Journal of Network and Systems Management (2021) 29:43
https://doi.org/10.1007/s10922-021-09610-y

An Ensemble Classifier Based Scheme for Detection of False Data Attacks Aiming at Disruption of Electricity Market Operation

Prasanta Kumar Jena[1] · Subhojit Ghosh[1] · Ebha Koley[1] · Murli Manohar[1]

Received: 26 September 2020 / Revised: 25 April 2021 / Accepted: 24 May 2021 /
Published online: 7 June 2021

Abstract

Wide area monitoring and control of modern power network demand real-time estimation of state variables from sensor measurements. Maintaining a high degree of reliability and accuracy in the state estimation process is important in avoiding any disruption in the electricity market operation. The market operation in power networks aims at providing a win-win situation for both the utility and consumer. The exposure and vulnerability of cyber components in smart grids allow for manipulating the electricity market by falsifying the state variables. The attacker can cause intentional profit/loss to the utility/consumer by misdirecting the estimated states through the injection of false data into the sensor information. Hence, maintaining integrity in the market operation demands a mechanism for detecting false data injection attack (FDIA). This paper proposes a classification-based approach for detecting FDIAs aiming at electricity market disruption. For any variation in the predicted and real-time nodal electricity price, the proposed decision tree (DT) based ensemble classifier is executed using state information to identify the prevailing scenario as a contingency or FDIA. The effectiveness of the proposed scheme has been extensively validated for various contingency and FDIA scenarios in IEEE 14 bus, 39 bus, and 57 bus test power systems.

Keywords Attack detection · Classifier · Contingency · Cyber attack · Decision tree · FDIA · LMP · Machine Learning · Smart grid · State estimation

✉ Subhojit Ghosh
sghosh.ele@nitrr.ac.in

Extended author information available on the last page of the article

1 Introduction

The traditional power system throughout the world is undergoing rapid changes with the increasing digitalization of system components, extensive accumulation of distributed energy resources, prevalent demand response, and inclusion of advanced protection and control equipment [1, 2]. The overall operation of the complex system through innovative monitoring of the system states has led to the development of smart grid. Reliable and resilient operation of the smart grid relies on the internet of things (IoT) technology-based monitoring and computing operations carried out at the control center [3, 4]. The dependence on cyber infrastructure, intelligent devices, and information and communication tools (ICT) has increased the vulnerability of smart grids to intimidating cyber-attacks [5–7]. Among the different cyber-attacks reported in the literature, FDIA is considered the most pertinent and severe attack [8]. Through FDIA, the attacker attempts to introduce malicious fake data into meter measurements, thereby manipulating the state estimation process. The manipulation is aimed at causing volatility in smart grid operation, which may even lead to a regional blackout.

The stable, secure and reliable operation of the smart grid is highly dependent on the real-time monitoring of the power system scenarios. Supervisory control and data acquisition (SCADA) systems receive real-time measurements from remote terminal units (RTUs) and transmit suitable control commands to the circuit breakers (CBs) and transmission system operators via a communication network. The overall monitoring and control of the smart grid are carried out by the energy management system (EMS). In contrast, the market management system (MMS) is responsible for performing all the economic operations in the smart grid [9]. The operations performed under EMS and MMS are highly dependent on the reliability of the states estimated from the sensor measurements. In a classical power system, the control actions rely solely on sensor measurements. In a smart grid, the inclusion of state information allows for improved monitoring and control during stressed scenarios like loss of sensor measurements and sensor failure.

The MMS executes the financial operations through the day-ahead market (DAM) scheduling and real-time market (RTM) price allocation [10]. The DAM involves purchasing and selling wholesale electricity for the next operating day through virtual bidding by the generation utilities to avoid price volatility. The final nodal price is evaluated in RTM operation as per the operating parameters executed in real-time. The final price settlement is performed in RTM in terms of locational marginal price (LMP) [11]. The LMP changes as a result of variation in the estimated states of the power system. Any disturbance arising out of load-generation imbalance or contingencies will impact the MMS operation through state variables. However, any contingency will have a more significant impact on the system states and hence, LMP. The possible contingencies include a change in network topology, generation failure and sudden loss of loads. To maintain accuracy in the state estimation task, bad data detection (BDD) methods are employed to filter out spurious meter measurements and/or noise arising out of faulty meters [12].

The attacker can exploit the high dependence of MMS on sensor information to cause financial mismanagement in the grid operation by falsifying the measurement data. The vulnerable cyber layer allows an attacker to manipulate the market operation by increasing/decreasing the profit/loss margin to a targeted utility/consumer. It can be achieved by injecting false data into the measurement within the endurable limits of BDD, thereby falsifying the state estimation process and LMP calculation. Precise information regarding the grid configuration eases the task of an attacker to launch FDIA aiming at the disruption of MMS operation [13]. The electricity market's financial misconduct due to FDIA causes economic loss to the consumers and power suppliers due to incorrect electricity price. Simultaneously, the attacker makes a profit from the gap in the actual and manipulated price. Thus, ensuring integrity and dynamic balancing in the market operation demands a mechanism for the early detection of FDIA.

In spite of the wide volume of work reported for detecting FDIA, no technique has addressed the detection of FDIAs aiming at the disruption of MMS in the smart grid. Early detection enables the independent system operator (ISO) to take necessary control action towards eliminating the financial mismanagement in the electricity market. Motivated by the significance of maintaining fairness in MMS operation and the possibility of launching FDIA in the cyber layer, the present work proposes a classification-based scheme for detecting market-oriented FDIAs. Considering the significant impact of state variables on the market operation, the proposed scheme is formulated by mapping the system states with the prevailing market scenario (healthy, power system contingency or FDIA). The proposed FDIA detection scheme is processed in two stages. In the first stage, the LMP is monitored in real-time to detect any substantial variation between the day-ahead price allocated and the real-time settlement price. For any significant variation in the price, the classifier is executed in the second stage to detect the prevailing scenario as a contingency or FDIA. The classifier is executed by feeding the states estimated in real-time as input. Before detecting FDIA in real-time, the classifier is trained with a simulated dataset comprising state variables for various contingency cases and FDIAs.

Considering the wide range of operating scenarios of the grid and the ineffectiveness of a single stand-alone classifier in providing accurate and unbiased results for complex multi-dimensional datasets, an ensemble of multiple classifiers has been used to solve the classification problem [14]. The ability of DTs in attaining high accuracy with increased robustness and low computational cost has been utilized to implement the ensemble classifier by a set of DTs. The effectiveness of the ensemble classifier in achieving high classification accuracy for small and multi-dimensional data set has been outlined in [15]. Unlike an isolated classifier with the possibility of overfitting and biasness towards a particular class, in ensemble, the output of a set of classifiers is aggregated using a majority voting strategy to provide the final output.

The proposed scheme has been extensively validated for varying operating scenarios of IEEE 14 bus, 39 bus and 57 bus power systems. The effectiveness of the scheme has been analysed in terms of its ability to discriminate between contingency and FDIA. The major highlights/ contributions of the proposed work can be summarized as

1. Assessing the impact of FDIA towards financial mismanagement of electricity market operation in terms of deviation in nodal price between day-ahead and real-time electricity market.
2. Formulating the market oriented FDIA detection scheme as a classification problem.
3. Solving the classification task using an ensemble of DTs while considering the real-time estimated state variables as discriminatory attributes.
4. Validation of the proposed attack detection scheme for varying scenarios of contingency and FDIA in IEEE 14 bus, 39 bus and 57 bus test power system.

The rest of the paper is organized as follows. In Sect. 2, the fundamentals of state estimation process and FDIA have been discussed. Section 3, demonstrates the electricity market mechanism in smart grid along with the formulation and impact of FDIA in causing financial mismanagement. In Sect. 4, the proposed FDIA detection scheme is discussed, and the methodology is validated in Sect. 5. Section 6 summarizes the concluding remarks of the proposed work.

