Ensuring Cybersecurity of Smart Grid against Data Integrity Attacks under Concept Drift

Mostafa Mohammadpourfard [a,*], Yang Weng [b], Mykola Pechenizkiy [c], Mohsen Tajdinian [d], Behnam Mohammadi-Ivatloo [e]

[a] Department of Electrical and Computer Engineering, Sahand University of Technology, Tabriz, Iran (e-mail: mohammadpourfard@sut.ac.ir). (*Corresponding author)
[b] School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, USA (e-mail: yang.weng@asu.edu).
[c] Department of Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands (e-mail: m.pechenizkiy@tue.nl).
[d] School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran (e-mail: tajdinian.m@shirazu.ac.ir).
[e] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran (e-mail: bmohammadi@tabrizu.ac.ir).

Abstract: For achieving increasing artificial intelligence in future smart grids, a highly accurate state estimation (SE) is needed as a prerequisite for many other key functionalities for successful monitoring and control. With increasing interconnection of utility network and internet, traditional state estimators are vulnerable to complex data integrity attacks, such as false data injection (FDI), bypassing existing bad data detection (BDD) schemes. While researchers propose detectors for FDI, such countermeasures neglect power state changes due to contingencies. As such an abrupt physical change negatively affects existing FDI detectors, they will provide incorrect classification of the new instances. To resolve the problem, we conducted analysis for a fundamental understanding of the differences between a physical grid change and data manipulation change. We use outage as an example and propose to analyze historical data followed by concept drift, focusing on distribution change. The key is to find critical lines to narrow down the scope. Techniques such as dimensionality reduction and statistical hypothesis testing are employed. The proposed method is tested on IEEE 14 bus system using load data from the New York independent system operator with two different attack scenarios: 1) attacks without concept drift, 2) attacks under concept drift. Numerical results show that the new method significantly increases the accuracy of the existing detection methods under concept drift.

Keywords: Smart grid, data integrity attacks, line outage, machine learning.

1 INTRODUCTION

The future smart grid is highly dependent on communication networks to facilitate power system operation and control. However, this strong reliance makes the grid vulnerable to a wide variety of malicious attacks which reduce the reliability of smart grids and cause severe operational failures and substantial financial loss [1]–[3]. For example, in the December 2015 Ukrainian electrical grid cyber-attack, the adversary opened several circuit breakers and caused a power outage of approximately 225,000 customers [4]. Among possible cyber-attacks, false data injection (FDI) [5] is one of the most critical ones which makes SE results inaccurate [6]. Wrong estimates could mislead system operators to take incorrect actions, endangering the security of the power system with catastrophic consequences such as blackouts [7]. This is because the provided network status information through state estimation is used in other functions of energy management systems, such as transmission stability analysis, load shedding, etc [8]. Specificially, adversaries can launch such an attack by altering readings of multiple sensors and Phasor Measurement Units (PMUs) to introduce malicious measurements and finally inject arbitrary errors to state estimates without being detected by BDD methods [9], [10].

1.1 Related Work

To safeguard the system operation and control against FDI attacks, several detectors and mitigation methods have been developed. These countermeasures can be divided into two categories: 1) protection-based; and 2) detection-based. The protection-based approaches try to alleviate FDI attacks through identifying and protecting critical meters [11], [12]. Drawbacks of these approaches include high implementation cost for large-scale power systems, the drop of measurements redundancy and the unassured effectiveness of the protection all of the time [13].

The detection-based methods rely on anomaly detection techniques to recognize maliciously altered measurements that do not fit the distribution of historical measurements utilizing graph theory, Kalman filter, classification algorithms, statistical threshold testing [13]–[17], etc. Specifically, the second solution type estimates the underlying data distribution of the historical data and use the estimated distribution as a reference to capture future attacks that will lead to extreme deviation from the reference distribution. For a detailed review, a qualitative comparison between relevant literature and the proposed method, 4 properties are studied for each method as shown in Table 1. These properties are:

1) Dealing with contingencies (DC): This property indicates the ability of methods in detecting attacks when the system is under a line outage contingency.
2) Attack localization (AL): This property indicates whether the proposed method can determine the location of attacks or not.
3) Applicable to large-scale power systems (ALSPS): This property indicates whether implementing the method in large-scale power systems is computationally complex and costly or not.
4) Requiring external devices (RED): This property indicates whether the method, to protect the system or to detect the attacks, relies on external devices such as secure PMUs, etc. or not. This reliance is a drawback since compromise of the device will negatively affect the method.

Table 1: Summary of related work

Approach	References	Main idea	DC	AL	ALSPS	RED
Protection	[11], [12], [19], [20]	Defending against the FDI attacks by determining the minimal set of measurements that needs to be protected and ensuring the security of them by encryption, tamper-proof communication systems, etc. For example, reference [11] uses greedy approaches to select a subset of measurements and to promote the secure PMUs deployment to defend against such attacks.	No	No	No	Yes
Detection	[21]	A sequential detector based on the generalized likelihood ratio is developed to detect FDI attacks.	No	No	Yes	No
Detection	[6], [16], [17], [22], [23]	Supervised learning algorithms such as k-Nearest Neighbor, Support Vector Machines, reservoir computing (multilayer perceptron + delayed feedback networks) are used to predict class label of new observations using historical data as training set.	No	Yes	Yes	No
Detection	[1]	A robust detector is proposed by checking the measurement statistical consistency using a subset of secure PMU measurements. Specifically, they derive the Neyman-Pearson detector for an FDI detection.	No	No	Yes	To some extent
Detection	[24], [25]	A discrepancy between the calculated Markov graph of the bus phase angles and the power grid graph can lead to anomaly detection. To find the attacked nodes, the approaches are dependent on selecting a threshold based on historical data.	No	Yes	Yes	No
Detection	[13]	A Kullback-Leibler distance (KLD) threshold is set using historical data to detect FDI attacks since they believe the KLD of the attacked measurement will be larger than the normal one. Proper selection of this threshold affects the accuracy of detection.	No	Yes	Yes	No
Detection	Proposed Method	Serves as a complementary tool for the existing detectors by ensuring their robustness against concept drift.	Yes	Yes	Yes	No

As shown in Table 1, one common drawback of the existing FDI attack detection approaches is reliance on the assumption of a time-invariant historical data and the assumption of static discovered patterns, which makes them more suitable for stationary data. This means they are designed for a specific system configuration and have not considered the impact of topology reconfigurations. However, in practice, the data tends to change over time and the underlying distribution of data are not stationary. In other words, they have been developed based on the training set which is obtained from the system without topology changes. This is while the concept of data (underlying distribution of unseen data) could unpredictably drift after a line outage contingency as shown in Fig. 1. As a result, predictions made by the models developed/trained on such historical data will be no longer accurate since the distribution of data changes and old observations became irrelevant to the new ones.

