IET Generation, Transmission & Distribution

Special Issue: Advanced Data-Analytics for Power System Operation, Control and Enhanced Situational Awareness

IET Journals
The Institution of Engineering and Technology

Deep learning based method for false data injection attack detection in AC smart islands

ISSN 1751-8687
Received on 1st March 2020
Revised 23rd April 2020
Accepted on 4th May 2020
E-First on 1st July 2020
doi: 10.1049/iet-gtd.2020.0391
www.ietdl.org

Moslem Dehghani[1], Abdollah Kavousi-Fard[1] ✉, Morteza Dabbaghjamanesh[2], Omid Avatefipour[3]

[1]*Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran*
[2]*Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX, USA*
[3]*Department of Electrical and Computer Engineering, University of Michigan-Dearborn, USA*
✉ *E-mail: kavousi@sutech.ac.ir*

Abstract: This paper investigates the false data injection attacks (FDIA) in an AC smart island and the detection solution of the attack on distributed energy resources in a smart island. In this study, a new scheme of FDIA detection is proposed based on wavelet singular values as input index of deep learning algorithm. In the proposed method, switching surface based on sliding mode control breaks down for adjusting accurate factors of wavelet transform and then features of wavelet coefficients are extracted by singular value decomposition. Indexes are determined according to the wavelet singular values in switching surface of voltage and current which defines the input indexes of deep machine learning and detecting FDIA. This cyber-protection plan has been put forward for cyber diagnostic and examined in different types of attacks happening in voltage and current signals derivation of measuring sensors as well as sending and receiving data from communication and control systems. The main priority of the suggested detection plan is the high capability to detect FDIA with a high accuracy. To show the effectiveness of the proposed method, simulation studies are performed on AC smart island in MATLAB/Simulink environment.

1 Introduction

In situations where the main grid and electricity utilisation are far apart, like distant islands and segregated communication stop, it is not economically efficient or is practically tough to supply electricity by transmission lines. In such situations, the advisable way is to provide power in the islanding mode of a microgrid (MG) incorporating renewable energy resources such as wind turbines, photovoltaic, and fuel cells. In such a smart island (SI), the MG distributed generation units are in the charge of voltage, frequency, and current control, fault protection and cyber-attack detection [1].

SI is an efficient way of merging renewable energy resources, storage devices, and new electronic loads that can work apart from the utility grid in an islanding. In addition, the nature of operating units at AC paradigm resulted in a brilliant way to boost the performance [2]. To enhance the robustness and scalability, distributed controllers are recommended in micro-grids for avoiding the only failure point compared with the centralised communication, due to their extremely dependable process when link crashes [2]. Furthermore, distributed control philosophy is an economical option as it is easily able to accommodate through lower volume of data transformation with not involving considerable traffic countering the intensive communication [3]. In MGs, helpful secondary controllers are used for different purposes like average voltage adjustment and proportionate load sharing [1].

SI security involves two main aspects: physical security and cyber security. Physical security represents the ability of a SI to maintain a normal working state in the presence of severe disturbances. Cyber security refers to the security of the communication networks and computer systems which support the SI operation. In recent years, cyber security has become a significant threat to smart city and SI system due to the pervasive application of information technologies. Moreover, weaknesses in cyber security can also threaten the physical security of the SI due to the deep integration of the physical and cyber systems [4].

Cyber-attacks are capable of undermining or even totally disrupting the control systems underlying electric power grids. It was traditionally believed that cyber-attacks were incapable of threatening the security of industrial systems. However, cyber-attacks have resulted in many security challenges in recent years and have become a critical concern for both industrial control system users and vendors [4].

Conventional MG operation is running via a central supervisory controller and data acquisition that is appropriate for high-level operations containing worldwide optimisation and agent obligation [5]. This controller suggests operative coalition of important sub-systems which is necessary to have secure performance [6]. Focus of control functions, presuming complete accessibility to information of system, is considered as advanced objectives at islanding MGs (IMGs) and isolated power systems that do not have accessibility to outside grids. Nevertheless, this concentrated plan is in danger of alone-spot devastations in physical and cyber-attacks that intrude by the IMG's process [4]. Data transferring using the applied communication networks are in peril to attack due to the lack of firewalls and absence of good encryption keys in communication protocols which have not been updated to the latest version to countermeasure the exposure of cyber-attacks [7]. New MGs are applied as commercial off-the-shelf calculating platforms which reportedly permeated through attackers lately [8]. Cyber menaces on MGs demand immediate attention as a result of the maritime systems nature, lots of them are far from the land and includes long-range communication [9]. Cyber-attacks are able to damage the system through growing the operating costs, interference with acute loads, and creating total system collapse.

Generally, individual sensors at wide extend networks are the essential aim of security compromises. The compromised agent is able to simply reach to the data in a compromised node. Although authentication approaches stand upon cryptography or security gateway design, like that was explained in [10], they are impenetrable owing to the calculation and storage restrictions of the system. Current studies in the smart grid background mostly emphasis on the network security of the cyber elements [11], advanced misfit diagnostic mechanism [12], and secure control theories stand upon techniques of different state estimation [13].

Several research studies have been conducted on suggesting different false data injection attacks (FDIAs) approaches and expanding similar detection mechanism. Abdollah *et al.* [14] investigated the attack strategy on the basis of Gaussian process and this strategy was applied based on machine learning for detecting the attack. Liu *et al.* [15] developed an FDI detection

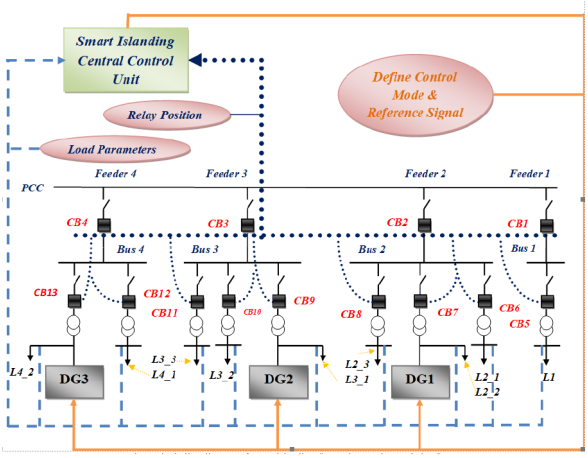

Fig. 1 *Single line diagram of smart islanding*

mechanism based on using attributes of the low measurement dimensions and lack of attack. In [16], authors integrated the biggest weighted remaining procedure and tantamount measurement conversion with the aim of finding the FDIAs.

Only a few studies have been published on AC-based FDIA, because of recognising complication of non-linear systems [17]. A graph-based algorithm is introduced in study [18] with the aim of identifying a series of endangered sensors that is enough to create an uncontrollable attack. However, this method was presented only for a single attack and different types of attacks were not considered. Rahman and Mohsenian-Rad [17] studied AC-based FDIA on the basis of linearisation around the target state default that SE is attained through a special algorithm, which can be very sensitive in operation. Soltan *et al.* [19] studied physical attacks and joint cyber on power grids, which has been developed to a simplified model of AC power flow and to the fault data injection ways.

In our suggested real-time detection method, deep learning algorithm is employed to detect the manner schema of the FDIAs by applying the historical measurement data and we use detector features to recognise the attacks of the FDI in real time.

Deep learning methods are recently suggested to catch the higher order statistical structure of the intricate data through ordering the detector features in layers. One of the essential deep learning techniques which is widely used is deep belief network which is created with a group of restricted Boltzmann machines (RBMs) [20].

Choosing appropriate variables with the aim of implementing the neural networks in an efficient way is vital. The control of the MGs is usually carried out using voltage and current measurements and potential attackers consider them as direct targets. Through injecting fault data to the voltage and current measurements, the attacked variable will be launched with the aim of controlling application and also it can disarrange MG application control.