1.1 Related Work

A number of FDIA detection schemes have been proposed in the literature. A cumulative sum (CUSUM) algorithm based on the generalized likelihood approach for FDIA detection has been presented in [16]. The other notable detection techniques include schemes based on collaborative intrusion detection [17], sparse optimization [18], unknown input observer (UIO) [19], Go decomposition approach [20], spatial-temporal correlations [21] and Kullback-Leibler distance estimation [22]. In addition to the analytical approaches, the effectiveness of machine learning in identifying intricate patterns from the multidimensional data of complex systems has been utilized for detecting cyber-attacks. In [23], an intrusion detection system has been proposed using the deep learning technique. In [24], a random forest classifier based attack detection scheme based on information and logs obtained from phasor measurement units (PMUs) has been proposed. Compromised meters in data integrity attacks have been identified using artificial intelligence in [25]. In [26], neural network and naive Bayes classification scheme is proposed for anomaly detection in load forecasting during cyber-attacks. In [27], the intrusion detection capability has been enhanced by employing a two-level hybrid anomaly detection mechanism for the smart grid. The cyber-attack is classified against normal fault scenarios by training the classifier with a dataset comprising physical and cyber layer information [28]. In [29], the consumer energy consumption pattern is classified from a malicious pattern to detect energy theft using an ensemble-based classifier. In [30], dynamic state prediction using an ensemble-based DT approach is proposed to compare the predicted result with the real-time values for security analysis of a power system network. Although many works have been reported on the detection of FDIA launched to disrupt different operations carried out at the control center, no work has addressed the detection and impact of FDIA on MMS operation. Since the electricity market operation is not continuously monitored by protective relaying equipment, it is difficult to detect any malicious intrusion in the power network in real-time. However, the intrusion is reflected in the form of irregularity in MMS operation, which further causes economic losses/profit to the utility/consumer. Thus, availing the economic benefits of smart grid demands a reliable electricity market operation with robustness against possible manipulation of sensor information. The summary of all the notations/variables used in describing the proposed work is given in Table 1.

Table 1 Summary of notations/variables

Notation/variable	
A	Bus-branch incidence matrix
c	Arbitrary integer vector
C_{gi}	Generation Cost for the generator (gi)
D	Total number of loads
e	Error vector
dev	Nodal price deviation
FP, FN	False positive and false negative
FPR	False positive rate
H	Measurement Jacobian matrix
k_m	Compromised set of meters
m, n	Total number of sensors and states
N_b	Total number of buses
N_g	Total number of generators
N_{tl}	Total number of transmission lines
P_{fl}	Power flow in a transmission line
P^*_{fl}	Optimum branch power flow for the anticipated load
$P^{min}_{fl}, P^{min}_{fl}$	Minimum and maximum power transmission limits
P^*_{gi}, L^*_d	Optimum power generation allocated for the anticipated Load (L^*_d)
$P^{min}_{gi}, P^{min}_{gi}$	Minimum and maximum power generation limits
r	Measurement residual
R	Covariance matrix
TP, TN	True positive and true negative samples
TPR	True positive rate
x	System states
x_{attack}	System states post cyber-attack
Z_a	Attack vector
ζ	Network topology
Z_{meas}	Measurement vector
λ	Lagrangian multiplier for nodal electricity price
$\lambda_{rt}, \lambda_{da}$	Real-time and day-ahead nodal electricity price

2 False Data Injection Attack on State Estimation

2.1 State Estimation and Bad Data Detection

State estimation is the core operation performed at EMS to identify the precise operational state of the system. The measurement set (Z_{meas}) is collected from the sensors deployed throughout the power system network and is transmitted to the SCADA system through the communication channel. The measurement set comprises of bus voltage magnitude (V_b), real and reactive bus power injection (P_{inj}, Q_{inj}), and the real (P_{fl}) and reactive (Q_{fl}) branch power flows. The states estimated in a N_b bus power system consists of $2N_b - 1$ states i.e. N_b voltage magnitudes and $N_b - 1$ bus angles [12]. The measurement set can be represented as

$$Z_{meas} = Hx + e \tag{1}$$

The states estimated using weighted least squares (WLS) algorithm can be evaluated as

$$\hat{x} = (H^T RH)^{-1} H^T RZ_{meas} \tag{2}$$

Since the measurements are sensitive to noise, calibration of instruments and several internal/ external factors or any inconsistency in estimation pose a threat to the integrity of estimated states. Hence, it is subjected to bad data detection (BDD) test to identify the occurrence of any abnormality in the measurement set. According to the largest normalized residual (LNR) test, a sensor measurement is considered valid only if the measurement residuals, i.e. the difference between the measured value and estimated value, are found to be less than the threshold value (α) i.e. [12]

$$r = ||Z_{meas} - H\hat{x}|| \leq \alpha \tag{3}$$

However, if an intruder can obtain the information regarding H matrix, a measurement set can be developed that can pass the BDD test. Manipulated sensor measurements passing the BDD test are known as false data.

2.2 False Data Injection Attack (FDIA)

Cyber-attack aims at manipulating the sensor measurements by injecting falsified sensor information into the communication channel during transmission from the remote terminal unit (RTU) to the control center. It can either be a random cyber-attack or a cyber-topology attack , where an arbitrary set of measurements or network connectivity status is manipulated to disturb the state estimation process. In FDIA, an intruder possessing information of the system configuration develops a properly formulated measurement dataset, that can falsify the estimated states.

In a measurement set Z_{meas}, the attack vector Z_a can be injected such that $Z_{attack} = Z_{meas} + Z_a$; so that the states can be deviated to $x_{attack} = \hat{x} + x_a$. The states estimated after the launch of FDIA is given as

$$
\begin{aligned}
x_{attack} &= (H^T R H)^{-1} H^T R Z_{attack} \\
&= (H^T R H)^{-1} H^T R (Z_{meas} + Z_a) \\
&= \hat{x} + x_a
\end{aligned}
\tag{4}
$$

and the measurement residual evaluated after the FDIA is given as

$$
\begin{aligned}
r(x_{attack}) &= ||Z_{attack} - Hx_{attack}|| \\
&= ||Z_{meas} - H\hat{x} + (Z_a - Hx_a)||
\end{aligned}
\tag{5}
$$

If Z_a is injected equal to Hx_a, then $r(x_{attack}) = r(\hat{x})$, and hence, it can pass the BDD test. This allows the intruder to send false information for desired value of deviation in estimated states [31].

3 Formulation of False Data Injection Attack (FDIA) for Electricity Market Disruption

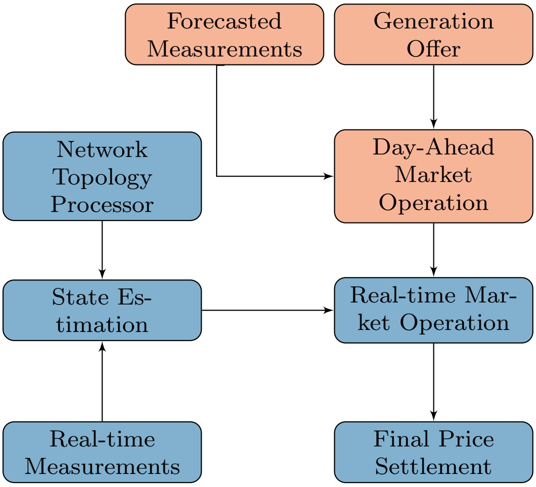
3.1 MMS Based Electricity Market Operation in Smart Grid

The deregulation in the electricity market industry imparts economic benefits to the end-user by providing cheaper electricity. It also provides bilateral involvement of utility and customer in the energy buying and selling mechanism to increase the satisfaction level at each stage of the power transmission process. The market operation performed by ISO is executed in two stages. The DAM is performed by solving OPF for the forecasted network scenario. The solution schedules the hourly energy price as per the generation offers, purchase/selling through bids by market participants, and physical constraints of the power network. It is solved by formulating the power delivery as a cost-minimizing optimization problem satisfying load generation balance while considering the physical and operating system constraints. The second stage of market settlement is performed in real-time and solved as per the states estimated from the sensor measurements obtained from the SCADA system and the network configuration identified by the network topology processor (NTP). The state estimation directs the optimal power flow (OPF) to allocate the generation schedule, and branch power flows, considering the system constraints and dispatch schedule cleared during the DAM settlement. Figure 1 depicts the block diagram of the state-dependent MMS operation performed to allocate the final nodal price in the real-time market settlement.