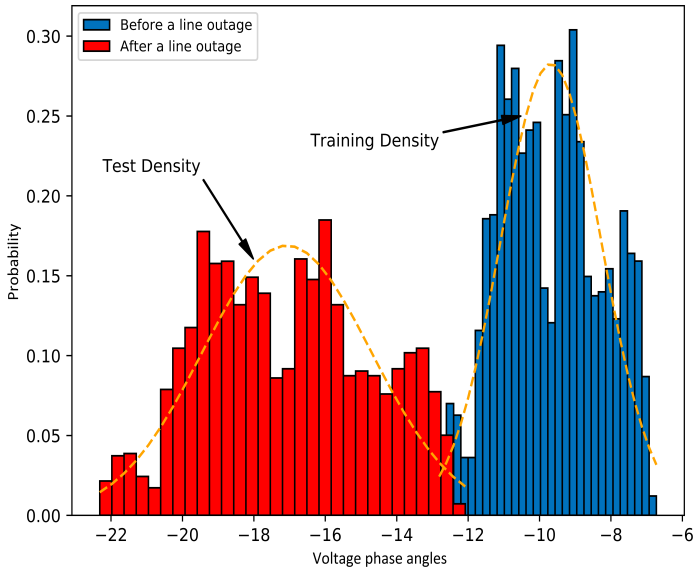

Fig. 1: Data distribution of a bus phase angles before and after a line outage

1.2 Concept Drift

In machine learning, unpredictable shifts in the underlying distribution of historical data over time are referred to as concept drift [18]. We apply the following framework in order to examine the issue of learning under concept drift. Specifically, let $\mathbf{S}_i$ be a p-dimensional measurements at time point i that is labeled with $y_i \in \{normal, attack\}$. The pair $(\mathbf{S}_i, y_i)$ is called an labeled instance. $\mathbf{TS} = \{\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_t\}$ is referred to as historical or training data instance and $\mathbf{S}_{t+1}$ as the target or testing instance. The goal of the machine learning-based detector is to predict the label of y_{t+1} the new example $\mathbf{S}_{t+1}$. To this end, a Lt learner is created using all or a subset of the available labeled historical data **TS**. Afterwards, L_t is applied to predict the label for $\mathbf{S}_{t+1}$. Each $\mathbf{S}_i$ is generated by a source $Source_i$. Source is defined as a distribution (probability density function) over data. As mentioned, current developed FDI attack detectors presume that the training and the future test data come from the same distribution. In other words, they believe that all data are taken from the same source, i.e. $Source_1 = Source_2 = \cdots = Source_{t+1} =$ **Source** and the concept is stable. However, the underlying data distribution can change dynamically over time due to topology changes as shown in Fig. 1. Therefore, at any stage the testing data may come from a different distribution as compared to the training data. For any two time points i and j, if $Source_i \neq Source_j$, it is stated that concept drift exists. It is worth noting that random noise is not known to be a form of drift since the data source is still the same. In this paper, the "concept" refers to the distribution of measurements, and the term "concept drift" indicates that the distribution of measurements is significantly changed. Hence, We assume that the source of the future data instance $\mathbf{S}_{t+1}$ is unknown and our main focus is on concept drift stems from line outage contingencies.

1.3 Contributions

To address this issue, we propose a paradigm to enable the existing attack detection approaches to handle concept drift stem from line outage contingencies. The key idea of the proposed method is to find the critical line outage contingencies leading to significant change in the underlying data distribution and including their historical data in the training phase of the attack detection method. Critical branch outage contingency set is constructed by comparing the probability distributions of the reduced drifted historical data with the reduced historical data of the network when there is no contingency. It is shown through simulations that the proposed method significantly increases the accuracy of the existing methods under concept drift. The contributions of this paper are as follow:

1) We thoroughly investigate the effect of concept drift on smart grid cybersecurity, i.e., show what are the differences between a physical grid change and data manipulation change and how drifting from the baseline network topology due to a line outage can affect the attack detection methods.
2) A detailed analysis of the behavior of a supervised learning technique in identifying attacks under concept drift is conducted.
3) The proposed method makes the existing FDI detectors robust to physical changes and works by identifying the critical line outage contingencies leading to significant data distribution change. Furthermore, the computed critical line set by the proposed method is compared with those constructed by power system indices.

The rest of the paper is as follows. In Section 2, we discuss the attack model. Section 3 explains the proposed method. Section 4 shows the test results of different attacking scenarios. Section 5 concludes the paper.

2 Problem Formulation

Continuous and uninterrupted operation of power systems requires an accurate knowledge of the system states to correctly apply active controls. SE processes the gathered measurement data from remote terminal units (RTUs) and estimates the system status using DC state estimation or AC state estimation. In this paper, we adopt the widely used DC power flow model [2], [3], [6]–[9], [16], [17], [24], [25] for power system state estimation. For smart grid under consideration, it is assumed that the underlying transmission grid consists of a set $\mathcal{N} = \{1, 2, \cdots, n\}$ of buses, a set $\mathcal{M} = \{1, 2, \cdots, m\}$ of meters and a set $\mathcal{B} = \{1, 2, \cdots, b\}$ of transmission branches. State estimation is used to estimate state variables $\mathbf{x} = [\theta_1, \theta_2, \cdots, \theta_n]^T$ based on m installed meter measurements $\mathbf{z} = [z_1, z_2, \cdots, z_m]^T$ under independent random measurement noises $\mathbf{e} \in \mathbb{R}^{m\times 1}$, which are assumed to follow Gaussian distribution with zero mean and diagonal matrix representing the covariance matrix of the measurement errors $\mathbf{R}$ [26].