Moreover, if both voltage and current measurements are controlled in a separate way, in case of system attack, the suggested deep learning anomaly detection strategy is able to diagnose different attacks on current measurements and voltage measurements these attacked variables can to be recognised directly with no expansion of detection strategy. The proposed method is able to inhibit more complication of the FDIAs detection where suggested method is considered for each agent. It is assumed that when both load alternation and cyber-attack do not occur at the system, SI acts as normal. In this case study, strategy of detection is able to diagnose among a cyber-attack and load variation. Also, the attack can be exactly identified with this method as part of a cyber-attack. At the end, the suggested FDIA detection in a SI is an offline digital time-domain manner and simulated in MATLAB with the purpose of proving the efficiency, productivity, precision, and validity of the proposed method.

This paper is organised as follows. Section 2 introduces SI model. Section 3 illustrates the state estimation and false data injection method, respectively. Section 4 defines wavelet transform (WT), singular value decomposition (SVD), deep learning, and suggested FDIA detection scheme. Section 5 presents case study and investigates proposed FDI detection mechanism under various case studies. Finally, the paper is concluded in Section 6.

2 Smart island model

2.1 Smart island

SI is an island where information and communication technology (ICT) is used to increase operational efficiency, share information with everyone as well as improving service delivery.

The exact definition of a SI raises depending on the target community. The main goal of SI is to increase the efficiency of its smart operations, economic growth as well as attracting more and more tourists. On a SI, it is trying to increase the level of well-being on these islands by using data analytics as well as new technologies. The success of SI depends on a strong relationship between DGs on the island. SI uses a combination of internet of thing (IOT) technologies, software solutions, user interface and communication networks. A SI is more depended on IOT equipment than anything else. IOT is a network of interconnected devices that communicate with each other and exchange data. The data collected is stored on a server or cloud in an IOT network. A SI must have a smart energy to attract more tourists and increase the level of well-being and spend times in the best way. So, smart energy is one of the major challenges of the SI.

Saving energy and increasing productivity are the main goals of SI. To achieve these goals, various technologies such as smart sensors, smart lights, smart grids are used. Smart grids are used to monitor energy consumption in different locations, balance and supply and save the energy.

What is being studied in this paper is investigating FDI attacks and detecting the attacks on the current and voltage sensors and controller in a SI where data is sent and received to the generation units and central control unit [21].

Fig. 1 illustrates block diagram of SI with central control unit that consists of DGs, various loads, relays, sensors, and central control units. Whole part of the SI, which is connected to the central control unit and measurement data, sends and/or receives information via fibre cable or wireless which can be an attack surface.

3 State estimation and FDIA

3.1 State estimation

State estimation is considered as one of the essential parts of central control unit in the SI that is able to calculate the operational state in each bus from different meter measurements. The outcome of state estimation is applied in higher layer applications including: contingency selection, security assessment and security restriction economic dispatch, and so on. The AC on the basis of state estimation model is shown as below:

$$x = h(x) + e \tag{1}$$

Where the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ defines the real time measurement data that consists of power injection and power flow measurements at buses and transmission lines, respectively; the vector $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)^{\mathrm{T}}$ defines the system state that consists of the phase angle and voltages at buses; the vector $\boldsymbol{e} = (e_1, e_2, \ldots, e_n)^{\mathrm{T}}$ defines the measurements noise, which is supposed to be Gaussian distribution [22]; $n > m$; $h(x)$ defines the functional dependency among state variables and measurements. The topology structure which is derived from the switch/breaker equipment, determines the accurate type of $h(z)$.

The weighted least squares strategy solves the model (1). In this way, solving the below optimisation problem can help to obtain estimated state variables $\hat{\boldsymbol{z}}$ vector:

$$\min \quad J(z) = \frac{1}{2}(\boldsymbol{x} - h(z))^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{x} - h(z)) \tag{2}$$

where $\boldsymbol{W}$ defines a diagonal matrix expressed as $= \mathrm{diag}(\delta_i^{-2}, 0)$, where δ_i^{-2} defines the measurement errors variance related with the ith meter ($1 \le i \le m$).

By solving (2) (applied an iterative algorithm), the estimated state variable is computed. The validation of $\hat{z}$ is defined through bad data detection that applies largest normalised residual test and $J(\hat{z})$ defines the performance index [22]

$$J(\hat{z}) < C \tag{3}$$

In which $J(\hat{z})$ is considered to track a chi-square distribution, with threshold C set to several predesignated importance level. The estimated state variable $\hat{x}$ can be applied in other application, just when model (3) is assumed. Since the primary sources, which are provided by switch/breaker and meters equipment and processed by the state, are from both digital and analogue measurements, any destructive behaviour in contrast breakers/switch and meters equipment may result in security problem in the power grids. The cyber topology attacks and FDIA which are mentioned below can be considered as a special attack that are able to impress the state estimation outcomes through manipulating measurements of the switch/breaker and meters equipment.

3.2 False data injection attacks

The FDIA, which is first suggested by Liu *et al.* [22], can be considered as a cyber-attack in which state estimation outcomes are destroyed through injecting false data into meter measurements in a precise and harmonious way. A successful FDIA warrants that the state estimation residual drops under the hypothesis test threshold, despite the existence of malicious injection data.

In summary, the secret attack manipulated the state estimation input as $x_{\mathrm{bad}} = \boldsymbol{x} + \boldsymbol{a}$, where $\boldsymbol{a}$ is considered as the vector of malicious injection datum. So, model (1) can be formulated as below:

$$x_{\mathrm{bad}} = h(z) + \boldsymbol{a} \tag{4}$$

The estimated state variable $\hat{\boldsymbol{z}}$ has a deviation according to the iterative algorithm, stated as $\hat{z}_{\mathrm{bad}} = \hat{\boldsymbol{z}} + c$. While the false estimated state variable $\hat{z}_{\mathrm{bad}}$ crosses the bad data detection, the attack has been successfully launched, assuring

$$J(\hat{z}_{\mathrm{bad}}) < C \tag{5}$$

Model (4) should be solved by attacker which assures test (5) to initiate this attack. An accurate vector of $\boldsymbol{a}$ is a key to this problem and several ways are able to gain this.

An attack is assured test (5) in this condition, if the attacker is able to assure $(\hat{z}_{\mathrm{bad}}) = J(\hat{z})$, this attack is named FDIA; otherwise, it is named a generalised FDIA. The first one is on the basis of DC layout, whereas the second one is mostly on the basis of AC layout. In this study, an AC layout for the state estimation is considered with greater relation to applied grid function.

4 Detection mechanism of FDIA

In this part, the FDIA detection mechanism applied the advantages of WT and SVD to extract the detailed components to use as an input index of deep learning. At first, the structure of the suggested method is defined. Finally, the performance of the method is stated with short statements to the technique engaged.

4.1 Wavelet singular values

4.1.1 Wavelet transform: Usual 1D decomposition consists of time-domain or frequency-domain resolution which is generally not able to obtain the attack pattern and it is hard to get the composed cyber-attack detection of the AC SI merely based on the data provided through frequency or time domain. Time frequency is depicting as an effective method in decomposing sensor signals to detect the attack by providing a vision into the main data in the time context. Different methods, such as S-transform (ST), short-time Fourier transform (STFT), and WT, are able to figure out time-frequency imaging. Nevertheless, due to the fixed frequency resolution of STFT and the frequency dimness for wide frequency band of ST, WT is adopted to transform 1D fluctuation signals processing in this study. WT is a useful way to extract time-frequency in detail, due to lower time decomposition and higher frequency decomposition in low frequency section [23]. WT is named a microscope of signal decompositions. It is able to demonstrate the information of low frequency in extensive scale and locally ascertain the feature of high frequency in small scale. WT is perceived through computing the internal yield of the condition resolution signal $z(t)$ and the wavelet foundation function $\theta_{a,\tau}(t)$. Therefore, WT is specified as follows [24, 25]:

$$WT_z(\alpha, \tau) = \int_{-\infty}^{+\infty} z(t)\theta_{\alpha,\tau}(t)\mathrm{d}t = \int_{-\infty}^{+\infty} z(t)\theta\left(\frac{t-\tau}{\alpha}\right)\mathrm{d}t \tag{6}$$

where α and τ are the dilation factors (it is implemented fundamental wavelet stretching) and the translation factor (it is reflected wavelet function displacement) in transformation, respectively.