The nodal price calculated can be formulated as an optimization problem that aims at minimizing the following generation cost and is formulated as [13]

$$
\min_{P^*_{gi}} \quad \sum_{i=1}^{N_g} C_{gi}(P^*_{gi}) \tag{6}
$$

Fig. 1 Block diagram of electricity market operation



subjected to

$$\begin{aligned} &\lambda : \sum_{i=1}^{N_g} P_{gi}^{*} = \sum_{d=1}^{D} L_d^{*} \\ &P_{gi}^{min} \le P_{gi}^{*} \le P_{gi}^{max}, \quad \forall i = 1, 2, \dots N_g \\ &P_{fl}^{min} \le P_{fl}^{*} \le P_{fl}^{max}, \quad \forall fl = 1, 2, \dots N_{tl} \end{aligned} \tag{7}$$

The nodal price calculated during RTM can deviate from that of DAM allocated price by a physical or cyber-attack which falsify the sensor measurements or topology status. The attack results in an OPF solution different from real scenario. The incremental cost of generation and transmission resulting due to the new OPF creates unethical market mismanagement by increasing the cost of energy delivery.

3.2 False Data Injection Attack (FDIA) in Electricity Market

In this section, the possible attack scenarios aiming at the disruption of electricity market operation has been analysed. The analysis is carried out in terms of quantifying the impact of the attack on the deviation in nodal price between DAM and RTM settlement. As mentioned earlier in MMS, financial misconduct can be achieved by manipulating the sensor measurements while bypassing BDD. For launching the FDIA, a strong adversary possessing the following capabilities has been considered.

1. Complete knowledge about the system configuration.
2. Information regarding the optimal states and dispatch schedule for the day-ahead market settlement.
3. Accessibility and hence ability to manipulate all the sensor measurements. In case of a set of sensors being protected, accessibility is assumed for the unprotected sensors.

With the knowledge of system configuration and line parameters, the intruder can construct the H matrix (1). Thus, an attack model can be formulated to inject a false data vector such that the residual condition of (3) can be satisfied, thereby circumventing the BDD test. The difference between nodal price evaluated during DAM (λ_{da}) and RTM (λ_{rt}) reflects the profit acquired by the market participant by selling/ buying electricity during virtual trading. The deviation in nodal price at bus i is

$$dev_i = \lambda_{rt_i} - \lambda_{da_i} \tag{8}$$

The deviation dev_i can be manipulated by injecting false data to cause intentional profit/loss to a utility. Such attacks refrain the ISO from taking desired financial decisions and hence, the integrity of MMS is compromised.

3.3 Cyber Attack Formulation

The FDIA can be launched in the communication channel by injecting false data to either manipulate the sensor measurements (random cyber-attack) or/and network connectivity status (cyber topology attack). However, to design the corresponding attack vector, a set of constraints needs to be satisfied depending on the capability of the adversaries to access and manipulate sensor data. The formulation of possible attack vectors for different scenarios of intruder accessibility are dealt in the subsequent sub-sections.

3.3.1 Random Cyber Attack

The random cyber-attack aims at manipulating only the analog sensor measurements by injecting false data into the communication channel. The attack can be executed under the following constraints of intruder accessibility.

(a) Intruder having access to all the sensors All the sensors are assumed to be vulnerable for injecting false data into the measurement vector. The attack vector Z_a can be generated with the knowledge of measurement Jacobian matrix H. With the injected attack vector, the measurement vector is manipulated as [8]

$$Z_a = Hc \tag{9}$$

where c is an arbitrary integer vector. The generation of Z_a can be framed as an optimization problem which aims at manipulating the measurement vector Z_{meas} for maximizing dev_i, such that the branch power flows are constrained within the limits. Thus, for manipulating the MMS operation, the objective function can be framed as

$$\max_{Z_a}(dev_i) \tag{10}$$

subject to

$$P_{gi}^{min} \leq P_{gi} + \Delta P_{gi} \leq P_{gi}^{max}, \forall i = 1, 2, \dots N_g$$
$$P_{fl}^{min} \leq P_{fl} + \Delta P_{fl} \leq P_{fl}^{max}, \forall fl = 1, 2, \dots N_{tl}$$

(b) Intruder having access to limited sensors Often the availability of limited budget and protection of certain strategic sensor through necessary security protocol hinders launch of FDIA on all the sensors. For such cases of limited sensors accessibility, the attack vector can be generated if the following condition is satisfied [32].

$$k_m \geq (m - n + 1) \tag{11}$$

where k_m is the set of compromised meters out of m sensors and n is the total number of states. Under such condition the attack vector can be formulated as per (9).

$$Z_a = Hc$$

where $c = b * a$ such that a is an integer vector and b is the binary vector in which the protected and unprotected sensors are represented by '1' and '0' respectively, i.e.

$$\begin{aligned} b_i &= \begin{cases} 1 & \mathit{if}\ i \in k_m \\ 0 & \mathit{if}\ i \notin k_m \end{cases} \\ c &= [a]_{m*1}[b]_{1*m} \end{aligned} \tag{12}$$

3.4 Cyber-Topology Attack

Based on the digital information received regarding the status of circuit breakers/switches, the processor continuously monitors the network information through NTP. Any inconsistency in analog/digital sensor information can be easily identified from the estimated states. However, a well-formulated cyber topology attack aiming at manipulation of network topology can evade possible attack detection by falsifying both the analog and digital sensor information. The NTP processes the network configuration in real-time using the line status information as [33]

$$A[i,j] = \begin{cases} 1 & \mathit{if\ line\ connects\ from\ i\ to\ j} \\ -1 & \mathit{if\ line\ connects\ from\ j\ to\ i} \\ 0 & \mathit{if\ no\ line\ connection\ exists} \end{cases} \tag{13}$$

For manipulating the network topology, the adversary falsifies the circuit breaker/switch status by altering 1s to 0s in matrix A or vice versa. The change in the information fed to the NTP is simultaneously accompanied by injecting false data in the analog measurements to mask the impact of topology attack at the control centre. The false data injection vector for cyber topology attack can be formulated as

$$Z_a = -\sum_{(i,j)\in\Delta\zeta} P_{fl,i-j} A(i,j) \tag{14}$$

where $\Delta\zeta$ represents the changed network topology.

For initiating an attack under the above scenarios, an adversary can obtain the system matrix by accessing parameter and configuration information from an internal employee or using the day-ahead market data available on the website.

Depending upon the objective to disrupt the market operation and constraints on the part of an adversary, either of the above three attacks can be executed. The BDD passed attack vectors can mislead the ISO by falsifying estimated states, thereby refraining the operator to take any control action against the unethical practice. This urges a mechanism to detect the attack by discriminating any disturbance at the control centre as a contingency or FDIA scenario.

3.5 Impact of False Data Injection Attack (FDIA) on Electricity Market Operation

The impact of FDIA on the state estimation task and further on the market operation has examined in this section. For the IEEE 14 bus test system shown in Fig. 2, the measurement set for state estimation comprises all the bus power injections and branch power flows obtained by solving the OPF under various operating conditions. The DAM operation is executed under a normal or anticipated contingency scenario (pre-attack), while the RTM operation is performed considering various FDIAs (post-attack) as discussed in Sect. 3.3. For the IEEE 14 bus test system (Fig. 2) with the attack vector Z_a (9), the states estimated (bus angles) for actual (Z_{meas}) and falsified (Z_{attack}) measurements are shown in Table 2. The nodal electricity prices at all the buses under attack scenarios are estimated higher than the healthy condition. Hence, for the attack scenario, the market participant can buy electricity at various nodes during the DAM scheduling and sell during RTM to make an unethical profit. For both the examined state vectors, the corresponding market operation is performed, and the nodal price at different buses are depicted in Fig. 3.