The n state variables are phase angles at all the buses except the slack bus where the phase angle is set to be zero (bus 1 in this paper, i.e., $\theta_1 = 0$). The installed meters measure buses active power injections and branch active power flows. Given the DC power flow model, the relationship between meter measurements $\mathbf{z}$ and state variables $\mathbf{x}$ is:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the linear measurement function. Based on the weighted least squares (WLS) approach, the estimated system state $\widehat{\mathbf{x}}$ can be calculated as:

$$\widehat{\mathbf{x}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\ \mathbf{z}. \tag{2}$$

After estimation, Euclidean norm of measurement residual $\rho = \|\mathbf{z} - \mathbf{H}\widehat{\mathbf{x}}\|_2$, is used to identify bad measurements [27]. This is achieved by comparing ρ with a predetermined threshold τ, and bad measurements are assumed to exist if $\rho > \tau$, otherwise the measurement set is taken as a normal one. However, this assumption is breached through a newly introduced attack called FDI [5]. In this research, it is proved that if the attacker has knowledge of the system structure $\mathbf{H}$ and can manipulate multiple meter measurements at the same time, can pass the bad data detection test. Let $\mathbf{a} \in \mathbb{R}^{m\times 1}$ denotes the cyber attack vector. Then, the manipulated measurement with the malicious data is given by $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$ and could bypass the BDD under the condition $\mathbf{a} = \mathbf{H}\mathbf{c}$, where $\mathbf{c} \in \mathbb{R}^{n\times 1}$ is the maliciously injected error on the system state. Sending $\mathbf{z}_a$ to the state estimator will result in false estimates $\widehat{\mathbf{x}}_a = \widehat{\mathbf{x}} + \mathbf{c}$ where $\widehat{\mathbf{x}}$ is the true estimates of the system. The attacked measurement $\mathbf{z}_a$ will bypass BDD since the measurement residual of $\mathbf{z}_a$ will not lead to a change in the measurement residual ρ and is the same as that of $\mathbf{z}$:

$$\rho_a = \|\mathbf{z}_a - \mathbf{H}\widehat{\mathbf{x}_a}\|_2 = \|\mathbf{z} + \mathbf{a} - \mathbf{H}(\widehat{\mathbf{x}} + \mathbf{c})\|_2 = \|\mathbf{z} - \mathbf{H}\widehat{\mathbf{x}} + (\mathbf{a} - \mathbf{H}\mathbf{c})\|_2 = \|\mathbf{z} - \mathbf{H}\widehat{\mathbf{x}}\|_2 = \rho. \tag{3}$$

3 Methodology

This section proposes a methodology to handle concept drift due to branch outage contingencies. The proposed approach would serve as a complementary tool to the existing detection methods by enabling them to detect stealthy attacks after concept drift, which could otherwise go undetected. Focus of this paper is on classification which is a supervised learning task. More specifically, the k-Nearest Neighbor (k-NN) [28] is used to detect FDI attacks and show the effectiveness of the proposed method since it has better performance in managing data distribution changes compared to other algorithms due to its lazy nature [23]. The overview of the proposed method is illustrated in Fig. 2. In this paper, the system topology without any contingency is considered as a baseline concept. We further define the network under a branch outage contingency as a new concept. In particular, the system topology could drift from the baseline to concept C_i, where $i = 1, \cdots, Q$. Concept C_i differs from the baseline in the network topology and there exists a line outage contingency in the system. Different algorithms have been used in the proposed method. First in this section, these algorithms are briefly described. Afterward, the proposed approach is presented.

3.1 Employed Algorithms

3.1.1 Principal Component Analysis (PCA)

In addition to complex dependence, high-dimensionality is another challenge for modeling stochastic data. Dimensionality reduction methods can make the transformation from the original high-dimensional space to the suitable lower dimensional space while maintaining the main characteristics of the original data set. PCA is a statistical method to transform a dataset with original features to a dataset of linearly uncorrelated features using an orthogonal transformation. The number of linearly uncorrelated features (i.e. principal components) is usually less than the number of original features. This transformation has a property that the variance of original data in principal components (PCs) is decreasing. In other words, most information about the original dataset is preserved by the first PC and the least information is preserved by the last [29]. Therefore, PCA can be used to reduce the dimension of the original dataset

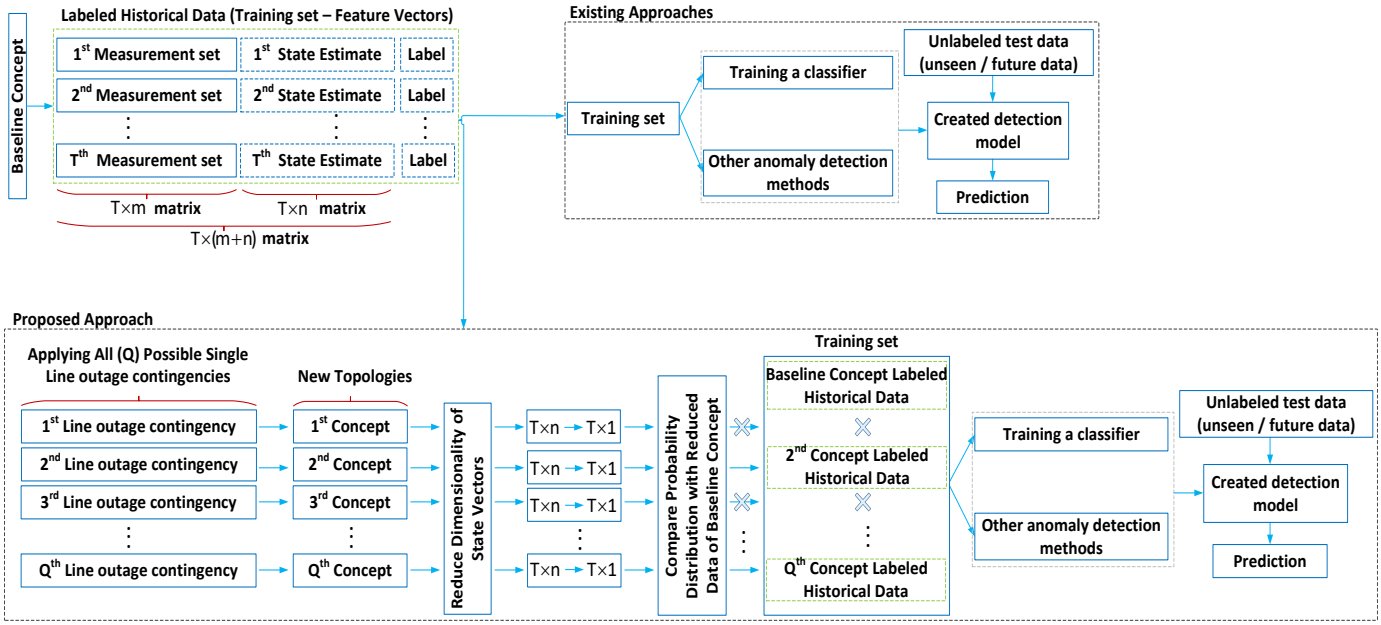

Fig. 2: Diagram of the proposed method

while preserving the most important information. The steps of PCA are displayed in Algorithm 1 (aassuming that $z^{(m,n)}$ is the original dataset where m is the number of samples and n is the number of features).

