protecting host computer networks and monitoring networks, the host-based IDS [10, 11] are used largely and also to spot the burglar attack activities, malicious actions nodes in the system and application anomalies. It classifies and prevents the risky incidents to investigate and observe the actions happening over the host system. The network-based IDS [12, 13] openly examines the traffic in the network system. For supervising the intrusive actions, based on its invariable changing state, it collects system's data initially. Then, finding and identifying the intrusions in the SCADA network becomes a significant and tedious task. Perhaps, it has its own disadvantages like complex installation and frequent network overheads. The major issue with the existing approaches is its high fake alarm rates and classifying unknown attacks and its feature absence in the intrusion attack library. To get rid of these issues, this paper proposes an enhanced IDS system and an optimization techniques using machine learning algorithm. The prime motto of this work is to pick the finest attributes for

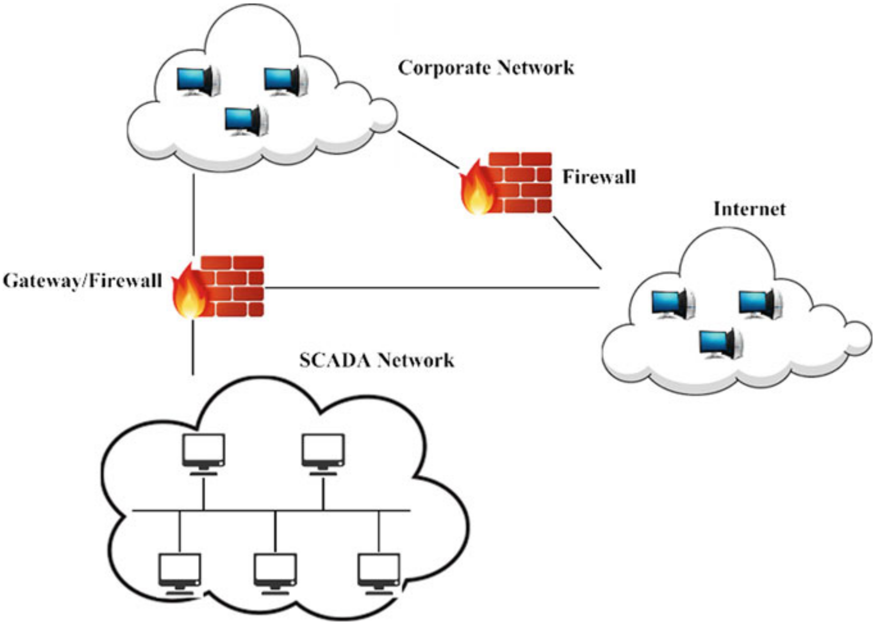


Fig. 1 Anomaly detection in SCADA network

optimization using WI-CS algorithm. In order to categorize the attacks or intrusions accurately in a SCADA system, a GNN categorization method is developed. The grading or hierarchical method is used in order to calculate the performance of the developed system. Moreover, this research work uses the real-time power system dataset which has been subjected to intrusions. The other sections of this chapter are structured as follows: Sect. 2 is more of literature survey that takes hold of previous research methods done in SCADA IDS. Section 3 is all about the full description of the developed WI-CS and GNN method, and the Sect. 4 assesses the enactment of the developed methods with already available methods for claiming the improved performance. The section V concludes in a way how this work is carried forward for future enhancements and concluded.

2 Related Works

This related work is more of a literature survey that explains previous research works done by various authors in SCADA IDS. A novel IDS idea given by Almalawi et al. [14], which identified only the usual and dangerous criteria that degrade the system efficiency. The prime focus of this paper is followed by:

- State recognition

- Extracting own decision statements on detection
- Decrease of high false positive rate
- Compute to assess criticality.

The major drawback of this proposed method is that the work did not dig deep into the constancy modifications in the network. Mitchell and Chen [15] proposed some major takeaways such as:

- To create a logical methodology for counteracting the attack
- To identify and find out the dissimilar failures types in industry control systems (IPs).

The major drawback of this research is, CPS system's solidity is been questioned during the intrusion times. Based on telemetry study, an approach to find out anomalies in SCADA system was developed by Ponomarev and Atkison [16]. In this paper, using master–slave process, the traffic is well divided between sender and receiver. The telemetry-based IDS supervised all packets in the ICS system, so it detected the irregularity in traffic. To spot the control-related attacks in SCADA system, a semantic analysis framework was designed by Lin et al. [17]. In this research, with a power flow analysis, the control results are optimized. A regular expression (RegEx) was proposed by Liu and Torng [18]. The supplies contain nondeterministic finite automata (NFA) size and deterministic finite automata (DFA) speed which is made more scalable and adaptable. Moreover, the condition copying and evolution copying were arrested using the projected automata mode. Marchang et al. [19] developed a mobile ad hoc network (MANET) to decrease the period of lively moment without compromising their efficiency. Here, a behavioral observance of nodes is calculated at a specific check point to reduce the energy consumption of the nodes in the network. Dirichlet-based detection scheme (DDOA) was introduced by Li et al. [20] mainly to identify the outliers residing in the smart grid. They done it by using the real-time IEEE power system database and by using a software tool called power world simulator. Ambusaidi et al. [21] proposed a method using square vector to find out the outliers in SCADA. They used KDD and NSL dataset for feature classification and optimization.