4.1.2 Singular value decomposition: It can be assumed that the original discrete signals $X = [x(1), x(2), \ldots, x(M)$ are gathered. The Hankel matrix is able to construct on the basis of the phase space reconstruction opinion as follows [25]:

$$\boldsymbol{Y} = \begin{bmatrix} x(1) & x(2) & \ldots & x(m) \\ x(2) & x(3) & \ldots & x(m+1) \\ \ldots & \ldots & \ldots & \ldots \\ x(M-m+1) & x(M-M+2) & \ldots & x(M) \end{bmatrix} \tag{7}$$

where $1 < m < \boldsymbol{M}$, let $n = \boldsymbol{M} - m + 1$, then $\boldsymbol{Y} \epsilon R^{n*m}$. The attacker orbit matrix is reconstructed by this matrix. The matrix $\boldsymbol{M}$ exposes the dynamic features of the attacker in the reconstruction space by reconstructing the specifications of the attractor. So, $\boldsymbol{Y}$ is able to express as $\boldsymbol{Y} = \boldsymbol{D} + \boldsymbol{W}$, that $\boldsymbol{D}$ indicates the $(M - m + 1)*m$ matrix of the smooth signal in the reconstruction space and $\boldsymbol{W}$ indicates the $(M - m + 1)*m$ matrix of the noise interposition signal.

The SVD is applied to the mentioned matrix $\boldsymbol{Y}$, the following relational equation is obtained:

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{8}$$

In (8), $\boldsymbol{U}$ and $\boldsymbol{V}^{\mathrm{T}}$ are $(\boldsymbol{M} - m + 1)*$ $(\boldsymbol{M} - m + 1)$ and $m*m$ matrices, respectively, $\boldsymbol{S}$ is a diagonal matrix of $(M - m + 1)*m$, the main diagonal elements are $\delta_i (i = 1, 2, \ldots, j)$ and $j = \min((M - m + 1), m)$, namely:

$$\boldsymbol{S} = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_k) \tag{9}$$

In (9), $\delta_1, \delta_2, \ldots, \delta_k$ are the singular values of matrix $\boldsymbol{Y}$, and $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_k \geq 0$ is satisfied, $\boldsymbol{V}^{\mathrm{T}}$ and $\boldsymbol{U}$ define the right and left singular matrix.

4.1.3 Deep learning algorithm: Deep learning algorithm is based on a neural network which consists of several hidden layers among the input and output layers. It is able to model intricate non-linear relevance between various kinds of variables. These network parameters are achieved through unsupervised learning for input data layer by layer, and then supervised learning is applied for fine-tuning. Deep learning patterns conduct to have more intricate details at higher output layers, and the learned intricate details will be invariable by alternating of input [26]. In this study, the deep auto-encoder (DAE) learning type is applied to exploit the relevance and details from the SI's voltage and current that are measured by sensors, send/receive to each other and controllers. For a DAE network, its training process includes two phases, pre-training and fine-training, that are used to gain the DAE network layout parameters.

4.2 Pre-training of the DAE

The DAE is a deep learning network that consists of diverse RBM stacks. In the DAE, each RBM output is assumed as a new input of a higher level RBM to gain the transmission of learning outcomes layer by layer. Each hidden layer is repeated several times to initialise the parameters.

The workmanship of the DAE network consists of two operations, encoding and decoding. In the encoding function, at first, the input X is converted to construct a series of details for subsequent layer-wise conversions, and the intricate details are gained in upper layers. Eventually, the code Y is obtained by the encoding function.

Similarly, the code Y is converted back to the main input iteratively through RBMs and the renovation of X, $\hat{X}$ is produced in the decoding function. The encoding and decoding mechanisms are shown in Fig. 2.

As it can be seen in Fig. 3, the RBM contains two-layer network of a random Markov kind that has N visible modules $v_i = \{0,1\}^N$ and M hidden modules $h_j = \{0,1\}^M$. The energy pattern is presented to the related energy of the common structure modules in the RBM, though

$$E(v,h) = \theta - \sum_{i=1}^{N} a_i v_i - \sum_{j=1}^{M} h_j b_j - \sum_{i=1}^{N}\sum_{j=1}^{M} w_{ij} v_i h_j \tag{10}$$

That $\theta = \{w_{ij} v_i h_j\}$ displays the weight among visible module i and w_{ij} displays the hidden module j. b_i displays the bias of visible module and a_j displays the hidden module.

The common distribution among modules on the basis of the energy pattern in the RBM is defined as follows:

$$\begin{aligned} P(v,h);\theta &= \frac{1}{z(\theta)}\exp(-E(v,h);\theta) \\ ;z(\theta) &= \sum_{v}\sum_{h} E(v,h);\theta \end{aligned} \tag{11}$$

$z(\theta)$ defines the normalising constant. The network presents each input vector probability values through the energy function. The probability can be raised through alternating parameter θ in (10) to set the energy value.

The hidden modules contingent distributions h and input vector v in the RBM are defined as

$$\begin{aligned} P(h_j = 1|v) &= f\left(\sum_{i=1}^{N} W_{ij} v_i + b_j\right) \\ P(v_i = 1|h) &= f\left(\sum_{j=1}^{M} W_{ij} h_j + a_i\right) \\ f(x) &= \frac{1 - e^{-2x}}{1 + e^{-2x}} \end{aligned} \tag{12}$$

where (x) defines an activation function which is derived as tan*h* function. The activation function essence is to keep the specifications of the activated neuron and its mapping. In this study, tan*h* function is replaced as the activation function for sigmoid. If the input is among [−1, 1], the sigmoid value is alternated sensitively. While the input is out of or close to the interval, the sensitivity of sigmoid value will be reduced. When the network precision diminishes, the sigmoid value will be in the saturated status. Further, the output convergence of tan*h* is faster than function sigmoid. *Tanh's* input and output is able to keep a non-linear steady ascent and drop relevance to face the BP network gradient solution [27].

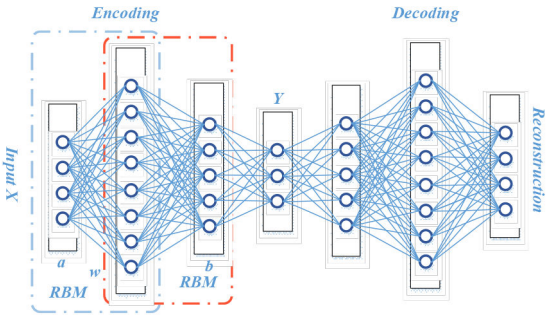

Fig. 2 *DAE network framework*

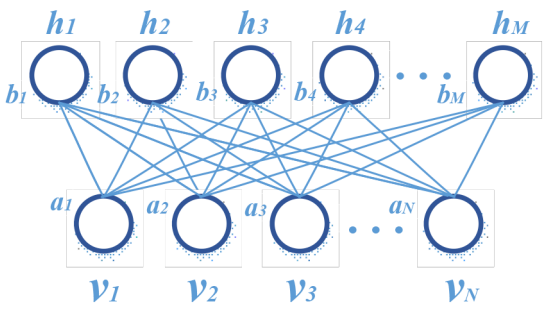

Fig. 3 *RBM framework*

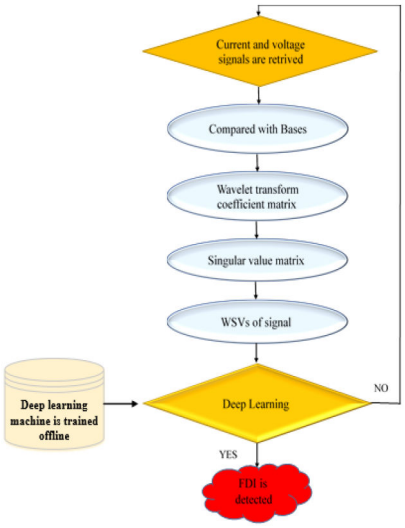

Fig. 4 *Suggested FDIA detection method*

In every RBM network, the hidden layer modules activation data is indicated as the input data extracted details. In other words, the RBM learning purpose is to gain the parameter θ to regain the main input information perfectly. Hence, the visible layer likelihood function v is made to gain the parameter θ in the following equation:

$$L(\theta; v) = \prod_{v} P_\theta(v) = \frac{\sum_h e^{-E_\theta(v,h)}}{\sum_{v,h} e^{-E_\theta(v,h)}} \tag{13}$$

As can be seen in (13), the algorithm is derived from both sides, and the derivative of logarithm function is offered by respect $\theta = \{w, b, a\}$ as follows:

$$\frac{\partial l\, n\left(L(\theta; v)^\partial\right)}{\partial \theta} = \frac{\sum \ln(P_\theta(v))}{\partial \theta} = \sum \left\{ E_{P_\theta(h|v)}\left[-\frac{\partial E_\theta(v,h)}{\partial \theta}\right] - E_{P_\theta(v,h)}\left[-\frac{\partial E_\theta(v,h)}{\partial \theta}\right]\right\} \tag{14}$$

Next, the gradient is estimated by the contrastive divergence (CD) method. The one-step CD learning is used to update the parameter θ [28]. The loss function is able to define as (15) to diminish the input X information lost and gain precise parameters in the DAE

$$J_{\mathrm{AE}}(\theta) = \frac{1}{N}\sum_{x \in X} R_e\left(X, f\left(\hat{X}\right)\right) \tag{15}$$

The parameter θ is gained through applying the gradient descent manner to reduce the loss function. Therefore, the update procedure can be explained as follows:

$$w^{k+1} = w^k + \varepsilon \frac{\partial J_{\mathrm{AE}}(\theta)}{\partial w} \tag{16}$$

$$a^{k+1} = a^k + \varepsilon \frac{\partial J_{\mathrm{AE}}(\theta)}{\partial a} \tag{17}$$

$$b^{k+1} = b^k + \varepsilon \frac{\partial J_{\mathrm{AE}}(\theta)}{\partial b} \tag{18}$$

That ε defines the learning rate. Then, an appropriate parameter θ is able to obtain through the RBM's hierarchical training procedure.

After training every RBM, the learnt data from the central control unit data of SI components stands in the hidden layer to be able to apply as the higher layer input to construct required data. These hidden layer parameters are obtained in order to finish the whole DAE network training θ. During the training phase, the central control unit's unlabelled data for long-term common operating states in the SI are chosen as the training data after the details are extracted by WT and SVD. The variable vector of every detail i is specified as X_i

$$X_i = [x_{i1}\ x_{i2} \ldots\ x_{im}] \tag{19}$$

i defines the SI parts name, x_{ij} defines the jth parameter in the central control unit variable vector of the detail i. For example, the SI variable vector consists of voltage, current, and so on to be able to reflect the status of SI.

To reduce computation error caused by the numerical differences in parameters of various kinds of SI parts and to keep the main information structure invariant, the central control unit information is preprocessed into the interval [0, 1] through normalisation.

4.3 Fine-tuning of DAE parameters

Each hidden layer's biases and weight of multi-layer RBMs are updated after the DAE hidden layers are trained with the RBMs, and the DAE layout framework will be constructed. The DAE network pre-training is an unsupervised learning of the central control unit information. Therefore, the learning outcomes are able to apply as a priori values for DAE network supervised learning.

The fine-tuning training merely requires local search on parameters gained through pre-training since the optimisation convergence time is significantly simplified in this procedures. Eventually, the parameters gained through these training procedures are better than being trained alone through the BP method [29]. After fine-tuning, the DAE network optimised parameters are retrieved.

For DAE networks of SI parts, the input to output mapping defines one-to-one, and any segment has the similar physical sense. Fig. 2 is exposed $\hat{X}$ and it defines the input reconstruction X, that is corresponded to the central control unit variables in X.

Hence, the status of SI components is able to evaluate through analysing the relevance among X and $\hat{X}$ (Fig. 4).

4.4 Proposed detection mechanism

Firstly, the current and voltage ingredients are regained and is decomposed with WT to extract precise factors and then detailed factors are analysed with SVD to extract singular values. Finally, these singular values are applied as the input indexed to deep learning to detect the FDI attacks.

In this work, the input to wavelet singular values (WSVs) approach contains 200 patterns. WSVs are extremely sensitive to the signal magnitude alternating and these WSVs, as input indexes of deep learning, are able to detect several FDI attack and diagnose load alternating from FDI attacks. Fig. 1 indicates the procedure of suggested method of attack detection.

5 Case studies

5.1 Cyber-physical model

An AC SI which consists of mth distributed generation units is displayed in Fig. 5. These units are linked to each other in parallel. Therefore, the voltage-frequency mode to stabilise the SI voltage and current control mode to share and divide load among DG units is defined by central control unit and send/receive data via fibre cable and wireless network [1]. Fig. 6 shows a typical layout of a SI unit and cyber-attack surfaces.

According to the sliding mode controller, the index of FDIA detection mechanism in voltage and current parameters is described as [1]

$$S_V = \tilde{\tilde{x}} - \lambda \tilde{x},\ \tilde{x} = x - x_{\mathrm{base}} \tag{20}$$

where S_V is the voltage index which is used to extract the detailed signal by WT and SVD, then the extracted signal is applied as the

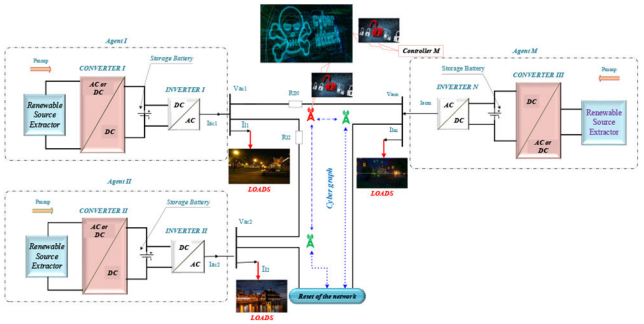

Fig. 5 *Cyber-physical layout of AC SI: blue arrows display the cyber layer and black lines display the physical circuit*

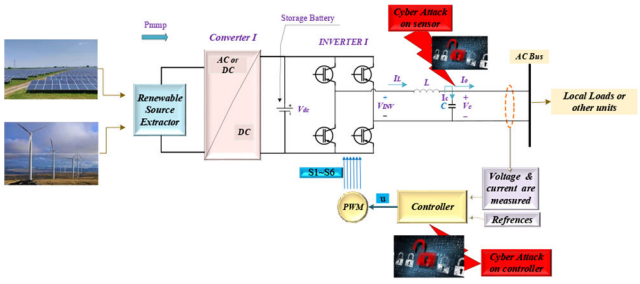

Fig. 6 *Typical layout of a unit of SI and cyber-attack points*

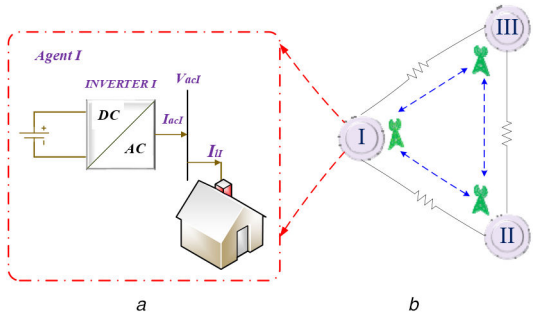

Fig. 7 *Two system models are illustrated*
(a) Agent layout, ***(b)*** Cyber-physical SI consists of three AC DG units