Algorithm 1 PCA Steps that PCA follows to reduce dimensions of a dataset

1: **Input** : Original dataset $z^{(m,n)}$; m = number of samples, n = number of features.
2: Calculating mean of each dimension:
$\mu^n = mean(z^{(m,n)})$;
3: Calculating covariance matrix:
$\sum = \frac{(z-\mu)^T(z-\mu)}{(n-1)}$, where μ is a $m \times n$ rows matrix which vector μ^m is repeated in each row.
4: Eigenvalues and eigenvectors of $\sum$ are computed and sorted according to deceasing eigenvalue. PCs are eigenvectors of the covariance matrix.
5: The first KP eigenvectors are selected as the KP principal components ($KP < n$).
6: **Output**: Principal Components (reduced dataset)

3.1.2 *Two-sample Kolmogorov-Smirnov Test (KS test)*

KS test [30] is a non-parametric test used to compare the data distributions of the values in two data in vectors $\mathbf{S1} = \{s1_1, s1_2, \cdots, s1_n\}$ and $\mathbf{S2} = \{s2_1, s2_2, \cdots, s2_m\}$. The null hypothesis ($H_0$) is that both data vectors belong to the same distribution of data; Otherwise, they are coming from different distributions. The KS test distance is defined as the maximum absolute difference between cumulative distribution functions of the distributions of the two sample data vectors **S1** and **S2**. Assuming that a sample $\mathbf{Y} = y_1, y_2, \cdots, y_m$ is given, its cumulative distribution function can be defined as follows:

$$F_{Y,m}(y) = \frac{\#i : y_i \leq y}{m}, \tag{4}$$

where $\#i : y_i \leq y$ denotes the number of elements in the set satisfying the property $y_i \leq y$, for all possible i. The KS test statistic for two sets with n and m samples is:

$$D_{n,m} = \sup_{-\infty<s<\infty} |F_{S1,n}(s) - F_{S2,m}(s)|, \tag{5}$$

where $F_{S1,n}$ and $F_{S2,m}$ are the empirical cumulative distribution functions of the first and the second sample respectively and $\sup_s$ is the supremum of the set of distances. The null hypothesis that the two sets of samples are from the same distribution, is rejected at the significance level α if the following necessary condition is satisfied:

$$D_{n,m} > c(\alpha) \times \sqrt{\frac{n+m}{n \times m}}, \tag{6}$$

where $c(\alpha) = \sqrt{-\frac{1}{2} \times \ln \alpha}$ and n and m are the sizes of first and second sample respectively. The significance level is usually set to 5%, 1% or 0.1% [31]. Lower values give more confident decisions in accepting/rejecting the null hypothesis [32]. Thus, 0.1% is selected as α in this paper.

3.1.3 k-NN

Let L be a set of class labels (normal versus tampered) and let $\mathbf{S}_i$ be a p-dimensional data instance of a training set $\mathbf{TS} = \{\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_t\}$ at time point i that is labeled with some $l \in L$. More specifically, the attack detection problem using machine learning algorithms can be defined as a binary classification problem:

$$l_i = \begin{cases} 0, & \text{if } \mathbf{c} = 0 \\ 1, & \text{if } \mathbf{c} \neq 1 \end{cases}, \tag{7}$$

where $l_i = 0$ means that there is no attack and $l_i = 1$ means the ith measurement is manipulated.

k-NN is a classification strategy in which the decision to classify the sample is made on the basis of the main class of the K-next data points of the sample. No training phase is needed as no classification model is necessary for the classification process to be carried out. In fact, the calculations and decisions for the test sample $\mathbf{S}_{t+1}$ are made during the classification process. The decision is taken by comparing the test sample to all the samples collected in order to determine that k-samples are nearest to the test sample. In other words, the k-NN classifier aims to predict the label of a new instance $\mathbf{S}_{t+1}$ according to the labels of a predefined (k) number of the training set closest in distance to $\mathbf{S}_{t+1}$. The commonly used dissimilarity measures is Euclidean distances which is defined as the 2-norm of the displacement vector between two instances as follows [28]:

$$d(\mathbf{S}_1, \mathbf{S}_2) = \mathbf{S}_1 - \mathbf{S}_{22} = \sqrt{\sum_{j=1}^{p} (\mathbf{S}_{1(j)} - \mathbf{S}_{2(j)})^2}. \tag{8}$$

To predict the class label of the new instance $\mathbf{S}_{t+1}$, the set of its k-NNs, $\aleph(\mathbf{S}_{t+1}) = \{\mathbf{S}_{i(1)}, \mathbf{S}_{i(2)}, \cdots, \mathbf{S}_{i(k)}\}$, is constructed by computing the Euclidean distances between this sample and all instances in the **TS**, where $\mathbf{S}_i \in \mathbf{TS}$ and $i(1), i(2), \cdots, i(t)$ are defined as follows:

$$d(\mathbf{S}_{t+1}, \mathbf{S}_{i(1)}) \leq d(\mathbf{S}_{t+1}, \mathbf{S}_{i(2)}) \leq \cdots \leq d(\mathbf{S}_{t+1}, \mathbf{S}_{i(t)}), \tag{9}$$

where t is the number of training samples. Then, the class label of the majority of the neighboring instances is assigned as the class label of the $\mathbf{S}_{t+1}$. Due to its simplicity, k-NN does not need to make any assumptions about the distribution of datasets. The best choice of k depends on the data; in general, higher values of k reduce the impact of noise on classification, but make the boundary between classes less distinct. In this paper, k is set to 3 which is calculated by searching $k \in \{1, 2, \cdots, \sqrt{t}\}$ using leave-one-out cross-validation [33].

3.2 Proposed Method

As discussed in the previous section, the current methods are non-robust to topology changes and will incorrectly label the samples (normal and attack ones) after a contingency. To enable robustness for the existing methods, we propose to use not only the historical data (measurement set and state vectors) of the baseline concept, but also historical data of the critical concepts stem from the line outage contingencies leading to significant change of data distribution. Contingency analysis is a powerful tool for transmission power system (TPS) which is performed to evaluate the outage events in TPS and it is a critical part in security assessment [34]. The objective of the existing line outage contingency selection methods is identifying the contingencies which may lead to unreliability [35]. The majority of the contingency selection methods are based on the evolution of some Performance Index (PI) derived from DC or fast decoupled load flow solution for each contingency [36]. This is while the objective of the proposed method is to identify line outages changing the underlying data distribution dramatically.