Hasan et al. [22] proposed a trust-based monitoring system for incoming and outgoing traffic. The main objective is to get rid of excessive capital and operational cost used by the nodes and he has done it by calculating trust values if every node in the system. Yang et al. [23] proposed a research methodology for the attacks having dissimilarity and parameter variation among themselves. He used behavioral approach of the nodes and try to whitelist the genuine nodes. Based on this, the author can separate the inliers and outliers from the network. A three-layer IDS for providing protection to SCADA was developed by Samdarshi et al. [24]. Based on partition system and dividing the IoT devices based on their router gateway, it is the major takeaway from this research. The core intent of this research is to use various cybersecurity applications use the trust system placement scheme. Sayegh et al. [25] proposed a way for IDS based on patterns and traffic of the incoming nodes in the network. This research includes the below five stages:

- Intrusion whiffing
- Attribute extraction
- Training phase
- Testing phase
- Anomaly detection.

Now, the relationships among the packet were anticipated to recognize its authentication. The problem of attack isolation and intrusion detection was investigated by Amin et al. [26]. The objectives of this research are as follows:

- It provides solutions for anomaly detection
- It considers the sensor data for attribute classification.

Maglaras et al. [27] proposed a vector classification which is used to improve the efficiency by a numerical method. In addition by using clustering methodology, it classifies the SCADA alarm into low, medium and high. The major problem of this technique is the need to reduce the fake alarm rate. Yang et al. [28] proposed an IDS to track malicious interventions in the SCADA network. The objective of this work was based on the signature-based rules and it detects the unknown attacks and the major benefit of this work was to detect the outliers accurately. Yasakethu and Jiang [29] proposed a methodology using Markov model that finds out the intruders. It also uses Markov matrix plot for data representation. This is again based on a set of pre-defined rules and if the nodes violate that it would be blacklisted. Finally, the confusion matrix results how the improved the true positive rate is and a reduced fake positive rate. A relation-based intrusion detection (RID) system was proposed by Wang et al. [30] to find a SCADA network's false data injection attacks. The major claims made in this research are of:

- The system state of RID is observed; the inconsistent situation is detected and the compromised origins were inferred.
- The intrusion detection model is used to attain real-time exposure on resource-constrained machines.

A model-based IDS was proposed by Goldenberg and Wool [31] which was extremely responsive that flags the irregularities with the help of the Modbus scheme. In the SCADA system, for finding the integrity attacks, an unsupervised anomaly-based IDS was developed by Almalawi et al. [32]. The major ideas of this research were as given below:

- SCADA system's steady and conflicting states were identified.
- From the identified states, it took out the proximity of finding a policy.

Here, based on the finest inconsistency, threshold conflicting annotations were divided from the steady annotations. A survey of several detection methods for finding the irregularities in the system was done by Ahmed et al. [33]. This survey contains the below groups: organization, numerical, information theory and groups. Besides, it also discusses the challenges in research while collecting dataset that is used for detecting system intrusions. Santhosh and Sachin [34] have detailed about multi-criteria decision-making (MCDM) technique which works based on a ranking methodology based on preferences. The work offers a decisive route or the mobile nodes to travel by combining the aforementioned MCDM technique with intuitionistic fuzzy soft set (IFSS).

3 Proposed Method

This part of the chapter discusses the detailed explanation of the proposed optimization technique for anomaly detection. The ultimate target of this research is to correctly spot the intrusions in a system with the help of extracted features using a learning methodology. WI-CS and GNA-NN are the two types of algorithms proposed in this research. In the first stage, the real-time data derived from the Mississippi University SCADA test bed is fed into the system, in which the features and the groups are arranged. After that, the features are extracted using the projected WI-CS-based optimization technique. Then, it selects the finest attributes and training data which are given as input to classification. In this phase, the HNN-based categorization algorithm is introduced to categorize the attack types. Figure 2 shows the flow graph of the research method and it consists of the below stages:

- Initialization of dataset
- Arrangement of attribute
- Initialization of cluster
- Optimization of feature and selection of attribute
- Classification.