4 Ensemble Classifier for FDIA Detection

In this section, the market-oriented FDIA detection task is formulated as a classification problem. The DAM and RTM operational procedure in allocating nodal electricity price is processed through a classification scheme to detect intrusion of

Fig. 2 IEEE 14 bus system

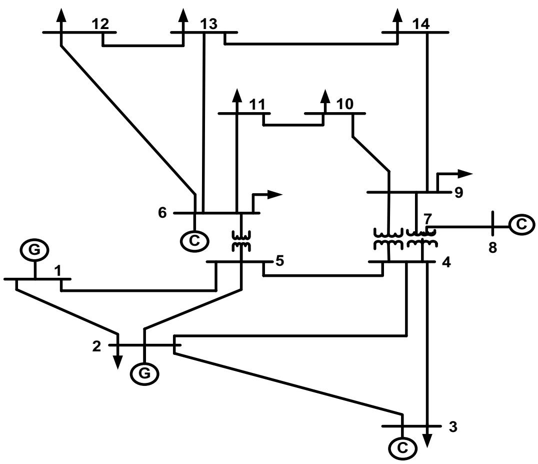


Table 2 State estimation under pre-attack and post-attack scenario for IEEE 14 bus test system

Bus	Bus angle (Degree)	
	Pre-attack	Post-attack
1	0.00	0.00
2	– 4.02	– 6.51
3	– 9.93	– 12.59
4	– 8.66	– 12.74
5	– 7.43	– 12.20
6	– 12.63	– 15.19
7	– 11.19	– 14.26
8	– 10.41	– 11.27
9	– 13.00	– 16.32
10	– 13.03	– 17.49
11	– 13.23	– 15.58
12	– 13.53	– 15.98
13	– 13.58	– 16.08
14	– 14.27	– 17.67

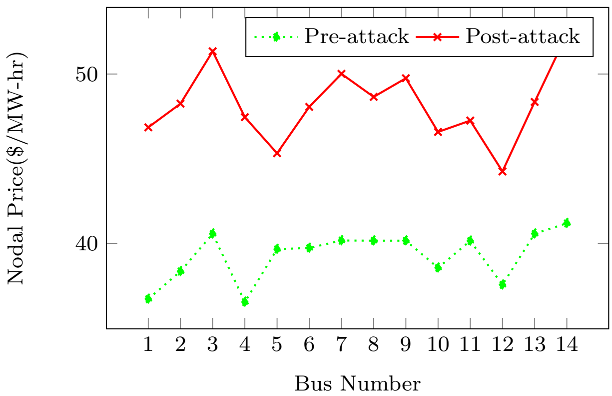


Fig. 3 Nodal price at different buses for Pre-attack and Post-attack states of Table 2

any inconsistency in the data received at the control centre (Fig. 4). Any marginal deviation (dev_i) between the DAM and RTM allocated nodal price is processed by the classifier to categorize the prevailing scenario as an actual or FDIA induced contingency.

4.1 Decision Tree (DT) Based Ensemble of Classifier

In recent times, DT has emerged as a powerful machine learning tool for solving complex classification problems in multi-dimensional space. The learning capability of DT has been successfully applied to solve complex regression and

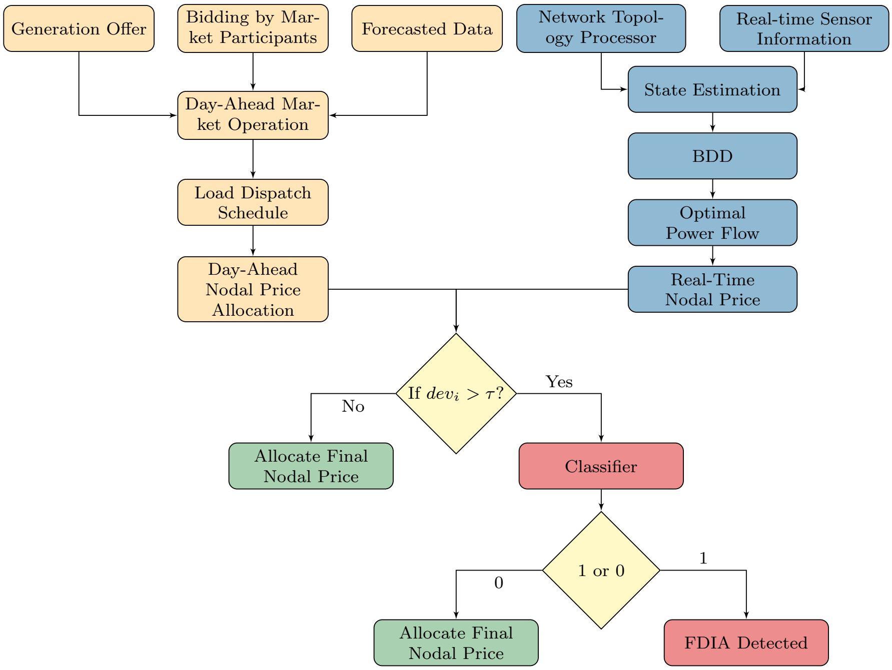


Fig. 4 Flow chart for proposed methodology

classification tasks in different domains of power systems, i.e. cybersecurity, transient stability, intrusion detection, and islanding detection. Being a rule-based approach, DT is more transparent and human friendly as compared to black-box solutions like neural networks. For a given dataset, a DT considers all the possible mapping in feature space to designate a particular class/ category for a given input. Further, logical operations to correlate the features with the class allow for straightforward interpretation and easier real-time implementation on a digital platform. The effectiveness of DT in achieving high classification accuracy for high dimensional complex dataset has been outlined in the literature. Motivated by the same, the present binary classification problem of categorizing the status of the power network as attack/healthy scenario has been performed using DT.

However, in spite of their effectiveness in achieving high classification accuracy, quite often, DTs tend to overfit, which leads to improper generalization during validation [34]. Also, often DTs fail to incorporate intricate information embedded in the dataset during the mapping because of low bias and high variance property. The same leads to wider variation in the classification accuracy for a minor change in the learning variable. Such instability in the prediction due to the biasness of individual DTs can be overcome by incorporating the learning ability of a set (ensemble) of DTs. With the consideration of the mapping characteristics of a set of DTs, the limitation pertaining to the weak learning ability of an individual standalone DT is avoided. Thus, by employing an ensemble of DTs, the high variance and overfitting of an individual classifier are nullified by utilizing the mapping characteristics of multiple DTs [35].

For a given dataset, the classification accuracy is improved by combining the individual output of all the classifiers using a voting strategy [36]. Among the ensemble algorithm, the bootstrap aggregation/bagging scheme has been widely used for complex classification tasks aiming at low variance because of its reduced sensitivity to noisy dataset and simpler implementation. The bagging algorithm initiates by partitioning the training set into a group of subsets of the same size by sampling with replacement [37]. Further for each of the *N* subsets (Fig. 5), a DT model is fitted to the desired level of classification accuracy. The output (predicted class) derived from all the '*N*' DTs are combined using a weighted algorithm. Each DT contributes to the final output in direct proportion to its classification performance. The overfitting of an individual DT does not impact the final output, since the same is offset by other DTs of the ensemble.