Finding critical concept set is important since drifting to those concepts is expected to change the data distribution significantly making a classifier learned before the drift unable to label the incoming samples after a drift correctly. The proposed method chooses critical concepts based on the shift in the distribution of the underlying data. To find such concepts using the proposed method, all concepts to which the network topology that drift from the baseline due to branch failure contingency have been specified except that the load-flow solution is unobtainable and the system is diverged. To this end, we first generate data of all concepts $C_{i=1,2,\cdots,Q}$ by applying all of the possible line outage contingencies one by one. This means we remove a line and repeatedly run the power flow and obtain the sequence of normal system state vectors of that concept $\mathbf{xm} = \{\mathbf{x}_1; \mathbf{x}_2; \cdots; \mathbf{x}_w; \cdots; \mathbf{x}_T\}$, where w is the time index and T is the total number of the collected data points.

To be able to run KS test and to reduce the computation complexities, we use PCA to reduce the dimensions of system state vectors and construct a new vector space $1-$D ($\mathbf{xm} \in \mathbb{R}^{T\times n} \rightarrow \mathbf{xm} \in \mathbb{R}^{T\times 1}$). Our tests show that the PCA not only maintains most of the variability of the original data but also maximizes the separation between the normal and anomalous operation points which helps increase the attack detection rate. Afterward, the probability distribution of the reduced system state vectors of the derived concepts are compared with the reduced data of the baseline concept using KS test. This is because we assume that some line outages will lead to significant data distribution change and the reduced **xm** of the new concept will have a different distribution than the baseline concept one. If the KS test rejects the null hypothesis that the two underlying one-dimensional probability distributions are the same, that concept will be considered as a critical concept.

Finally, we create a training set which comprises historical data of the baseline concept and the identified critical concepts. Now, this training set could be used to develop a robust detection method. As one will see in the simulation result, the proposed paradigm ensure the robustness of current approaches against concept drift and increases their accuracy significantly. This is because we update the **TS** with samples of identified critical concepts and then train the forecast model on the basis of newly updated data.

3.3 Power System Indices for Contingency Ranking

As one will see in the simulation result, to show the effectiveness of the proposed method, the computed critical line set by the proposed method is compared with those constructed by power system indices. Reference [35] ranks the contingencies and finds the severest contingencies by a newly proposed index called root sum square index (RSSI). The higher the RSSI is, the closer the system is to instability. Reference [37] find the critical contingency set by calculating fast voltage stability index (FVSI). FVSI value 1.0 indicates that the particular line is in its instability point which may lead to voltage collapse in the entire network. Reference [38] utilizes a conventional linear sensitivity method and use load power margins as an index to compute the most severe branch outage contingencies.

3.4 Performance Evaluation

F-measure and false positive rate are used to evaluate the performance of the proposed method. F-measure is defined as follows [28]:

$$FM = (\frac{2 \times P_r \times R_e}{P_r + R_e}), \tag{10}$$

where P_r is the precision, R_e is the recall and are computed as follows:

$$P_r = (\frac{TP}{TP + FP}) \qquad R_e = (\frac{TP}{TP + FN}), \tag{11}$$

where true positive (TP) is the number of attack samples correctly detected and localized, false positive (FP) is the number of incorrectly detected and localized attacks, true negative (TN) is the number of correctly rejected normal samples, and false negative (FN) is the number of missed attacks. FM $= 1$ indicates that each sample labeled normal is actually a normal instance, and each measurement classified as an attack is actually a manipulated one.

4 Numerical Results

The simulations are implemented on the IEEE 14 bus system. To highlight the effectiveness of the proposed paradigm in managing concept drift originated from the line outage contingencies, two evaluation scenarios have been designed. Scenario S1 focuses on the baseline concept where FDI attacks are conducted when the system is not affected by contingencies. Scenario S2 is about classifying the attacks when the network topology drifts from the baseline to concept $C_{i=1,2,\cdots,Q}$. As we already discussed, concept C_i is the network under a line outage contingency. These concepts are shown in Fig. 3. For IEEE 14 bus system, DC load flow converged for 19 line outage contingencies ($Q = 19$).

4.1 Data Preparation

The historical data have been preprocessed by MATPOWER [39]. The proposed method is implemented using MATLAB 2017 and Python 3.

Real-world Load Data: To simulate the power system behavior in a more realistic pattern, real-world load data has been integrated into the MATPOWER framework. The load data from New York independent system operator (NYISO) [40] are adopted as the real power profile in the subsequent simulations. NYISO contains online load flow profiles for 11 regions recorded every five minutes. This implies that there are about 288 values for each day. The load data used in this paper is for the first week of January 2016 (January 1, 2016 to January 7, 2016). This means $T = 2045$ load values are obtained as normal samples for each region.

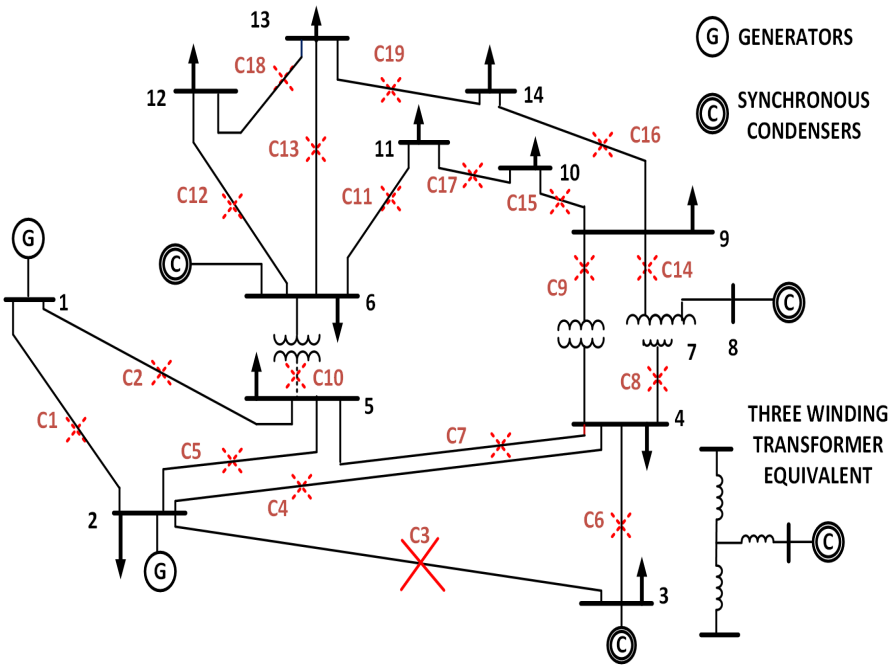

Fig. 3: IEEE 14 bus test system.