In this analysis, the SCADA test bed dataset is fed as the input to evaluate the performance of the proposed IDS, which is formulated based on the PRC technique. A unique dataset feature contains 37 power system events that accompanied 15 different sets, which are used to fabricate the dataset. These features are further classified on an event basis like a natural event, attack event and no events. To get more in detail, these conditional scenarios are furthermore subdivided into other categories in a random manner such as three-class, multi-class datasets and binary. The integration of a huge number of sensors is available in the SCADA system. As a result, the memory intricacy of the system is amplified. Based on the response from all sensors, feature selection is executed to resolve the above-mentioned problem. This enhances the efficiency and, decreases the storage intricacy. The conditions are created, derived from every attribute, because of that one can use a vast data sources. With the aim of finding variations among parameters, GNN structure is developed. Huge attributes are used for accurate detection throughout the progression of classification.

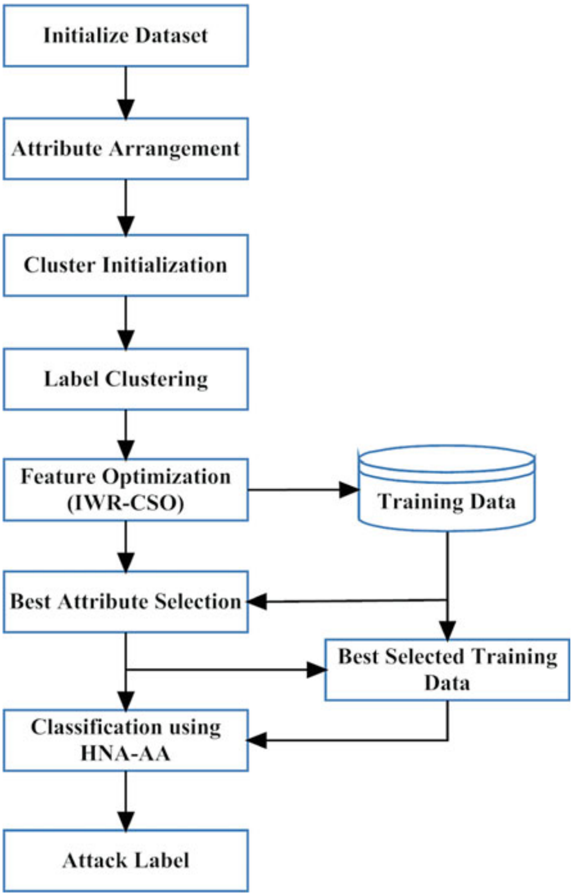


Fig. 2 Flow graph of the research method

3.1 Optimization of Feature

Here, all the extracted features are scanned for any possible intrusion of an anomaly. The main motto of our work is to select the appropriate attributes for optimization. It finds outs and senses the irregular node behavior based on preferred attributes. The developed WI-CS is an optimization-based technique which is used to cluster and organize the node attributes. Generally, fitness and a cost function comprised to form input. Optimization is the course of regulating the inputs and uniqueness of a machine to get the utmost outcome. Here, we use cuckoo search algorithm which is inspired by the brood parasitism of cuckoo birds. The cuckoos laying their eggs in the nests of other host birds. If a host bird discovers the eggs in the nests are not their own, it will throw away these alien eggs away. Else it will abandon its nest and build a new nest somewhere. Based on this observation, the cuckoo search algorithm is described by the following rules:

- Every cuckoo lays solely 1 egg at a time and the eggs are exactly set in a nest (randomly selected).
- The nest having better quality eggs which are carried onto the next phase.
- The number of nest is fixed and the quality of nest is static and is not alterable.