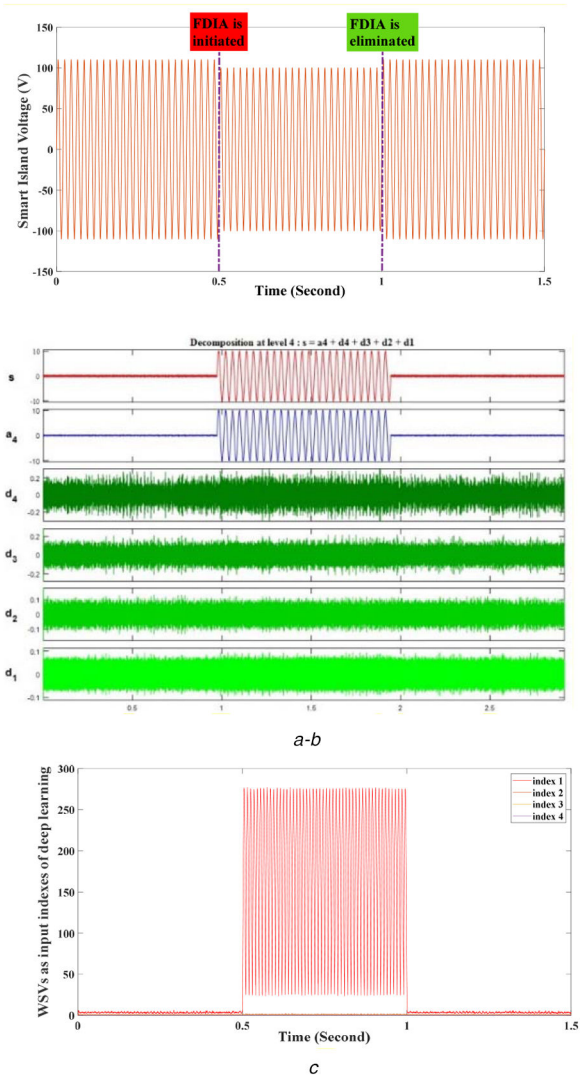

Fig. 8 *FDIA based on altering the amplitude of voltage reference signal*
(a) SI voltage, ***(b)*** Wavelet decomposition, ***(c)*** Input indexes of deep learning on basis of WSVs

input of deep learning to detect FDIA. λ is a positive number, x and x_{base} are the SI voltage and base voltage. The base voltage has a constant domain and frequency

$$S_I = z - z_{\text{base}} \tag{21}$$

where S_I is the current index which is used to extract the detailed signal by WT and SVD, then the extracted signal is applied as the input of deep learning to detect FDIA. z and z_{base} are the smart islanding current (product by any distributed generation) and the base current of loads which is defined via central control unit.

Fig. 3 illustrates an AC SI which is simulated in this study under several case studies. DC sources or uninterruptable power supply is coupled to the DC/AC inverters and inter-connected via timelines to create the SI physical layer. Every DC/AC inverters have a local primary and secondary controller. A communication network undirected cyber plan is investigated in this study that sends and receives data through its neighbours. Also, loads and the converter output of any agent are coupled to each other.

The diagram of communication is illustrated as a graph by links and edges by an adjacency matrix $\boldsymbol{B} = [b_{ij}] \epsilon R^{M*N}$, the proposed communication weights is shown as follows:

$$b_{ij} = \begin{cases} > 0, & \text{if } (z_i, z_j) \epsilon A \\ 0, & \text{else} \end{cases} \tag{22}$$

Wherever two nodes are linked together via an edge (A), z_i and z_j are the local node and neighbouring node, respectively. The communication weights illustrate data exchange only between two related nodes and can be demonstrated through a matrix with incoming data, $X_{\text{in}} = \sum i \epsilon M^{a_{ij}}$.

So, if both matrices were equal together, the Laplacian matrix $\boldsymbol{L}$ is balanced, that $L = X_{in} - B$ and its parts are set by

$$I_{ij} = \begin{cases} {}^{\circ}(m_i), & i = j \\ -1, & i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where the degree of ith node is $\deg(m_i)$ and $L = [I_{ij}] \epsilon R^{M \times N}$.

Remark I: All agents are gained consensus through $z(k+1) - z(k) = -\mu L z(k)$ for a well-spanned matrix $\boldsymbol{L}$. Therefore, $\lim_{k \to \infty} z_i(k) = c$, $\forall\, i \epsilon M$, where c and μ are constant and positive values, respectively. M is considered as the agents' number.

5.2 Simulation results

Fig. 7 exposed two system models as agent layout (see Fig. 6*a*) and cyber physical layout of a SI which consist of three DG units that are linked together via resistive lines (see Fig. 6*b*) with $V_{\text{ref}} = 110\sin(2*pi*60)$.

The system characteristics and control parameters are listed in [1]. To understand the study better, each attack occurs in the case study are divided by a determined time-gap.

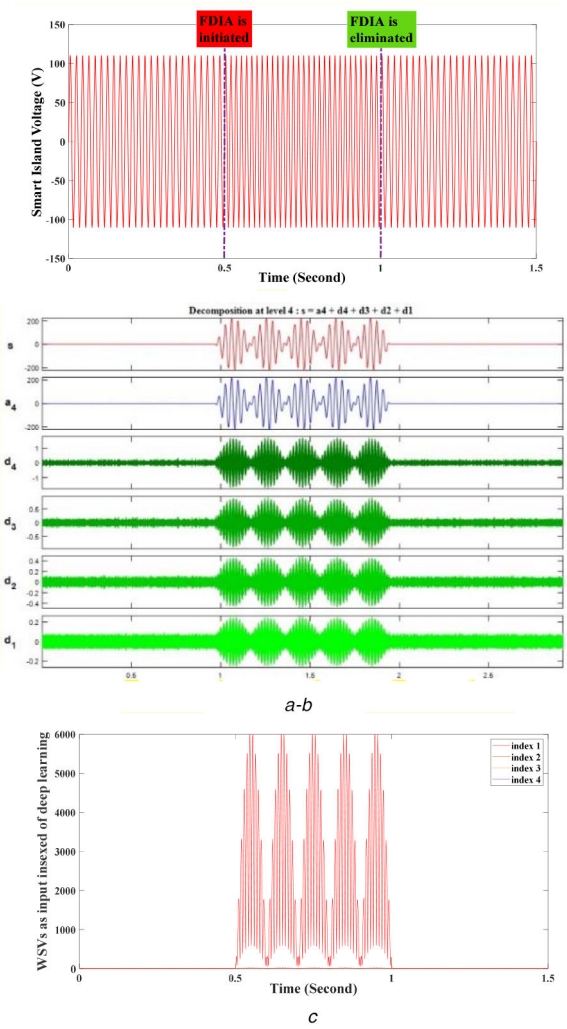

Fig. 9 *FDIA based on altering the frequency of voltage reference signal*
(a) SI voltage, ***(b)*** Wavelet decomposition, ***(c)*** Input indexes of deep learning on basis of WSVs

Case study I: In this state, the FDIA is on altering the amplitude of voltage reference signal. The system behaviour is investigated in a sample of this FDI attack type and deep learning indexes are extracted on basis of WSVs.