4.2 DT Based Ensemble Classifier Design for FDIA Detection

Following the formulation of the FDIA detection task as a classification problem, different disturbances in the smart grid market operation is classified as a FDIA or contingency using an ensemble of DT classifier. Figure 5 outlines the sequence of operations employed to classify a disturbance by the ensemble classifier as a FDIA or contingency. It is to be noted that, the classifier is activated only if a deviation

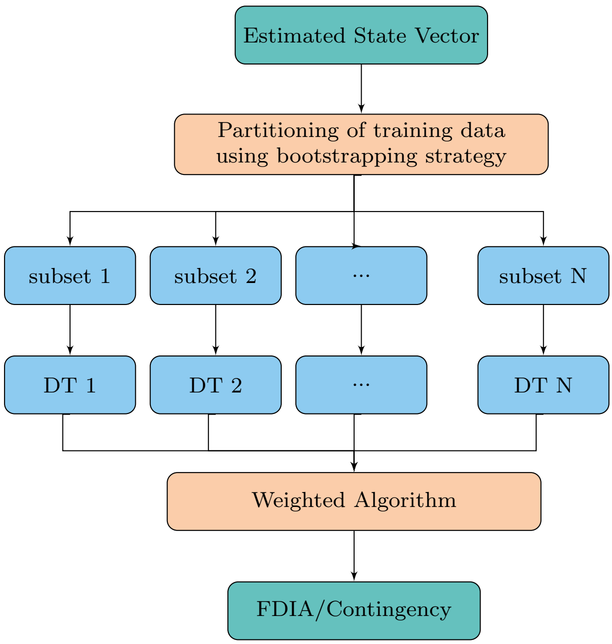


Fig. 5 Proposed DT based ensemble classifier for FDIA detection

in between the nodal price for day-ahead (λ_{da-i}) and real-time (λ_{rt-i}) is observed (Fig. 4). Following the partitioning of data samples, all the dataset is trained simultaneously by a set of DTs. After training to the desired level of pre-defined accuracy, the individual output of each DT is combined to form the ensemble of classifier. For arriving at the final output, the widely used weighted majority algorithm is adopted. For the present binary classification problem involving two classes i.e. FDIA and contingency, considering the output of the ith DT being represented as $d_{i,j} \in [0, 1]$, where $i = 1, 2, \ldots, N$ and $j = 1, 2$. For FDIA, the output is assigned as 1, while $d_{i,j} = 0$ for contingency. The values of $d_{i,j}$ derived for all the 'N' DTs are accumulated and fed to the weighted majority algorithm. For every classifier, a weight (w_t) is assigned in proportion to the classification accuracy. The class receiving the maximum outcomes is assigned as the final class of ensemble of classifier. The final output is represented as [34]

$$\sum_{i=1}^{N} w_t d_{i,j} = \max_{j=1}^{2} \sum_{i=1}^{N} w_t d_{i,j} \tag{15}$$

Post-training of an ensemble of DT, any new scenario of MMS operation satisfying $\lambda_{da-i} \neq \lambda_{rt-i}$ is processed by the ensemble classifier with the corresponding state variables for detection of FDIA (if any). For the scenario being classified as a contingency, the final nodal price as calculated using (7) is allocated by the operator. In the case of a FDIA, necessary control measures are initiated to mitigate the impact of the data falsification on the market operation. The pseudo-code for the proposed ensemble of DT classifier for FDIA detection is given in Algorithm 1.

Algorithm 1 Pseudocode for ensemble classifier

Input: $\hat{x}$=Training dataset, p=boot strapped data percentage, N=Number of base classifiers.

for $i = 1$to$N$ **do**

1.R_i=Randomly drawn sample of $p\%$ from the training dataset $(\hat{x})$ with replacement.

2. Process the classifier (DT_i) with R_i.

3. Aggregate the individual DT_i to the ensemble classifier.

end for

Output:

1.Choose the unlabelled dataset y_s.

2.Evaluate the ensemble $[DT_1, DT_2, ..., DT_i]$ on y_s.

$$d_{i,j} = \begin{cases} 1 & if\ Attack\ detected \\ 0 & if\ No\ attack\ detected \end{cases}$$

3.Total vote gained by a class

$$D_j = \sum_{i=1}^{N} w_t d_{i,j}$$

4.Identify the class (Attack/Healthy) receiving highest total vote.

5 Results and Discussion

The effectiveness of the proposed FDIA detection scheme is validated using IEEE 14 bus, 39 bus and 57 bus test power systems. The measurement dataset has been generated by simulating the system under varying contingency and attack scenarios using MATPOWER 7.0 [38]. The measurement set (Z_{meas}) is further used for state estimation using WLS algorithm. With the dataset of state vector for different contingency and attack scenarios, the ensemble classifier based FDIA detection scheme has been executed to identify any disturbance in the market operation as a FDIA or contingency.

As outlined in the previous section, for formulating the attack detection task as a binary classification problem, the state variables estimated from the sensor measurements are considered input to the classifier. The state variables for varying operating scenarios of the smart grid involving both healthy and cyber-attack cases have been estimated from the corresponding sensor measurements. Further, the estimated state variables have been used to generate the dataset for training and validating the DT based ensemble classifier. The choice of state variables for detecting any market-oriented cyber-attack is motivated by the fact that any deviation in the state vector will have a significant impact on the electricity market operation and hence, LMP. The increased dependence of market dynamics on the state estimation task allows for proper mapping between the state variables and the operating scenario (Attack/Healthy).

The training dataset for FDIA is generated using the formulations outlined in Sect. 3.3, considering load variation between 0.8 *p.u.* to 1.1 *p.u.* of the rated value. For contingency scenarios, the states are estimated from measurements during a change in the network topology. The effectiveness of the ensemble classifier has been evaluated for detecting manipulated sensor measurements. Further, the detection accuracy of the proposed classifier is compared with the classical standalone Support vector machine (SVM) and DT based classifiers. Support vector machine (SVM) has been widely used for binary classification problems with known labels. However, the SVM necessitates a large amount of data for training to achieve higher accuracy, and its applicability is restricted to datasets with higher noise or data overlapping [39]. For discrete applications, DTs are known to be more effective than SVM [40, 41].

5.1 Classifier Performance Metric

In this sub-section, the effectiveness of the proposed attack detection scheme has been quantified using different performance indices and metrics.

(a) Accuracy The accuracy of the classifier represents the efficacy of the ensemble algorithm in classifying a falsified data from healthy data obtained at the control centre. It is quantified as [42]

$$Accuracy = \frac{TP + TN}{Total\ data}$$

where TP and TN are the true positive and true negative samples, indicating the actual number of attacked and normal data being accurately classified, respectively.

The measurement data set comprising of healthy and attacked measurements are generated for different network configurations. The classification task is performed for the data encompassing both healthy and compromised sensor information. The training and testing dataset are respectively divided into 80% and 20% of the total data. The size of the training and testing data sets are demonstrated in Table 3. Each

Table 3 Size of different training and testing data samples

Type of data						
Test system	Healthy/contingency	Random Cyber	Cyber topology	Total data	Training data	Testing data
IEEE 14 Bus	230	100	70	400	320	80
IEEE 39 Bus	300	120	80	500	400	100
IEEE 57 Bus	350	140	110	600	480	120

Table 4 Training performance of DT based ensemble classifier

Performance measure (%)	IEEE 14 Bus	IEEE 39 Bus	IEEE 57 Bus
Accuracy	98.56	98.80	99.33
False Positive	0.50	0.40	0.16
False Negative	1.00	0.80	0.50

Table 5 Testing performance of DT based ensemble classifier in detecting cyber-attacks

Test system	Cyber-attack detection accuracy (%)		
	Random cyber	Cyber topology	Overall detection accuracy
IEEE 14 Bus	97.50	92.85	95.58
IEEE 39 Bus	97.77	91.66	95.33
IEEE 57 Bus	96.66	90.00	94.00

data sample comprises of estimated system states, i.e. 13, 38 and 56 load angles for IEEE 14 bus, 39 bus and 57 bus system, respectively. The training and testing results are estimated by means of 10-fold cross-validation, where the dataset for each fold are chosen randomly in 80 : 20 ratio. The ensemble classifier is trained for 20 learners with a maximum number of splits confined to 30. The training accuracy of the proposed ensemble classifier is identified to be 98.56%, 98.80%, and 99.33% for IEEE 14 bus, 39 bus, and 57 bus respectively (Table 4). The detection accuracy for the attack cases in the testing dataset is reported in Table 5 along with the performance on the type of attack (random cyber and cyber-topology attack).