In this research, huge attributes are used to develop a SCADA organization. Therefore, it is necessary to spot whether the attributes are grouped or not. If the attributes are grouped, the precise results can be achieved; or else, this directs to the misclassification rate. Then, the cost function is predicted based on the difference that caused by the anomalies compared with legal nodes. Also, based on the weight, the probability of the particle is predicted. Then, the fitness value falls somewhere between anomaly and non-anomaly node ranges. These calculated values would be considered for cuckoo's egg laying probability [35]. Then, the finest attribute is selected based on previous and current values. This repeats for every iteration. The finest attribute is picked once the fitness average rate becomes higher than the calculated cost estimated value. In the proposed algorithm, input feature matrix T, cuckoo particles P, initial radius r and cost C_{st} are initialized. The cost rate is predictable by the length and radius of searching particles.

```
Algorithm I –Weighted Intrusion based Cuckoo Search (WI-CS)
Input:      T Feature matrix;
Output:     ST Select Feature;
Step 1:     Initialize cuckoo particles and cost rate;
            $P = \{T_1, T_2 \ldots T_N\}$ // Cuckoo Particles;
            Cst = 0; //Initial cost value;
            r=1; //Initial Radius;
Step 2:     Estimate cost value as,
            $Cst = \{P(1), (g \times h)\}$
            Where,
            $g = 1 + \frac{r}{(N-1)\times\sum P_i'}$        // i = 1, 2 … N;
            r – Radius of searching;
            N – Length of particles;
            $h = 1 - \sqrt{\frac{P_1}{g}}$
Step 3:     Co-ordinates of particles;
            $x = P(Fitness, 1)$;
            $y = P(Fitness, 2)$;
```

```
Step 4:    Objective function;
           $O = \begin{Bmatrix} P_i & if\ x\ \leq y \\ 0 & else \end{Bmatrix}$ // Objective function;
Step 5:    Update radius;
           $r_1 = d_1 + \frac{(V_{0\,to\,r} \times (d_2 - d_1))}{r}$
           Where,
           $d = Min\,(Cst) \pm \left(\alpha \times \left(Max(Cst)\right) - \left(Min(Cst)\right)\right)$
Step 6:    Reproduce and Update Cuckoo particles;
           $for\ i = 1\ to\ M$ // 'M' – Number of iteration
           $if\ (Cst_i < Cst_{i-1})$
                   $C_{head} = P(indx)$ // Cluster head selection;
                   Where, $indx = \begin{cases} 1, if\ \left(P \times e^{-\beta N}\right) < 0 \\ 0, \qquad\qquad else \end{cases}$
      $X_{update}(i) = x(i-1) + \left(\left(Rand^{-\frac{1}{\alpha}}\right) * cos(Cst * 2 * pi)\right)$
      $Y_{update}(n) = y(i-1) + \left(\left(Rand^{-\frac{1}{\alpha}}\right) * cos(Cst * 2 * pi)\right)$
      $Y_{update}(n) = y(i-1) + \left(\left(Rand^{-\frac{1}{\alpha}}\right) * cos(Cst * 2 * pi)\right)$
        $P(m) = \left(1 - \frac{i-1}{(M-1)^{\frac{1}{\mu}}}\right)$ //Probability of laying eggs;
        $if\ P(Cst) < P(m)$ //Mutation
          $X_{Mutation}(i) = x(i-1) + \left(P(m) \times \left(Max(x) - Min(x)\right)\right)$
          $Y_{Mutation}(n) = y(i-1) + \left(P(m) \times \left(Max(y) - Min(y)\right)\right)$
                        $Cst_{mutation} = \{P(1), (g \times h)\}$
        End if;
           Update radius r;
      End i loop;
Step 7:  $\boldsymbol{ST = T\left(Cst > Average(Cst)\right)}$
```

3.2 Classification Using HNN

By using the projected HNN method, the attacking and non-attacking labels are properly categorized only next to optimize the features. In the proposed method, the inputs given are the label Lb, select training set Str and testing feature Sv. At the start, the preferred testing feature, φ, directionality D, and iteration(theta) are initialized as zero. The unsystematic values like Wxy and Wxh are initialized with the range of testing Sv and also training Str. The NN includes three dissimilar kinds of layers contains hidden, input and output layers. Normally, the final result can be obtained from the output layer. Here, net 1 and net 2 indicated the input–output layer, respectively. Then, the exponential is calculated for the layers *H1* and Y. Then, the temporary distance is calculated. The temporary distance s_i is obtained only after getting the calculated load value of φ that is upgraded by including the present value. After that, for the differential features, the directionality *D* is estimated, *D* and theta values are upgraded. The association *r* among the attribute sets are estimated, in case, it is larger than the *D* value, Lb(i) label is allocated as the classified label (CL).

Algorithm II – Hierarchical Neural Network (HNN)

```
Input: Str as choosen training set; Sv as choosen testing feature and Lb as Label;
Output: Classified Label CL;
Step 1: Initialization,
        $SV = \frac{SV}{norm\ (SV)}$;
        $Theta = 0, D = 0, \varphi = 0$;
Step 2: $for\ (i = 1\ to\ Row_{size}\ (Str))$
Step 3:         $Wxh = Rand(STr), Why = Rand(SV)$;
Step 4:         $net1 = SV \times Wxh - Theta$;
Step 5:         $H1 = \frac{e^{(net1)} - e^{(-net1)}}{e^{(net1)} + e^{(-net1)}}$;
Step 6:         $net2 = H1 \times Why - Theta$;
Step 7:         $Y = \frac{e^{(net2)} - e^{(-net2)}}{e^{(net2)} + e^{(net2)}}$;
Step 8:         $s_i = \sqrt{\left(\frac{\sum((Str(n,:) - Y)^2)}{Y}\right)}$;
Step 9:         $\Delta Y = (1 + Y) \times (1 - Y) \times (STr_i - Y)$
Step 10:        $\varphi = \varphi + s_i + Theta$;
Step 11:    $D = (\varphi * \Delta Y + s_i) * \varphi$; // Directionality for differential features;

Step 12:        $Theta = Theta * D$;
Step 13:        $\varphi = \varphi + Theta$;
Step 14:    $r = \frac{\sum_{j=1}^{N}(Y_i - \bar{Y}) \times (s_i - \bar{s})}{\sqrt{\sum_{j=1}^{N}(Y_i - \bar{Y})^2 \times \sum_{j=1}^{N}(s_i - \bar{s})^2}}$;// Correlation between the feature sets;
Step 15:    If $(r > min(D))$ // Feature verification condition;
Step 16:        $CL = Lb\ (i)$; // Classified label;
Step 17:    End if;
Step 18:  end i loop;
```