Table 2: NYISO load data characteristics

Region	Bus	Range (MW)	Mean(MW)	SD (MW)
CAPITL	Bus2	[11.76-21.70]	16.68	2.38 (14.29%)
CENTRL	Bus3	[51.23-94.20]	72.85	9.54 (13.10%)
DUNWOD	Bus4	[25.70-47.80]	35.35	5.11 (14.46%)
GENESE	Bus5	[4.14-7.60]	5.76	0.81 (14.18%)
HUD VL	Bus6	[5.44-11.20]	8.26	1.27 (15.40%)
LONGIL	Bus9	[15.39-29.50]	21.41	3.45 (16.15%)
MHK VL	Bus10	[4.60-9.00]	6.87	1.11 (16.22%)
MILLWD	Bus11	[1.68-3.50]	2.48	0.35 (14.12%)
N.Y.C.	Bus12	[3.54-6.10]	4.71	0.72 (15.32%)
NORTH	Bus13	[9.12-13.50]	11.07	0.89 (8.09%)
WEST	Bus14	[9.29-14.90]	11.98	1.31 (11.01%)

To generate data for SE, each load bus of the test system is linked with one region of NYISO using the shown map in Table 2 and then fit the normalized load data into the case file. For example, Fig. 4 shows the distribution of normalized NYISO load values for bus 3. Subsequently, we run a power flow based on the power profile above to obtain the true measurement sets. Therefore, we obtain $T = 2045$ normal measurement sets by repeatedly running the power flow. To mimic the effect of random errors, Gaussian noises with zero mean and the standard deviation of 0.02 are added to the measurements. After orchestrating FDI attacks (Section 2), the measurement data are given as inputs to the SE. In these simulations, active power measurements and system states are collected and considered as inputs to the k-NN algorithm.

4.2 Attack State Variable

To test the performance of the proposed method, false data injection attacks on each system state variable $\theta_2 - \theta_{14}$ are simulated. For each attack, one system state variable is decreased or increased by ten percentage of its original value which means two injection amounts 90% and 110% are simulated. 90% means that the manipulated state variable is 10% smaller than the true value.

4.3 Suggested Concepts by the Proposed Method

To find the critical concept set by the proposed method, all of the concepts to which the network topology may drift from the baseline due to branch outage contingency have been defined except the case load-flow solution is unobtainable and the program is diverged

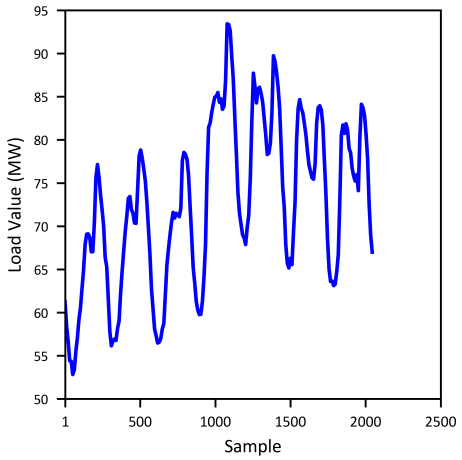

Fig. 4: Normalized load data for bus 3.

(line $7-8$ outage). These concepts are shown by C_i in Fig. 3. For example, the concept C_4 means the network under outage of branch $2-4$. Table 3 shows the computed critical set by the proposed method which drift to those concepts is expected to change the data distribution significantly. The table also shows the concepts which the baseline concept will drift to by occurrence of the computed the critical contingency set using power system indices. Fig. 5 shows an example of the proposed method in IEEE 14 bus system where the proposed method rejects the hypothesis that the distribution of the new concept (stem from a line outage) and the baseline concept are from the same distribution. $D_{n,m}$ is the largest vertical distance between the CDFs.

Table 3: Concepts stem from the occurrence of the obtained critical contingency set

Method	Index	Critical Concepts
Proposed	change of data distribution	$C_1, C_2, C_3, C_8, C_{10}$
Ref [35]	RSSI	C_1, C_2, C_3, C_4, C_6
Ref [37]	FVSI	C_1, C_2, C_3, C_4, C_5
Ref [38]	load power margins	$C_1, C_2, C_3, C_4, C_{14}$

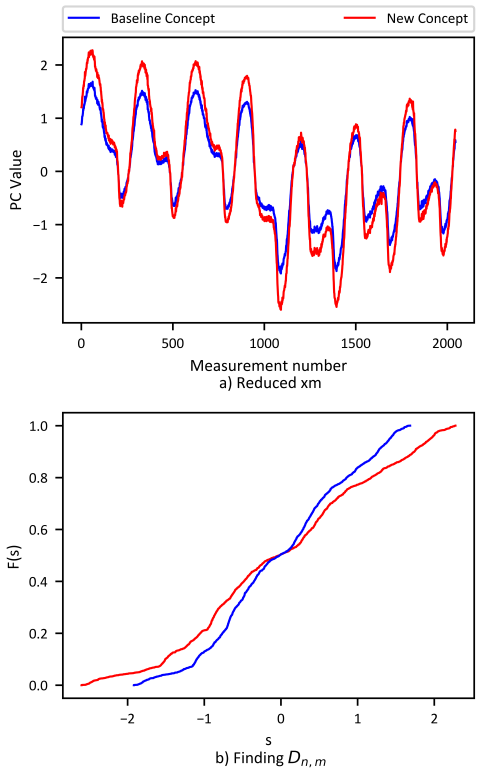

Fig. 5: An example of the proposed method

4.4 Scenario S1: Detecting Attacks in the Network without any Contingencies

In this scenario, it is supposed that there is no contingency in the network (baseline concept) and the system works well except for the seventh day. This means that the attacker has falsified state variables of one day and the measurements of that day is completely replaced by the attacked ones. Therefore, there are 288 attack samples for each attacking scenarios with incremental/decremental attack cost and overall 576 attack samples for each bus. Table 4 summarizes the test results. This table represents the results when the k-NN classifier is trained with just the baseline concept data and with the suggested critical concepts samples by the proposed method. This means we update the training set **TS** by samples of the identified critical concepts and then the forecast model is trained based on the newly updated data. The models were built using tenfold cross-validation.

To evaluate the effectiveness of the computed critical contingency set and the obtained critical concepts by the proposed method, the classifier is also trained based on the collected data of the concepts stem from the occurrence of the critical contingency set identified by references [35], [37], [38] which are shown in Table 3. As it is clear, all trained models were able to detect all attack samples correctly for this scenario.