FDIA is started and removed at time $t = 0.5$ s and $t = 1$ s, respectively. The amplitude of voltage reference signal is changed (reduced by 10%). The simulation results of this case study are shown in Fig. 8. The SI voltage is demonstrated in Fig. 8*a*. The WTs of signal and wavelet decomposition at several levels are indicated in Fig. 8*b*. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are shown in Fig. 8*c*.

Case study II: In this state, the FDIA is on altering the frequency of voltage reference signal. The system behaviour is investigated in a sample of this FDI attack type and deep learning indexes extracted on basis of WSVs.

FDIA is started and removed at time $t = 0.5$ s and $t = 1$ s, respectively. The voltage frequency is altered from 60 to 50 Hz. The simulation results of this case study are illustrated in Fig. 9, where the SI voltage is demonstrated in Fig. 9*a*.

The WTs of signal and wavelet decomposition at several levels are shown in Fig. 9*b*. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are shown in Fig. 9*c*.

Case study III: In this state, the FDIA is on adding a white noise to voltage reference signal. The system behaviour is investigated in a sample of this FDI attack type and deep learning indexes extracted on basis of WSVs. FDIA is started and removed at time $t = 0.5$ s and $t = 1$ s, respectively. The voltage of SI is altered by mixing a noise with the reference signal of voltage via attack in the controller. The simulation results of this case study are illustrated in Fig. 10. The SI voltage is demonstrated in Fig. 10*a*.

The WTs of signal and wavelet decomposition at several levels are indicated in Fig. 10*b*. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are shown in Fig. 10*c*.

Case study IV: In this state, the FDIA is on sensor voltage signal or sending voltage via fibre cable or wireless network. The system behaviour is investigated in a sample of this FDI attack type and deep learning indexes extracted on basis of WSVs.

FDIA is started and removed at time $t = 0.5$ s and $t = 1$ s, respectively. The voltage of SI is altered by an attack in sensor voltage signal or sending voltage via fibre cable or wireless network. The simulation results of this case study are illustrated in Fig. 11. The SI voltage is demonstrated in Fig. 11*a*.

The WTs of signal and wavelet decomposition at several levels are extracted. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are shown in Fig. 11*b*.

Case study V: In this state, the FDIA is on altering the load reference current signal on agents II and III. The system behaviour

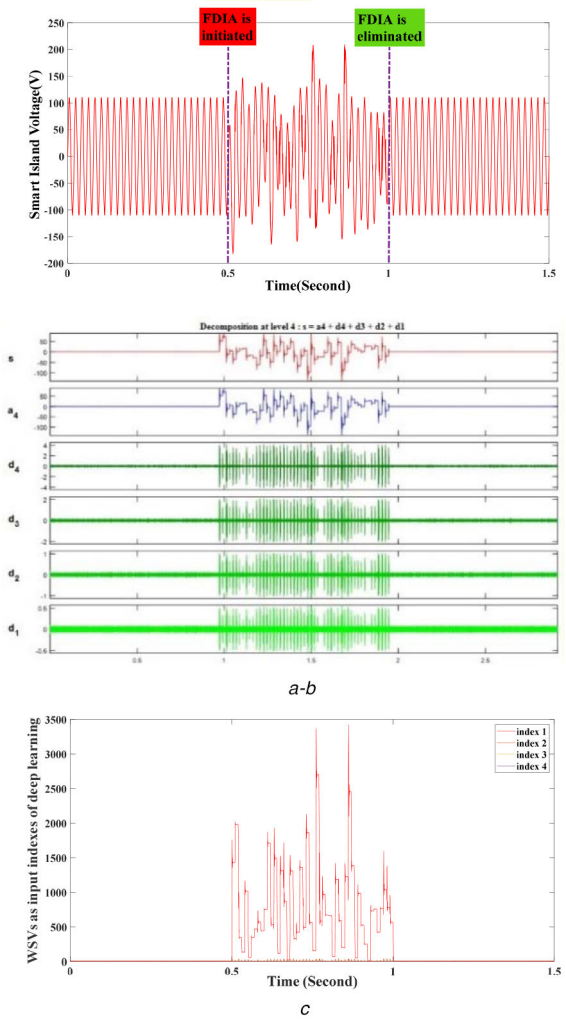

Fig. 10 *FDIA based on adding a white noise to voltage reference signal*
(a) SI voltage, ***(b)*** Wavelet decomposition, ***(c)*** Input indexes of deep learning on basis of WSVs

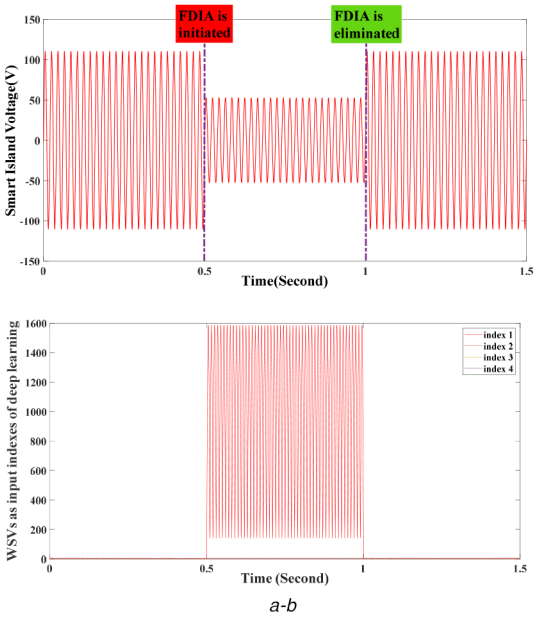

Fig. 11 *FDIA on sensor voltage signal (SI voltage diminished)*
(a) SI voltage, ***(b)*** Input indexes of deep learning on basis of WSVs

is investigated in a sample of this FDI attack type and deep learning indexes extracted on basis of WSVs.

FDIA is started and removed at time $t = 0.5\,\text{s}$ and $t = 1\,\text{s}$, respectively. The voltage of SI is altered by an attack in the load reference current signal on agents II and III. The simulation outcomes of this case study are illustrated in Fig. 12. Loads current is shown in Fig. 12*a*. The DGs current is depicted in Fig. 12*b*.

The WTs of signal and wavelet decomposition at several levels are indicated in Fig. 12*c*. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are shown in Fig. 12*d*.

Case study VI: In this state, the FDIA is on sensor current signal or sending current via fibre cable or wireless network in a way to deteriorate the current sharing profile among agents II and III. The system behaviour is investigated in a sample of this FDI attack type and deep learning indexes extracted on basis of WSVs.

FDIA is started and removed at time $t = 0.5\,\text{s}$ and $t = 1\,\text{s}$, respectively. The voltage of SI is altered by an attack in the sensor current signal or sending current via fibre cable or wireless network in a way to deteriorate the current sharing profile among agents II and III. The simulation results of this case study are shown in Fig. 13. The loads current is demonstrated in Fig. 13*a*. The DGs current is depicted in Fig. 13*b*.

The WTs of signal and wavelet decomposition at several levels are extracted. To obtain the efficient singular values in order to calculate the WSVs as input indexes of deep learning to detect cyber-attack, the wavelet coefficients ($d1$, …, $d4$) are engaged. The sample input indexes of deep learning based on WSVs are defined in Fig. 13*c*.

Case study VII: Loads alternation.

In this state, the system behaviour is investigated under various loads alternation and deep learning indexes extracted on basis of WSVs.