Table 1 Parameter dataset with and without intrusion

Factor		Dataset			
		OLIRD	OLORD	MLIRD	MLORD
Humidity factor	Without intrusion	44–52	32–65	42–55	44–75
	With intrusion	45–90	50–87	50–92	57–94
Temperature factor	Without intrusion	25–30	24–36	28–32	25–35
	With intrusion	25–55	26–38	25–55	28–50

4 Performance Investigation

This part of the chapter gives the outcome using detection rate and false alarm rate for both proposed as well as existing algorithms. Here, in the presented research, with the help of the Ns-2 simulator, one dataset is formed by designing the system organization with a hundred nodes. After that, MATLAB tool is used to detect the intrusion by the implemented dataset. For validating the proposed method, two scenarios were taken, i.e., with and also without attacks. The compared datasets [13] in this section are as follows, one-leap indoor real data (OLIRD), multi-leap indoor real data (MLIRD), one-leap outdoor real data (OLORD) and multi-leap outdoor real data (MLORD).

4.1 Description of Dataset

Table 1 illustrates the dataset formed and utilized in this proposed work. Temperature and humidity are some of the parameters considered for analysis. Those above-mentioned factors point out the significant states from various densities in two-dimensional (2D) space. It shows the various levels of discriminations which is near to the low, medium and large-risk conditions. Besides, considering without and with attack, the factors are evaluated. Now, the lowest and largest values of temperature and humidity factors are calculated. For example, 44–52 in OLIRD dataset, 44 is the lowest and 52 is the largest humidity level. The dataset including low and high values are tabulated below:

4.2 Metrics for Accuracy

In the proposed method by using detection and false positive rate, the accuracy metrics is measured for intrusion detection. Efficient data-driven clustering (EDDC) [36] and proposed WI-CS with GNN techniques the detection rate and fake positive rate are given in Table 2.

Table 2 Detection rate and false positive rate

Dataset	Detection rate (%)		False positive (%)	
	EDDC	WI-CS with GNN	EDDC	WI-CS with GNN
OLIRD	100	100	0	0
OLORD	96.88	97.57	1.94	1.14
MLIRD	100	100	0.31	0.18
MLORD	92.98	94.28	0.04	0

$$\text{Detection Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

where

TP True positive
TN True negative
FP Fake positive
FN Fake negative

The fake positive rate is calculated as follows:

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

where TP specifies the count of perfectly identified critical states, similarly FN specifies the count of the present but not identified critical states, FP specifies the wrongly flagged normal states as critical, and furthermore, TN specifies the count of properly identified normal states. Hence, it is observed from the analysis when compared to the other methods, the developed method gives the finest results for the datasets of four different cases.

4.3 Accuracy

$$\text{Accuracy} = \frac{\text{Number of critical states}}{\text{Total number of states}} \tag{3}$$

For different datasets, the accuracy values are shown in Table 3 for proposed as well as existing methods. For all datasets, high accuracy is obtained for the proposed method when compared to the existing method.

Table 3 Accuracy

Dataset	Accuracy (%)	
	Existing	Proposed
OLIRD	100	100
OLORD	97	98
MLIRD	100	100
MLORD	93	95

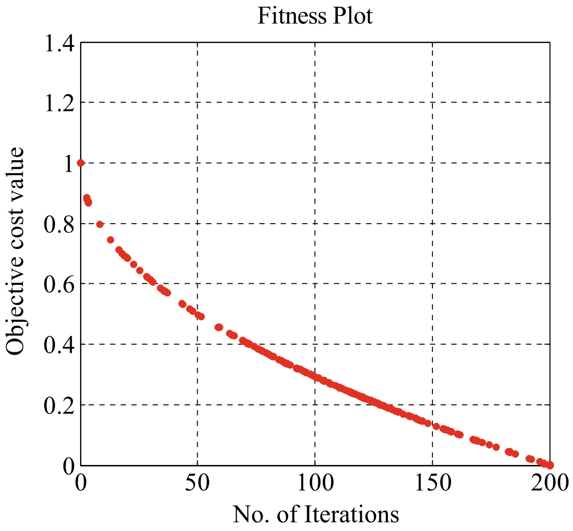


Fig. 3 Number of iterations versus objective cost value

4.4 Fitness Plot

Concerning the number of iterations, Fig. 3 shows the objective cost value. An increase in the numbers of iterations (0–200) will decrease the cost value as in progress. In the proposed method, the cost is reduced for intrusion detection which can be concluded from the above analysis.