Table 4: Evaluation of methods

No	Forecast Model	FM	FP
1	Trained with just baseline concept data - Currently utilized approach	1	0%
2	Trained with baseline concept data and the proposed critical concepts - Proposed paradigm	1	0%
3	Trained with baseline concept data and data of critical contingency set by [35] - Comparison purpose	1	0%
4	Trained with baseline concept data and data of critical contingency set by [37] - Comparison purpose	1	0%
5	Trained with baseline concept data and data of critical contingency set by [38] - Comparison purpose	1	0%

4.5 Scenario S2: Detecting Attacks under Concept Drift

This scenario focuses on the prediction of the labels of samples after a line outage using the trained classifiers in the scenario S1 shown in Table 4. To this end, the contingencies are applied one by one and false data are injected into the system after each contingency. The occurrence of each contingency leads to drift to a new concept from the baseline concept. The trained models are tested with each instance of the derived concepts. For this simulation, we have assumed that a line outage occurs at 12 PM of the sixth day and the system works under that contingency until the end of the seventh day. But, the adversary launches attacks to the measurements of the seventh day completely. This means there exists 288 attack samples with incremental/decremental attack cost for each state variable under each contingency and derived concept.

Table 5 represents the results for this scenario. Each row of Table 5 represents average F-measure and FP rate over the state variables of related concept with different trained models in scenario S1. For example, the FP rate for C_4 for the trained model No. 1 is 82% which means this model incorrectly classifies 1694 normal samples out of $T = 2045$ samples as attack ones under concept drift (line $2 - 4$ outage). Fig. 6 presents the average F-measure and FP rate over the different concepts of Table 5. As it is clear from the results, the trained model with just baseline concept (No. 1) is not able to yield its performance and cannot predict samples correctly after system reconfiguration in new concepts.

The dramatic loss of F-measure of the forecast model No.1 after topology changes because the underlying distribution of incoming data unpredictably drifts after topology reconfigurations and old observations became irrelevant to the new ones. This is while the trained model using the proposed paradigm (No. 2) is robust to changes and achieves a high F-measure in detecting attacks in different concepts. This is because the most critical concepts are part of the model. Hence, any deployed method against data integrity attacks should be able to address dynamically changing system configuration.

The results indicate that the trained model with our proposed concepts outperforms the models which are trained by the suggested concepts stem from identified critical contingency set by power system indices. The proposed paradigm has a higher F-measure and less false positive rate compared to those methods. The reason is that, unlike the power system indices, the proposed method selects the critical concepts based on the change in the underlying data distribution.

Table 5: Evaluation of the built models with test data

Test Data	No. 1		No. 2		No. 3		No. 4		No. 5	
	FM	FP	FM	FP	FM	FP	FM	FP	FM	FP
$C1$	0.7	71%	1	0%	1	0%	1	0%	1	0%
$C2$	0.4	50%	1	0%	1	0%	1	0%	1	0%
$C3$	0.75	86%	1	0%	1	0%	1	0%	1	0%
$C4$	0.78	82%	0.9	16%	1	0%	1	0%	1	0%
$C5$	0.94	17%	0.94	14%	1	0%	1	0%	0.94	14%
$C6$	0.92	31%	0.96	16%	1	0%	0.97	12%	0.92	28%
$C7$	0.28	97%	0.88	29%	0.28	100%	0.6	88%	0.6	86%
$C8$	0.64	77%	1	0%	0.64	77%	0.65	76%	0.88	9%
$C9$	0.93	0.09%	0.98	2%	0.93	0.09%	0.93	0.09%	0.93	0.09%
$C10$	0.36	74%	1	0%	0.42	100%	0.41	100%	0.44	100%
$C11$	1	0%	1	0%	1	0%	1	0%	1	0%
$C12$	1	0%	1	0%	1	0%	1	0%	1	0%
$C13$	0.9	0.24%	0.92	0.24%	0.9	0.24%	0.9	0.24%	0.9	0.24%
$C14$	0.93	0.5%	0.93	4%	0.94	0.5%	0.94	0.5%	1	0%
$C15$	1	0%	1	0%	1	0%	1	0%	1	0%
$C16$	0.99	0%	1	0%	0.99	0%	1	0%	0.99	0%
$C17$	1	0%	1	0%	1	0%	1	0%	1	0%
$C18$	1	0%	1	0%	1	0%	1	0%	1	0%
$C19$	1	0%	1	0%	1	0%	1	0%	1	0%

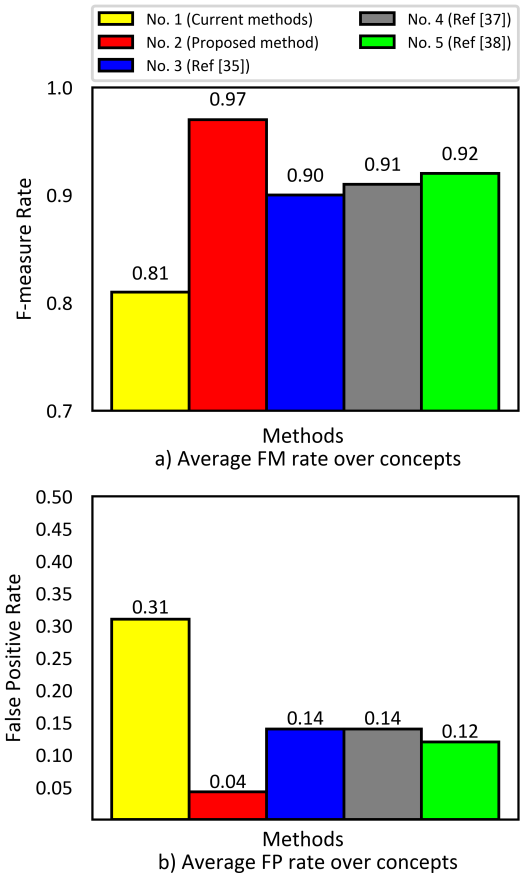

Fig. 6: Summary results of testing trained models with test data.

5 Conclusion

FDI attacks can present a serious threat to operation and control of smart grid. Existing detectors and mitigation methods are unable to manage concept drift since they assume that the training and the future unseen data come from the same distribution. We, therefore,

in this paper, proposed a paradigm to enable robustness for these algorithms. Specifically, instead of using only the baseline concept historical data, we proposed to systematically find the critical concept set which causes significant drift of data distribution from the baseline concept and then update the training set with samples of those concepts. Such a set is computed by using PCA and KS test. k-NN algorithm has been used to show the effectiveness of the proposed method. Unlike the power system indices to find the critical contingency set, the proposed method identifies the critical concepts based on the severity of the change in the underlying data distribution. Numerical results showed that the proposed method can achieve a high accuracy under concept drift and is able to ensure the robustness of the existing countermeasures against concept drift stem from line outages.

REFERENCES

[1] J. Zhao, L. Mili and M. Wang, "A Generalized False Data Injection Attacks Against Power System Nonlinear State Estimator and Countermeasures," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4868-4877, Sep 2018.

[2] M. Mohammadpourfard, A. Sami, and A. R. Seifi, "A statistical unsupervised method against false data injection attacks: A visualization-based approach," *Expert Systems with Applications*, vol. 84, pp. 242-261, May 2017.