To evaluate the impact of training data size on the classification accuracy, the ensemble algorithm has been executed by varying the sample size. The training dataset is divided into subsets comprising 20% of the entire samples. In each step, the classification accuracy has been evaluated by sequentially adding the data subsets into the training dataset. A similar process is carried out for the SVM and DT classifier on the benchmark test systems. The variation in the classification accuracy on the testing dataset with an increase in the size of the training dataset is depicted in Figs. 6, 7 and 8 for IEEE 14, 39, and 57 bus system, respectively. The increase in the classification accuracy with an increase in the number of training samples is because of the improved extent of generalization achieved between the input (state variables) and operating scenarios (healthy/ attack). The higher proportion of the training dataset allows for incorporating more diverse operating scenarios in the mapping performed by the classifier. It can be observed that the effectiveness of the ensemble classifier over DT and SVM is more pronounced even when the data size is less. For a lesser number of training samples, the ensemble approach avoids the limitation of base classifiers of increased biasness toward a particular class. The test results (Table 6) reflect the effectiveness of the proposed ensemble classifier in achieving high classification accuracy for a multi-dimensional dataset. The improvement over

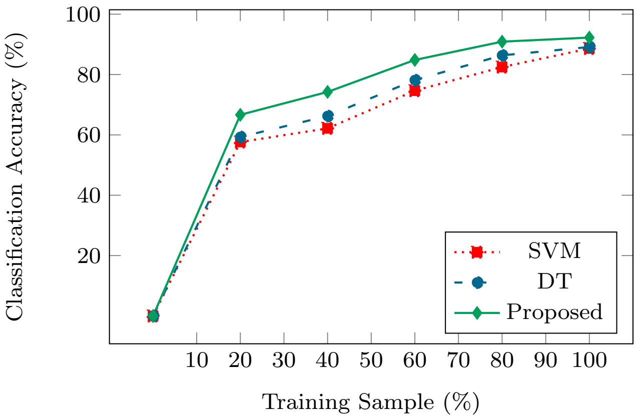


Fig. 6 Accuracy of the classifiers with varying training samples for IEEE 14 bus test system

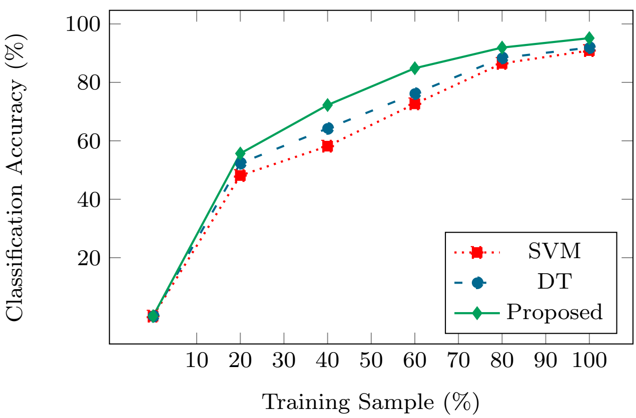


Fig. 7 Accuracy of the classifiers with varying training samples for IEEE 39 bus test system

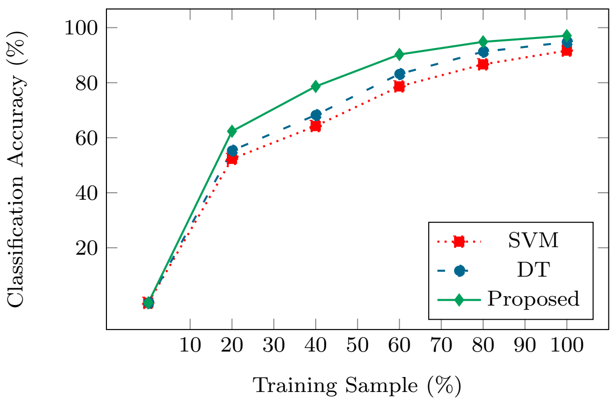


Fig. 8 Accuracy of the classifiers with varying training samples for IEEE 57 bus test system

Table 6 Comparison of performance metric of different classifiers

Classifier				
Test system	Performance measure	SVM	DT	DT based ensemble
IEEE 14 bus	Accuracy	88.63	89.13	**92.25**
	Precision	0.89	**0.90**	**0.90**
	Recall	0.87	0.87	**0.93**
	$F_1 Score$	0.88	0.88	**0.91**
IEEE 39 bus	Accuracy	90.13	91.88	**95.13**
	Precision	0.87	0.89	**0.92**
	Recall	0.89	0.90	**0.95**
	$F_1 Score$	0.88	0.89	**0.94**
IEEE 57 bus	Accuracy	91.16	94.75	**97.08**
	Precision	0.90	0.88	**0.97**
	Recall	0.88	0.86	**0.96**
	$F_1 Score$	0.89	0.87	**0.96**

The metric value for the best performing classifier is represented in bold

standalone DT and SVM classifier is attributed to the proposed scheme's ability to avoid possible over-fitting on the training data.

(b) $F_1 Score$ In addition to accuracy, the performance evaluation of the proposed scheme has been carried out for precision and recall. Precision indicates the effectiveness of the classifier in detecting attacks only. The higher value of precision pertains to a lesser number of false alarms from the detection mechanism, while recall refers to the number of attacks being not detected by the classifier. Using the precision and recall index, the effectiveness in attack detection is quantified by $F_1 Score$ as [39]

$$Precision(P_r) = \frac{TP}{Predicted\ Positive}$$
$$Recall(R_e) = \frac{TP}{Actual\ Positive}$$
$$F_1 Score = 2\frac{P_r R_e}{P_r + R_e}$$

The appropriateness of the proposed classifiers in maintaining integrity in market operation has evaluated and compared with DT and SVM based schemes using the above indices in Table 6. The increase in the classification accuracy and precision with the size of the network is due to the increased volume of the dataset used for training the classifier. The improved performance of the classifier on the testing dataset reflects its generalization ability for cases not learned during the training phase.

(c) Receiver Operating Characteristics (ROC) Receiver operating characteristics (ROC) plot is a vital tool for estimating the performance of a discrete classifier. It plots

the graph between true positive rate (TPR) and false positive rate (FPR) to establish the trade-off between sensitivity and specificity of the classifier. True positive rate (TPR) and false positive rate (FPR) are evaluated as [43]

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{TN + FP}$$

where TPR and FPR indicate the number of attacked samples being correctly detected and false alarms, respectively. The ROC plot for the ensemble scheme demonstrates how efficient the classifier is for detecting the attacked scenarios. The ROC plots for the proposed ensemble DT-based classifier in IEEE 14, 39, and 57 bus test system is shown in Fig. 9. A higher area under the curve of the ROC plot, authenticates the improved performance of the classifier.

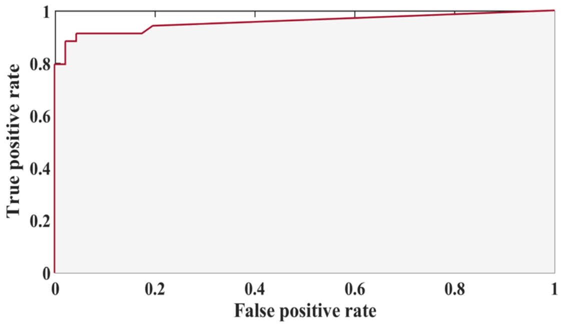
(d) Confidence interval Confidence interval is used to quantify the uncertainty in the response of a classifier. It is estimated by providing bounds to the estimated results (classification accuracy). A lesser interval corresponds to higher precision in the classifier performance. The confidence interval can be estimated as

$$Confidence\ interval = s * \sqrt{\frac{accuracy(1 - accuracy)}{n_s}}$$

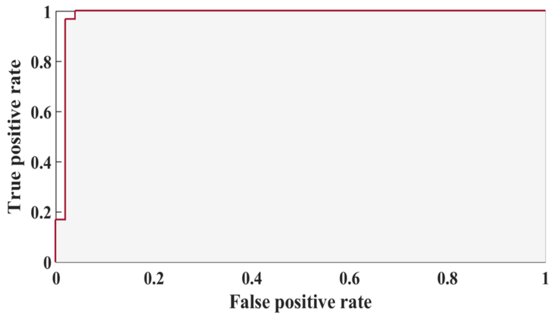
where n_s is the size of sample and s is the number of standard deviations for Gaussian distribution (for 95% confidence interval, $s = 1.96$) [44]. With the proposed scheme, the confidence interval for IEEE 14 bus, 39 bus, and 57 bus test system is estimated to be (0.92±0.05), (0.95±0.04), (0.97± 0.03). The same has been illustrated in form of error bars in Fig. 10.