4.5 Performance Rate

The incorporated WI-CS with SVM does not give the improved performance results, because of the restrictions of SVM but the proposed WI-CS is a competent optimization method. The major disadvantage of SVM is that it digs out an additional quantity of irrelevant features. Lacking of WI-CS, the time complexity is increased because of the whole data is necessary for analysis. During the normal attack detection and abnormal attack detection, it may guide to high misclassification rate. To overcome

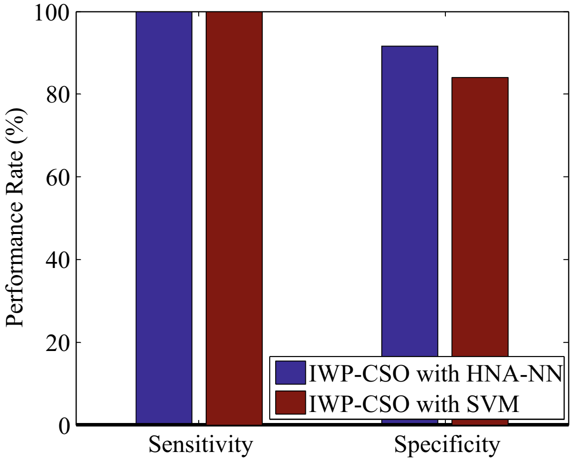


Fig. 4 Sensitivity and specificity

all these, WI-CS is created to obtain the optimal features. Similarly, to rise above the difficulty of SVM, the GNN method is collaborated with the WI-CS method. In general, to calculate the efficiency of the system, two major values are considered. One is sensitivity and the other one is specificity. It can be calculated as given below:

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{\text{Number of true positive assessments}}{\text{Number of all positive assessments}} \tag{4}$$

Likewise, specificity is calculated as given below:

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} = \frac{\text{Number of true negative assessment}}{\text{Number of all negative assessment}} \tag{5}$$

These values of IWP-CSO/WI-CS with SVM and proposed IWP-CSO/WI-CS with GNN techniques are shown in Fig. 4. When compared to other methods, the proposed method gives a high-performance rate which is obtained from the above analysis.

4.6 Jaccard and Dice Coefficients

The similarity between data can be measured mainly by using the Jaccard and dice coefficients. It is given as the fraction between the sizes of the intersection of a couple of set to the union of two sets. These coefficients are used in the paper [37, 38] for comparison of different algorithms like random forest, *K*-means, etc. The metrics can be calculated as given below:

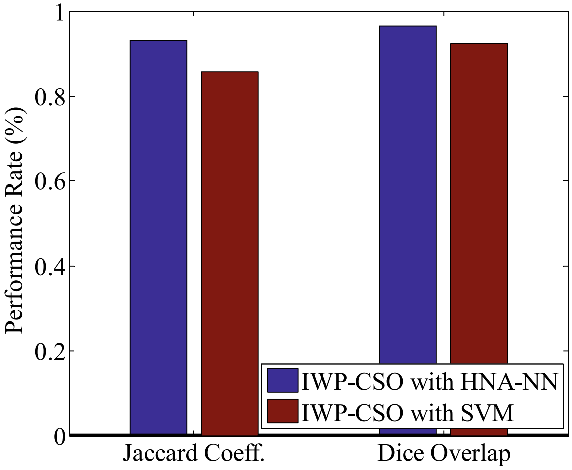


Fig. 5 Jaccard and dice coefficients

$$\text{Jaccard} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{6}$$

where X and Y specify a couple of datasets. Likewise, dice is furthermore a similarity measuring method that discovers the similarity between information. It can be calculated as given below:

$$\text{Dice} = 2\frac{|X \cap Y|}{|X| + |Y|} \tag{7}$$

Jaccard and dice of WI-CS with SVM and proposed IWP-CSO / WI-CS with HNN method are shown in Fig. 5. The high-performance results are obtained in the proposed method when compared to the existing method.

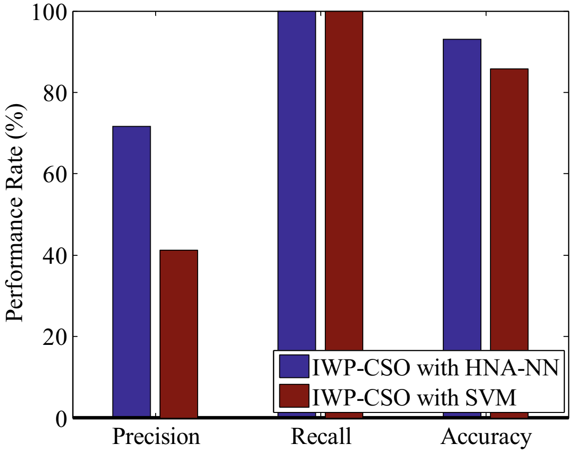
4.7 Precision, Recall and Accuracy

Figure 6 shows the performance rate of IWP-CSO with SVM and proposed IWP-CSO/WI-CS with HNN methods. These three rates are highly considered or the performance criteria. Precision and recall are calculated mainly based on true positive detection rate, whereas accuracy majorly depends on sensitivity and specificity values as mentioned below:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

Fig. 6 Performance rate



$$\text{Accuracy} = \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{TP} + \text{FN} + \text{FP})} = \frac{\text{Number of true correct assessment}}{\text{Number of all assessment}} \tag{10}$$

It is very clear that the proposed method gives a very high precision, recall, as well as accuracy measures than the existing methods.