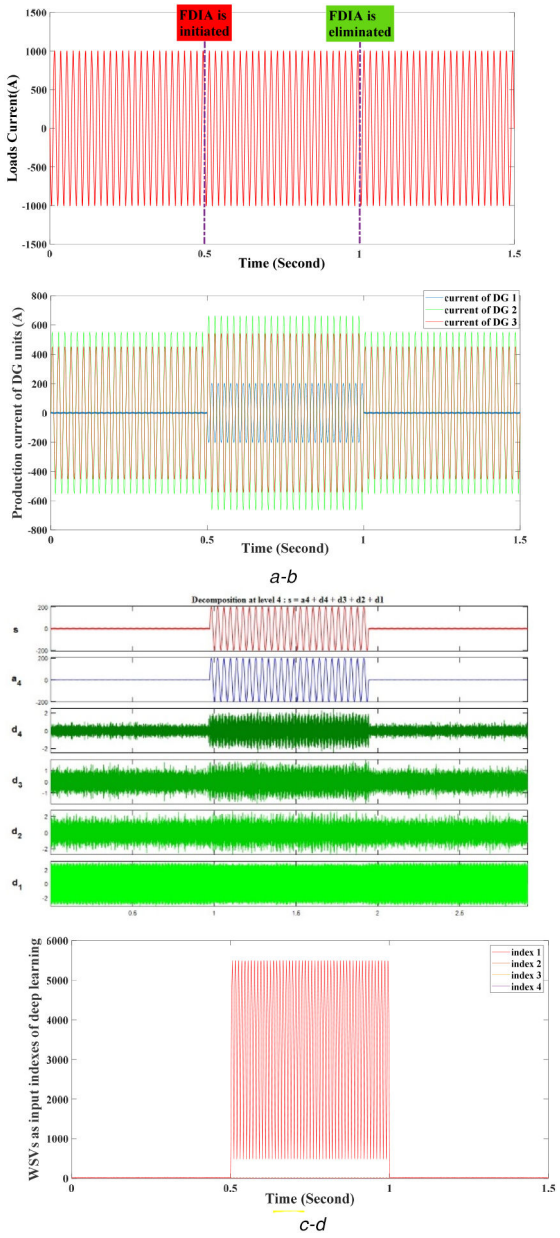

Fig. 12 *FDIA on altering the load reference current signal: current division on units 2 and 3 are deteriorated*

(a) Loads current, ***(b)*** Current of DGs, ***(c)*** Wavelet decomposition, ***(d)*** Input indexes of deep learning on basis of WSVs

Resistive, inductive, and non-linear loads are connected to the SI at time $t = 0.4$ s, $t = 0.8$ s, and $t = 1.2$ s, respectively. Linear loads are disconnected at time $t = 0.8$ s (as can be seen in Fig. 14*a*). The simulation results of this case study are illustrated in Fig. 14.

The sample input indexes of deep learning in normal condition without attack based on WSVs are shown in Figs. 14*b* and *c*.

Simulation results and discussion: A confusion matrix is used for performance evaluation, which represents the four possible outcomes when we compare the actual data point labels given by an expert to the corresponding data point results generated by a given classification algorithm. In this case, the four possible outcomes include: hit rate (*HR*), false alarm rate (*FR*), miss rate (*MR*), and correct reject (*CR*). To have a better realisation of these four criteria, the confusion matrix is provided in Table 1.

The suggested anomaly detection procedure can construct each of these four intentions in Table 1 as true negative (TN), false negative (FN), true positive (TP), and false positive (FP). These intentions are constructed on the basis of the actual system information and the suggested anomaly detection layout reaction.

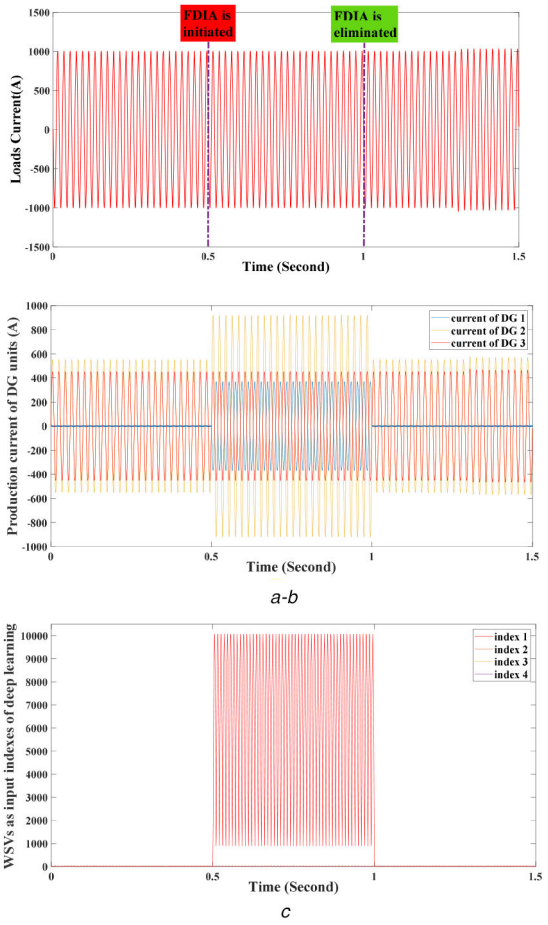

Fig. 13 *FDIA on sensor current signal: current division on units 2 and 3 are deteriorated*

(a) Loads current, ***(b)*** Current of DGs, ***(c)*** Input indexes of deep learning on basis of WSVs

To verify the validation of proposed deep learning in FDIA detection, several sample tests are used. The efficiency of suggested detection mechanism is evaluated by using the FDI attack model and the evaluation outcomes are offered. The performance of proposed detection mechanism is evaluated by using the FDI attack model and the evaluation outcomes are summarised in Tables 2 and 3. From Tables 2 and 3, it can be observed that the proposed mechanism is able to detect the FDI attacks with detection accuracy over 97%, that illustrates the efficiency of the suggested detection method on detecting the FDI attacks.

6 Conclusion

In this paper, a DAE deep learning network is presented to detect FDIA in an AC SI. Although several researches have been conducted for attacks and detections in DC systems, only a few works have focused on the AC peer that is widely adopted by SI. Proposed FDIA layout is focused on deep learning. The WSVs of compound WT and SVD to extract indexes are applied as input of deep learning to diagnose cyber-attack from normal conditions. Outcomes discover that WSVs are sensitive to sudden signals alternations and are capable of defining deep learning input indexes to detect FDIA. Therefore, this paper suggested a highly accurate and intelligent detection layout for securing the SIs in front of FDI attacks with detection accuracy over 97%. To evaluate the efficiency of our FDIA diagnostic mechanism, we performed a comprehensive series of simulations on an AC SI. The DAE network scheme is trained via normal and compromised condition of SI data. The planned detector is able to achieve outstanding attack diagnostic function. Also, it was explained how suggested model achieved superior performance in the face of malicious

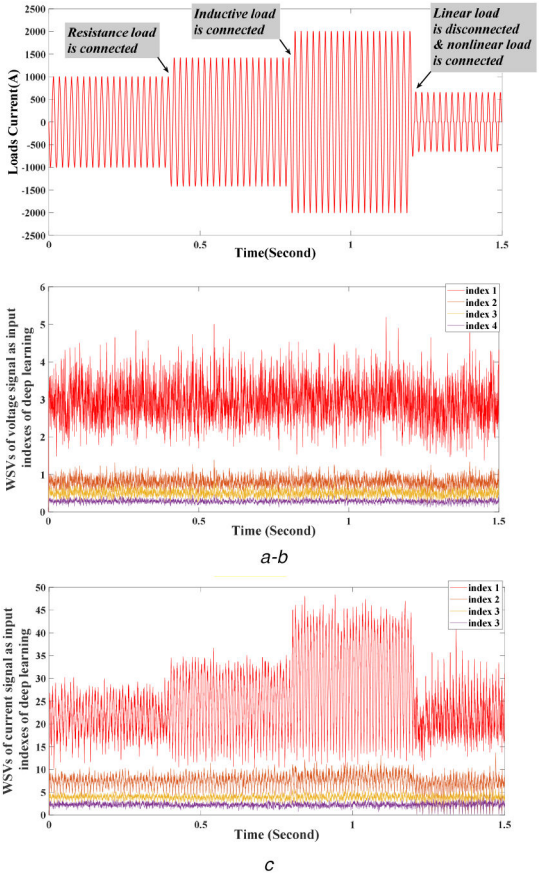

Fig. 14 *Loads alternation*

(a) Loads current, ***(b)*** Input indexes of deep learning on basis of WSVs of voltage, ***(c)*** Input indexes of deep learning on basis of WSVs of current