5.2 Complexity Analysis

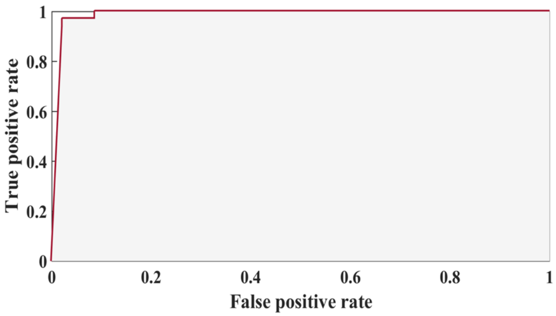
Effective performance of any cybersecurity mechanism necessitates a computationally cheap algorithm to detect any false data attack. The overall complexity of any classifier based attack detection is composed of the computational cost associated with two components i.e., feature extraction and classification. Unlike the reported techniques [29, 39], the proposed scheme involves estimation of features/attributes (state variables) using a non-iterative procedure, thereby avoiding the complexity associated with the feature extraction stage. The use of DT-based ensemble classifier for attack detection and classification allows for achieving high robustness and accuracy with the reduced computational cost compared to other classifiers like artificial neural network (ANN), SVM, and adaptive neuro-fuzzy inference system [34].



(a) IEEE 14 Bus



(b) IEEE 39 Bus



(c) IEEE 57 Bus

Fig. 9 ROC Curve of the proposed ensemble DT based classifier in standard IEEE 14, 39, and 57 bus test systems

6 Conclusion

With the aim of maintaining the integrity of market operation performed by ISO, this paper proposes a scheme for detecting FDIA intended at the disruption of the MMS in the smart grid. The dynamic behaviour of system states and its impact on market operation has been used to identify FDIAs. Detecting the factiously developed nodal price allows the ISO to take the necessary steps so as to avoid market mismanagement. The market behaviour over all possible operating scenarios has been considered

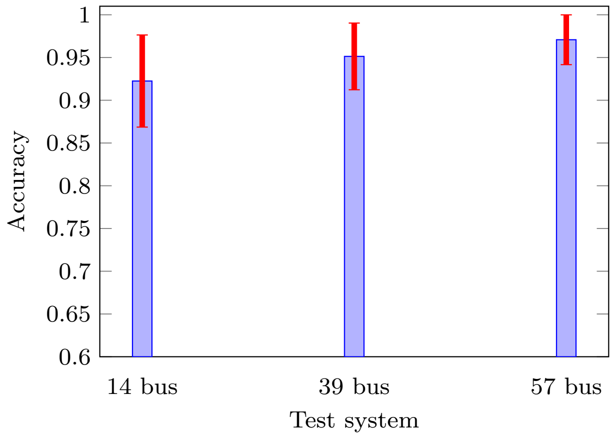


Fig. 10 Error bar plots of the proposed ensemble DT based classifier in standard IEEE 14, 39, and 57 bus test systems

to discriminate the FDIA from actual contingency using an ensemble classification-based approach. The scheme involves continuous monitoring of the market operation by analysing the deviation in the nodal electricity price. For any substantial deviation, the classifier identifies the corresponding scenario as an actual contingency or FDIA. The falsified data generated within the physical constraints are found to impact the market operation, while benefiting a particular market participant (utility/consumer) at the cost of others. The effectiveness of the proposed methodology has been extensively validated for IEEE 14, 39, and 57 bus systems. For different operations of the MMS, the proposed DT based ensemble classifier is found to effectively detect FDIAs of varying type and magnitude. It has been observed that, with the proposed scheme, the attack detection accuracy for IEEE 14, 39, and 57 bus test system is observed to be 92.25%, 95.13%, and 97.08%, respectively. The improvement in the performance over other states of the classifiers (SVM, DT) is found to be more significant for systems of increased size. The uncertainty of response with regard to a false alarm is found to lie under 5% for all the test systems. In the present work, the network scenario for the DAM and RTM is assumed to be similar, and the nodal price deviation between them has been considered as a threshold to activate the classifier. Future work in this direction would include a possible change in network configuration or sensor failure between the execution of DAM and RTM.

Acknowledgements This work was funded by Interdisciplinary Cyber Physical Systems (ICPS) division of Department of Science and Technology, Government of India under Research Grant DST/ICPS/Cluster/CS Research/2018(General).

References

1. Kounev, V., Lévesque, M., Tipper, D., Gomes, T.: Reliable communication networks for smart grid transmission systems. J. Netw. Syst. Manage. **24**, 629–652 (2016). https://doi.org/10.1007/s10922-016-9375-y
2. Ali, F., Bouachir, O., Ozkasap, O., Aloqaily, M.: SynergyChain: Blockchain-assisted Adaptive Cyberphysical P2P Energy Trading. IEEE Trans. Ind. Inform. (2020). https://doi.org/10.1109/TII.2020.3046744
3. de Jesus Martins, R., Knob, L.A.D., da Silva, E.G., Wickboldt, J.A., Schaeffer-Filho, A., Granville, L.Z.: Specialized CSIRT for incident response management in smart grids. J. Netw. Syst. Manag. **27**(1), 269–285 (2019). https://doi.org/10.1007/s10922-018-9458-z
4. Al Ridhawi, I., Otoum, S., Aloqaily, M., Boukerche, A.: Generalizing AI: challenges and opportunities for plug and play AI solutions. IEEE Netw. **35**(1), 372–379 (2021). https://doi.org/10.1109/MNET.011.2000371
5. Khaitan, S.K., McCalley, J.D., Liu, C.C.: Cyber Physical Systems Approach to Smart Electric Power Grid. Springer, New York (2015)
6. Gunduz, M.Z., Das, R.: Cyber-security on smart grid: Threats and potential solutions. Comput. Netw. **169**, 107094 (2020). https://doi.org/10.1016/j.comnet.2019.107094
7. Eder-Neuhauser, P., Zseby, T., Fabini, J., Vormayr, G.: Cyber attack models for smart grid environments. Sustain. Energy Grids Netw. **12**, 10–29 (2017)
8. Liu, Y., Ning, P., Reiter, M.K.: False data injection attacks against state estimation in electric power grids. ACM Trans. Inf. Syst. Security. **14**(1), 1–33 (2011)
9. He, H., Yan, J.: Cyber-physical attacks and defences in the smart grid: a survey. IET Cyber-Phys. Syst. **1**(1), 13–27 (2016)
10. Ott, A.L.: Experience with PJM market operation, system design, and implementation. IEEE Trans. Power Syst. **18**(2), 528–534 (2003)
11. Zheng, T., Litvinov, E.: Ex post pricing in the co-optimized energy and reserve market. IEEE Trans. Power Syst. **21**(4), 1528–1538 (2006)
12. Abur, A., Exposito, A.G.: Power System State Estimation: Theory and Implementation. CRC Press, Boca Raton (2004)
13. Xie, L., Mo, Y., Sinopoli, B.: False data injection attacks in electricity markets. 2010 First IEEE International Conference on Smart Grid Communications. pp. 226–231 (2010).https://doi.org/10.1109/SMARTGRID.2010.5622048.
14. Otoum, S., Kantarci, B., Mouftah, H. T.: A novel ensemble method for advanced intrusion detection in wireless sensor networks. In ICC 2020-2020 IEEE international conference on communications(ICC),2020. 1-6(2020).https://doi.org/10.1109/ICC40277.2020.9149413.
15. Shukla, S.K., Koley, E., Ghosh, S.: DC offset estimation-based fault detection in transmission line during power swing using ensemble of decision tree. IET Sci. Meas. Technol. **13**(2), 212–222 (2018)
16. Li, S., Yılmaz, Y., Wang, X.: Quickest detection of false data injection attack in wide-area smart grids. IEEE Trans. Smart Grid. **6**(6), 2725–2735 (2014)
17. Liu, X., Zhu, P., Zhang, Y., Chen, K.: A collaborative intrusion detection mechanism against false data injection attack in advanced metering infrastructure. IEEE Trans. Smart Grid. **6**(5), 2435–2443 (2015)
18. Liu, L., Esmalifalak, M., Ding, Q., Emesih, V.A., Han, Z.: Detecting false data injection attacks on power grid by sparse optimization. IEEE Trans. Smart Grid. **5**(2), 612–621 (2014)
19. Luo, X., Wang, X., Pan, X., Guan, X.: Detection and isolation of false data injection attack for smart grids via unknown input observers. IET Gener. Transm. Distrib. **13**(8), 1277–1286 (2019)
20. Li, B., Ding, T., Huang, C., Zhao, J., Yang, Y., Chen, Y.: Detecting false data injection attacks against power system state estimation with fast go-decomposition approach. IEEE Trans. Industr. Inf. **15**(5), 2892–2904 (2019)
21. Chen, P.Y., Yang, S., McCann, J.A., Lin, J., Yang, X.: Detection of false data injection attacks in smart-grid systems. IEEE Commun. Mag. **53**(2), 206–213 (2015)
22. Chaojun, G., Jirutitijaroen, P., Motani, M.: Detecting false data injection attacks in ac state estimation. IEEE Trans. Smart Grid. **6**(5), 2476–2483 (2015)
23. Ferrag, M.A., Maglaras, L., Moschoyiannis, S., Janicke, H.: Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. J. Inf. Secur. Appl. **50**, 102419 (2020)