4.8 False Detection Rate

It is very important to get rid of genuine nodes which are misclassified as anomalies. To find out that we calculate false detection rate (FDR) as given below:

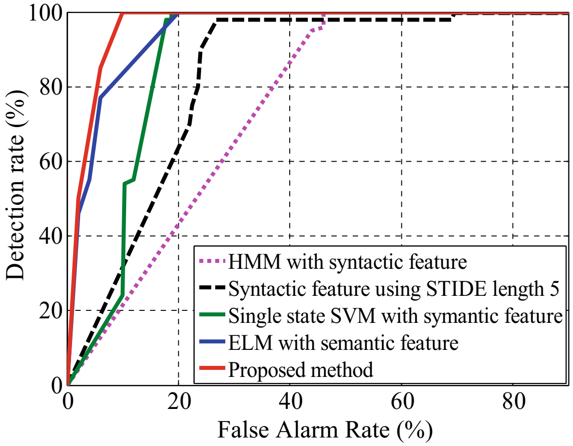
$$\text{False Detection Rate} = \frac{\text{number of honest users misidentified}}{\text{number of nodes identified as attackers}} \tag{11}$$

The false detection rate of an existing method as well as the proposed method is shown in Fig. 7, in which the x-axis corresponds to the alarming rate (in %) and the y-axis corresponds to the detection rate (in %). The proposed method gives a very high detection rate than the existing methods which is obtained from the analysis.

5 Conclusion and Future Work

Nowadays, even SCADA and ICS systems are more likely to be using a wireless sensor network. Even in the medical field, wireless sensor system became so inevitable nowadays. This study [39] deals with a detailed review of how ECG monitoring model works and about the evaluation of zig bee medical sensor networks in

Fig. 7 Detection rate



clinical networks. This shows really how much impact WSN has made in health care sectors. Hence, this chapter's security proposal is not just enough for the traditional SCADA network but also for SCADA that is using WSN. This chapter presents an improved WI-CS and GNN techniques for a SCADA system to filter the anomalies in it. The major motto of this work is to find out the outliers in this network and to classify the unknown attacks in the SCADA system. In this, the real-time dataset of the SCADA network is fed as input, and appropriate parameters are chosen. Then, in the training dataset, the attributes are selected which are optimized by employing the proposed WI-CS technique. After that, the filtering the finest attributes they are optimized by the GNN classification algorithm that guesses the intruder and also non-intruder label. The investigational results estimate the performance of the proposed method considering sensitivity, false detection rate, precision, recall, Jaccard, accuracy, dice and specificity. Furthermore, SIRD, SORD, MIRD and MORD which are the four different datasets are considered in this research for showing the improved performance of the introduced system. From this study, when compared to the other algorithms, the developed WI-CS through HNN algorithm gives better results.

References

1. Aghajanzadeh N, Keshavarz-Haddad A (2015) A concise model to evaluate security of SCADA systems based on security standards. Int J Comput Appl 111
2. Shahzad A et al (2015) A SCADA intermediate simulation platform to enhance the system security. In: 2015 17th international conference on advanced communication technology (ICACT), pp 368–373
3. Debashis De et al (2020) Wireless sensor network: applications, challenges, and algorithms, Springer tracts in nature-inspired. Springer, pp 1–18
4. Wei H et al (2015) SOM-based intrusion detection for SCADA systems. In: Electronics and electrical engineering: proceedings of the 2014 Asia-Pacific electronics and electrical engineering conference (EEEC 2014), 27–28 Dec 2014, Shanghai, China, p 57