Table 1 Confusion matrix of the suggested detection layout

		Actual value	
Detection model response		Positives	Negatives
	Positives	hit rate TP	false alarm rate FP
	Negatives	miss rate FN	correct rejection rate TN

Table 2 Suggested detection layout

Label	Number of testing data	Identified to be compromised	Identified to be normal	Detection accuracy, %
compromised	1348	1312	36	97.36
normal	1174	14	1160	98.84

Table 3 Confusion matrix of the suggested detection layout

		Actual value	
Detection model response		Positives	Negatives
	Positives	97.36%	1.16%
	Negatives	2.64%	98.84%

attacks with various severities ranging from 10 to 100% data injection.

7 References

[1] Dehghani, M., Khooban, M.H., Niknam, T., *et al.*: 'Time-varying sliding mode control strategy for multibus low-voltage microgrids with parallel connected renewable power sources in islanding mode', *J. Energy Eng.*, 2016, **142**, (4), p. 05016002

[2] Khooban, M.-H., Dehghani, M., Dragičević, T.: 'Hardware-in-the-loop simulation for the testing of smart control in grid-connected solar power generation systems', *Int. J. Comput. Appl. Technol.*, 2018, **58**, (2), pp. 116–128

[3] Dragičević, T., Lu, X., Vasquez, J.C., *et al.*: 'DC microgrids—part II: A review of power architectures, applications, and standardization issues', *IEEE Trans. Power Electron.*, 2015, **31**, (5), pp. 3528–3549

[4] Ghiasi, M., Dehghani, M., Niknam, T., *et al.*: 'Investigating overall structure of cyber-attacks on smart-grid control systems to improve cyber resilience in power system', *IEEE Smart Grid Newsletter*, 2020, **1**, pp. 1–6

[5] Zhu, X., Xie, Z., Jing, S., *et al.*: 'Distributed virtual inertia control and stability analysis of dc microgrid', *IET Gener. Transm. Distrib.*, 2018, **12**, (14), pp. 3477–3486

[6] Mashayekh, S., Butler-Purry, K.L.: 'An integrated security-constrained model-based dynamic power management approach for isolated microgrids in all-electric ships', *IEEE Trans. Power Syst.*, 2015, **30**, (6), pp. 2934–2945

[7] Sun, C.-C., Hahn, A., Liu, C.-C.: 'Cyber security of a power grid: state-of-the-art', *Int. J. Electr. Power Energy Syst.*, 2018, **99**, pp. 45–56

[8] Babineau, G.L., Jones, R.A., Horowitz, B.: 'A system-aware cyber security method for shipboard control systems with a method described to evaluate cyber security solutions'. 2012 IEEE Conf. on Technologies for Homeland Security (HST), Waltham, MA, USA, 2012, pp. 99–104

[9] Wang, Q., Tai, W., Tang, Y., *et al.*: 'Review of the false data injection attack against the cyber-physical power system', *IET Cyber-Phys. Syst.: Theory Appl.*, 2019, **4**, (2), pp. 101–107

[10] Khan, R., Mclaughlin, K., Laverty, D., *et al.*: 'Design and implementation of security gateway for synchrophasor based real-time control and monitoring in smart grid', *Ieee Access*, 2017, **5**, pp. 11626–11644

[11] Xu, R., Wang, R., Guan, Z., *et al.*: 'Achieving efficient detection against false data injection attacks in smart grid', *IEEE Access*, 2017, **5**, pp. 13787–13798

[12] Alcaraz, C., Fernandez-Gago, C., Lopez, J.: 'An early warning system based on reputation for energy control systems', *IEEE Trans. Smart Grid*, 2011, **2**, (4), pp. 827–834

[13] Farraj, A., Hammad, E., Kundur, D.: 'A distributed control paradigm for smart grid to address attacks on data integrity and availability', *IEEE Trans. Signal Inf.Process. Over Netw.*, 2017, **4**, (1), pp. 70–81

[14] Abdollah, K.-F., Su, W., Jin, T.: 'A machine learning based cyber attack detection model for wireless sensor networks in microgrids', *IEEE Trans. Ind. Inf.*, 2020, **16**, pp. 1–8

[15] Liu, L., Esmalifalak, M., Ding, Q., *et al.*: 'Detecting false data injection attacks on power grid by sparse optimization', *IEEE Trans. Smart Grid*, 2014, **5**, (2), pp. 612–621

[16] Luo, X., Wang, X., Pan, X., *et al.*: 'Detection and isolation of false data injection attack for smart grids via unknown input observers', *IET Gener. Transm. Distrib.*, 2019, **13**, (8), pp. 1277–1286

[17] Rahman, A., Mohsenian-Rad, H.: 'False data injection attacks against nonlinear state estimation in smart power grids'. 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, Canada, 2013, pp. 1–5

[18] Moslemi, R., Velni, J.M., Mesbahi, A.: 'A weighted graph-based method for detection of data integrity attacks in electricity markets', *IET Gener. Transm. Distrib.*, 2019, **13**, pp. 1–9

[19] Soltan, S., Yannakakis, M., Zussman, G.: 'Power grid state estimation following a joint cyber and physical attack', *IEEE Trans. Control Netw. Syst.*, 2016, **5**, (1), pp. 499–512

[20] Heaton, J.: 'AIFH, volume 3: deep learning and neural networks', 2015

[21] Mulquin, M.J.: 'Roles of IEC in supporting effective smart city standards', *IET Smart Cities*, 2019, **1**, (1), pp. 10–18

[22] Liu, Y., Ning, P., Reiter, M.K.: 'False data injection attacks against state estimation in electric power grids', *ACM Trans. Inf. Syst. Sec. (TISSEC)*, 2011, **14**, (1), pp. 1–33

[23] Singh, S., Kumar, N.: 'Detection of bearing faults in mechanical systems using stator current monitoring', *IEEE Trans. Ind. Inf.*, 2016, **13**, (3), pp. 1341–1349

[24] Biswal, B., Vyshnavi, E., Metta, S., *et al.*: 'Robust retinal optic disc and optic cup segmentation via stationary wavelet transform and Maximum vessel pixel sum', *IET Image Process.*, 2019, **14**, pp. 592–602

[25] Dehghani, M., Khooban, M.H., Niknam, T.: 'Fast fault detection and classification based on a combination of wavelet singular entropy theory and fuzzy logic in distribution lines in the presence of distributed generations', *Int. J. Electr. Power Energy Syst.*, 2016, **78**, pp. 455–462

[26] Lei, Y., Jia, F., Zhou, X., *et al.*: 'A deep learning-based method for machinery health monitoring with big data', *J. Mech. Eng.*, 2015, **51**, (21), pp. 49–56

[27] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks'. Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012, pp. 1097–1105

[28] Hinton, G.E., Salakhutdinov, R.R.: 'Reducing the dimensionality of data with neural networks', *Science*, 2006, **313**, (5786), pp. 504–507

[29] Chen, Y., Lin, Z., Zhao, X., *et al.*: 'Deep learning-based classification of hyperspectral data', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2014, **7**, (6), pp. 2094–2107