24. Wang, D., Wang, X., Zhang, Y., Jin, L.: Detection of power grid disturbances and cyber-attacks based on machine learning. J. Inf. Secur. Appl. **46**, 42–52 (2019)
25. Khanna, K., Panigrahi, B.K., Joshi, A.: AI-based approach to identify compromised meters in data integrity attacks on smart grid. IET Gener. Transm. Distrib. **12**(5), 1052–1066 (2018)
26. Cui, M., Wang, J., Yue, M.: Machine learning-based anomaly detection for load forecasting under cyberattacks. IEEE Trans. Smart Grid **10**(5), 5724–5734 (2019)
27. Otoum, S., Kantarci, B., Mouftah, H. T.: Mitigating False Negative intruder decisions in WSN-based Smart Grid monitoring. In 2017 13th International wireless communications and mobile computing conference (IWCMC). pp. 153–158 (2017). https://doi.org/10.1109/IWCMC.2017.7986278.
28. Wang, Q., Cai, X., Tang, Y., Ni, M.: Methods of cyber-attack identification for power systems based on bilateral cyber-physical information. Int. J. Electr. Power Energy Syst. **125**, 106515 (2020)
29. Gunturi, S.K., Sarkar, D.: Ensemble machine learning models for the detection of energy theft. Electr. Power Syst. Res. **192**, 106904 (2020). https://doi.org/10.1016/j.epsr.2020.106904
30. Mukherjee, R., De, A.: Development of an ensemble decision tree-based power system dynamic security state predictor. IEEE Syst. J. **14**(3), 3836–3843 (2020)
31. Liang, G., Zhao, J., Luo, F., Weller, S.R., Dong, Z.Y.: A review of false data injection attacks against modern power systems. IEEE Trans. Smart Grid **8**(4), 1630–1638 (2016)
32. Deng, R., Xiao, G., Lu, R., Liang, H., Vasilakos, A.V.: False data injection on state estimation in power systems-Attacks, impacts, and defense: A survey. IEEE Trans. Industr. Inf. **13**(2), 411–423 (2017)
33. Liang, G., Weller, S.R., Zhao, J., Luo, F., Dong, Z.Y.: A framework for cyber-topology attacks: Line-switching and new attack scenarios. IEEE Trans. Smart Grid **10**(2), 1704–1712 (2019)
34. Polikar, R.: Ensemble based systems in decision making. IEEE Circuits Syst. Mag. **6**(3), 21–45 (2006)
35. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
36. Manohar, M., Koley, E., Ghosh, S.: Reliable protection scheme for PV integrated microgrid using an ensemble classifier approach with real-time validation. IET Sci. Meas. Technol. **12**(2), 200–208 (2018)
37. González, C., Mira-McWilliams, J., Juárez, I.: Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests. IET Gener. Transm. Distrib. **9**(11), 1120–1128 (2015)
38. Zimmerman, R. D., Murillo-Sanchez, C. E.: MATPOWER (Version 7.0) [Software]. https://matpower.org (2019)
39. Ahmed, S., Lee, Y., Hyun, S.H., Koo, I.: Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest. IEEE Trans. Inf. Forensics Secur. **14**(10), 2765–2777 (2019)
40. Kamwa, I., Samantaray, S.R., Joos, G.: Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices. IEEE Trans. Smart Grid **1**(2), 144–158 (2010). https://doi.org/10.1109/TSG.2010.2052935
41. Dubey, R., Samantaray, S.R., Panigrahi, B.K., Venkoparao, V.G.: Data-mining model based adaptive protection scheme to enhance distance relay performance during power swing. Int. J. Electr. Power Energy Syst. **81**, 361–370 (2016)
42. Aloqaily, M., Otoum, S., Al Ridhawi, I., Jararweh, Y.: An intrusion detection system for connected vehicles in smart cities. Ad Hoc Netw. (2019). https://doi.org/10.1016/j.adhoc.2019.02.001
43. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
44. Mitchell, T.M.: Machine Learning. McGraw-Hill, Germany (1997)

Prasanta Kumar Jena received the B.Tech. degree in electrical and electronics engineering from B.P.U.T.University, Odisha, India, in 2008 and the M.Tech. degree in Power Electronics and Drives from KIIT University, Odisha, India, in 2014. He is currently pursuing the Ph.D. degree in Electrical

Engineering Department, NIT Raipur, India. His current research interests include cyber-physical security in smart grid and optimization techniques.

Subhojit Ghosh received the Ph.D degree from IIT Kharagpur in 2010. He has more than 15 years of teaching experience at BIT Mesra, NIT Rourkela and NIT Raipur. Presently he is working as Associate Professor in Electrical Engineering Department, NIT Raipur. His research interests include optimization techniques, control systems, renewable energy and power system protection.

Ebha Koley received the Ph.D degree from National Institute of Technology (NIT) Raipur in 2015, where she is presently working as Assistant Professor in Electrical Engineering Department. She has more than 10 years of industrial/teaching experience at Jindal Steel and Power Limited (JSPL), Raigarh and NIT Raipur. Her research interests include power system protection, microgrid and soft computing.

Murli Manohar received the B.E. (Electrical and Electronics Engineering) and M.E. (Power Electronics) degree in 2010 and 2015 respectively from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P), India. He is currently pursuing Ph.D. from National Institute of Technology (NIT), Raipur. His current research interests include microgrid protection, application of soft computing and data mining techniques in power system protection.

Authors and Affiliations

Prasanta Kumar Jena[1] · Subhojit Ghosh[1] · Ebha Koley[1] · Murli Manohar[1]

Prasanta Kumar Jena
pkj.ped@gmail.com

Ebha Koley
ekoley.ele@nitrr.ac.in

Murli Manohar
murlimanohar2311@gmail.com

[1] Department of Electrical Engineering, National Institute of Technology Raipur, Chhattisgarh 492010, India