5. Mcquillan JL, Lloyd CA (2016) SCADA intrusion detection systems. ed: US Patent 20,160,094,578
6. Liao H-J et al (2013) Intrusion detection system: a comprehensive review. J Netw Comput Appl 36:16–24
7. Das SK, Tripathi S (2019) A nonlinear strategy management approach in software-defined ad hoc network. In: Lecture notes in networks and system, pp 321–346
8. Manikandan ST (2014) Removal of selective black hole attack in MANET by AODV protocol. Int J Innov Res Sci Eng Technol 3(3):2372–2377
9. Shitharth S, Winston DP (2015) An appraisal on security challenges and countermeasures in smart grid. Int J Appl Eng Res 10(20):16591–16597
10. Ou C-M (2012) Host-based intrusion detection systems adapted from agent-based artificial immune systems. Neurocomputing 88:78–86
11. Shitharth S, Winston DP (2015) A comparative analysis between two countermeasure techniques to detect DDoS with sniffers in a SCADA network. Procedia Technol 21:179–186
12. Koc L et al (2012) A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. Expert Syst Appl 39:13492–13500
13. Selvarajan S, Shaik M, Ameerjohn S, Kannan S (2019) Mining of intrusion attack in SCADA network using clustering and genetically seeded flora based optimal classification algorithm. Inf Secur IET 14(1):1–11
14. Almalawi A et al (2016) An efficient data-driven clustering technique to detect attacks in SCADA systems. IEEE Trans Inf Foren Secur 11:893–906
15. Mitchell R, Chen R (2016) Modeling and analysis of attacks and counter defense mechanisms for cyber physical systems. IEEE Trans Reliab 65:350–358
16. Ponomarev S, Atkison T (2016) Industrial control system network intrusion detection by telemetry analysis. IEEE Trans Dependable Secur Comput 13:252–260
17. Lin H et al (2016) Runtime semantic security analysis to detect and mitigate control-related attacks in power grids. IEEE Trans Smart Grid 1–1
18. Liu AX, Torng E (2016) Overlay automata and algorithms for fast and scalable regular expression matching. IEEE/ACM Trans Netw 1–16
19. Marchang N et al (2016) A Novel approach for efficient usage of intrusion detection system in mobile Ad Hoc networks. IEEE Trans Veh Technol 1–1
20. Li B et al (2016) DDOA: a dirichlet-based detection scheme for opportunistic attacks in smart grid cyber-physical system. IEEE Trans Inf Foren Secur 1–1
21. Ambusaidi M et al (2016) Building an intrusion detection system using a filter-based feature selection algorithm. IEEE Trans Comput 1–13
22. Hasan MM, Mouftah HT (2016) Optimal trust system placement in smart grid SCADA networks. IEEE Access 4:2907–2919
23. Yang Y et al (2014) Multiattribute SCADA-specific intrusion detection system for power networks. IEEE Trans Power Delivery 29:1092–1102
24. Samdarshi R et al (2015) A triple layer intrusion detection system for SCADA security of electric utility. In: 2015 annual IEEE India conference (INDICON), pp 1–5
25. Sayegh N et al (2014) SCADA intrusion detection system based on temporal behavior of frequent patterns. In: MELECON 2014–2014 17th IEEE mediterranean electrotechnical conference, pp 432–438
26. Amin S et al (2013) Cyber security of water SCADA systems—part II: attack detection using enhanced hydrodynamic models. IEEE Trans Control Syst Technol 21:1679–1693
27. Maglaras LA et al (2014) Integrated OCSVM mechanism for intrusion detection in SCADA systems. Electron Lett 50:1935–1936
28. Yang Y et al (2013) Intrusion detection system for IEC 60870–5–104 based SCADA networks. In: 2013 IEEE Power Energy Soc Gener Meet: 1–5
29. Yasakethu S, Jiang J (2013) Intrusion detection via machine learning for SCADA system protection. In: Proceedings of the 1st international symposium on ICS and SCADA cyber security research 2013, pp 101–105

30. Wang Y et al (2014) SRID: state relation based intrusion detection for false data injection attacks in SCADA. In: European symposium on research in computer security, pp 401–418
31. Goldenberg N, Wool A (2013) Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems. Int J Crit Infrastruct Prot 6:63–75
32. Almalawi A et al (2014) An unsupervised anomaly-based detection approach for integrity attacks on SCADA systems. Comput Secur 46:94–110
33. Ahmed M et al (2016) A survey of network anomaly detection techniques. J Netw Comput Appl 60:19–31
34. Das SK, Tripathi S (2018) Intelligent energy-aware efficient routing for MANET. Wirel Netw 24(4):1139–1159
35. De D, Mukherjee A, Das SK, Dey N (2020) Nature-inspired computing for wireless sensor networks, Springer tracts in nature-inspired. Springer, pp 1–341
36. Suthaharan S et al (2010) Labelled data collection for anomaly detection in wireless sensor networks. In: Intelligent sensors, sensor networks and information processing (ISSNIP), 2010 6th international conference on, pp 269–274
37. Selvarajan S, Shaik M, Ameerjohn S, Kannan S (2019) Integrated probability relevancy classification (IPRC) for IDS in SCADA', design framework for wireless network. Lect Notes Netw Syst 82(1):41–64
38. Das SK, Samanta S, Dey N, Kumar R (2019) Design frameworks for wireless networks. Lecture notes in networks and system. Springer, Singapore, pp 1–439
39. Dey N et al (2017) Developing residential wireless sensor networks for ECG healthcare monitoring. IEEE Trans Consum Electron 63(4):442–449