[3] L. Xie, Y. Mo and B. Sinopoli, "Integrity Data Attacks in Power Market Operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659-666, Dec 2011.

[4] Electricity Information Sharing and Analysis Center. (2016) "Analysis of the Cyber Attack on the Ukrainian Power Grid," [Online]. Available: https://ics.sans.org/media/E-ISAC SANS Ukraine DUC 5.pdf

[5] Y. Liu, P. Ning and M. K. Reiter, "False Data Injection Attacks Against State Estimation in Electric Power Grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 1-33, Jun 2011.

[6] K. Hamedani, L. Liu, R. Atat, J. Wu, and Y. Yi, "Reservoir Computing Meets Smart Grids: Attack Detection Using Delayed Feedback Networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 734-743, 2018.

[7] M. Mohammadpourfard, A. Sami, and Y. Weng, "Identification of False Data Injection Attacks with Considering the Impact of Wind Generation and Topology Reconfigurations," *IEEE Transactions on Sustainable Energy*, vol. PP, no. 99, pp. 1-1, Dec 2017.

[8] R. Deng, P. Zhuang, and H. Liang, "CCPA: Coordinated Cyber-Physical Attacks and Countermeasures in Smart Grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2420-2430, Sep 2017.

[9] R. Deng, G. Xiao, and R. Lu, "Defending against false data injection attacks on power system state estimation," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 198207, 2017.

[10] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 12441253, 2013.

[11] T. T. Kim, and H. V. Poor, "Strategic Protection Against Data Injection Attacks on Power Grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326-333, 2011.

[12] J. Kim, L. Tong, "On phasor measurement unit placement against state and topology attacks," *IEEE International Conference on Smart Grid Communications*, pp: 396401, 2013.

[13] G. Chaojun, P. Jirutitijaroen and M. Motani, "Detecting False Data Injection Attacks in AC State Estimation," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2476-2483, Sep 2015.

[14] Y. Chakhchoukh, H. Ishii, "Enhancing robustness to cyber-attacks in power systems through multiple least trimmed squares state estimations," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4395-4405, 2016.

[15] M. G. Kallitsis, S. Bhattacharya and G. Michailidis, "Detection of False Data Injection Attacks in Smart Grids Based on Forecasts," 2018 *IEEE International Conference on Communications*, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aalborg, 2018, pp. 1-7.

[16] M. Ozay, I. Esnaola, F. T. Vural, S. R. Kulkarni and H. V. Poor, "Machine Learning Methods for Attack Detection in the Smart Grid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1773-1786, Aug 2016.

[17] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng and Z. Han, "Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1-9, Aug 2014.

[18] J. Gama, I. liobait, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 1-37, Apr 2014.

[19] S. Bi, and Y. J. Zhang, "Graphical Methods for Defense Against False-Data Injection Attacks on Power System State Estimation," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1216-1227, May 2014.

[20] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang and W. Zhao, "On False Data-Injection Attacks against Power System State Estimation: Modeling and Countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 717-729, Mar 2014.

[21] S. Li, Y. Ylmaz and X. Wang, "Quickest Detection of False Data Injection Attack in Wide-Area Smart Grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725-2735, Nov. 2015.

[22] J. Yan, B. Tang and H. He, "Detection of false data attacks in smart grid with supervised learning," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 1395-1402.

[23] M. Mohammadpourfard, Y. Weng, and M. Tajdinian, "Benchmark of machine learning algorithms on capturing future distribution network anomalies," *IET Generation, Transmission& Distribution*, vol. 13, no. 8, pp. 1441-1455, Apr 2019.

[24] R. Moslemi, A. Mesbahi and J. M. Velni, "A Fast, Decentralized Covariance Selection-Based Approach to Detect Cyber Attacks in Smart Grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4930-4941, Sep 2018.

[25] H. Sedghi and E. Jonckheere, "Statistical Structure Learning to Ensure Data Integrity in Smart Grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1924-1933, Jul 2015.

[26] A. Abur and A. G. Expsito, "Power System State Estimation: Theory and Implementation," 1st ed. New York, NY, USA: Marcel Dekker, Mar 2004.

[27] A. Monticelli, "State Estimation in Electric Power Systems: A Generalized Approach," New York, NY, USA: Springer, May 1999.

[28] H. Jiawei, J. Pei and M. Kamber, "Data mining: concepts and techniques", Waltham, MA, USA: Elsevier, 2012.

[29] I. Jolliffe, "Principal component analysis", New York, USA: Springer; 2nd edition, Oct 2002.

[30] P. Schmid, "On the Kolmogorov and Smirnov Limit Theorems for Discontinuous Distribution Functions," *The Annals of Mathematical Statistics*, vol. 29, no. 4, pp. 1011-1027, 1958.

[31] C. L. Chiang, "Statistical Methods of Analysis," Singapore: World Scientific Publishing Company, 2003.

[32] D. H. Johnson, "The Insignificance of Statistical Significance Testing," *The Journal of Wildlife Management*, vol. 63, no. 3, pp. 763-772, Jul 1999.

[33] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd ed. Springer, Feb. 2009.

[34] Z. Li, J. Wang, H. Sun and Q. Guo, "Transmission Contingency Analysis Based on Integrated Transmission and Distribution Power Flow in Smart Grid," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3356-3367, Nov. 2015.

[35] A. Doroudi, A. Motie Nasrabadi and R. Razani, "Two novel static and dynamic voltage stability based indexes for power system contingency ranking," *IET Generation, Transmission & Distribution*, vol. 12, no. 8, pp. 1831-1837, Apr 2018.

[36] S. R, R. S. Kumar and A. T. Mathew, "Online Static Security Assessment Module Using Artificial Neural Networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4328-4335, Nov 2013.

[37] I. Musirin and T. K. A. Rahman, "On-line voltage stability based contingency ranking using fast voltage stability index (FVSI)," *IEEE/PES Transmission and Distribution Conference and Exhibition*, Yokohama, Japan, 2002, pp. 1118-1123 vol.2.

[38] N. Yorino, H. Q. Li, S. Harada, A. Ohta and H. Sasaki, "A method of voltage stability evaluation for branch and generator outage contingencies," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 252-259, Feb 2004.

[39] R. D. Zimmerman and C. E. Murillo-Sanchez, "Matpower, a matlab power system simulation package," http://www.pserc.cornell.edu/ matpower/manual.pdf, 2010.

[40] NYISO, "Load data profile," [Online]. Available: http://www.nyiso.com. [Accessed Apr 2018].