





Article

Covert Cyber Assault Detection in Smart Grid Networks Utilizing Feature Selection and Euclidean Distance-Based Machine Learning

Saeed Ahmed , YoungDoo Lee , Seung-Ho Hyun  * and Insoo Koo 

School of Electrical Engineering, University of Ulsan, Ulsan 44610, Korea; saeed.ahmed21@gmail.com (S.A.); leedy1004@naver.com (Y.L.); iskoo@ulsan.ac.kr (I.K.)

* Correspondence: takeitez@ulsan.ac.kr; Tel.: +82-52-259-1277

Received: 11 April 2018; Accepted: 9 May 2018; Published: 12 May 2018



Abstract: Communications technologies are an integral part of efficient monitoring and reliable control in smart grids, but enhanced reliance on these technologies heightens the risk of cyber assaults. Recently, a new type of stealth, or covert, assault in smart grid networks has been discovered, which cannot be ascertained by legacy bad-data detectors using state estimation. Due to the delay-sensitive nature of smart grid networks, swift detection of abnormal changes is immensely desired. In this paper, we propose two Euclidean distance-based anomaly detection schemes for covert cyber-assault detection in smart grid communications networks. The first scheme utilizes unsupervised-learning over unlabeled data to detect outliers or deviations in the measurements. The second scheme employs supervised-learning over labeled data to detect the deviations in the measurements. Unlike the classic detection test, the proposed schemes tackle an unknown sample with low computational complexity, leading to a shorter decision time. To improve detection accuracy and further reduce the computational complexity and the associated time delay, we employ a genetic algorithm-based feature selection method to choose the distinguishing optimal feature data subset as input to both of the proposed schemes. The evaluation is carried out through the standard IEEE 14-bus, 39-bus, 57-bus and 118-bus test systems. Simulation results show that compared to the existing feature extraction-based detection schemes, the proposed schemes show significant improvement in covert cyber deception assault-detection accuracy.

Keywords: anomaly detection; cyber assaults; Euclidean distance; feature selection; genetic algorithm; smart grids; state estimation

1. Introduction

The emerging smart grid (SG) concept as a cyber-physical complex organization is being implemented through a composition of communications networks overlaying traditional power systems. Due to the electrical energy flow's vital dependency on communications technologies in the SG, its vulnerability to new malicious types of cyber attack is very high. Nation states are apprehensive about power grid privacy and security. Therefore, vigorous and secure communications management is essential to all aspects of the SG. The security, privacy and integrity of data and the information network have become a prime focus of research activities in SGs. Traditionally, bulk storage of the generated electricity is not possible, and hence, its generation should be closely equated to consumption; otherwise, there can be a deviation in the electrical quantities. Thus, the power control center (PCC) needs to monitor the power network closely to make sure that the operation of the power system is safe and reliable. State estimation (SE) is a fundamental approach employed in an energy management system (EMS) to monitor states in power networks.

Fundamental elements (generation, transmission and consumption) of a power system, along with communications links, are illustrated in Figure 1. Distributed sensors, actuators and meters designated as remote terminal units (RTUs) are employed in electric power grids to aggregate measurements, including bus power injections and branch power flows. These measurements are combined at the PCC via communications links and are further used to estimate the states (i.e., bus voltage angles). These state variables form the basis of suitable decisions by the EMS about auto-generation control (AGC) and optimal power flow (OPF) to keep electric power systems in a safe operating zone. On the one hand, the existence of a communications infrastructure is compulsory for the realization of efficient monitoring and intelligent control in the framework of an SG, but the communications infrastructure is prone to malicious cyber-assault threats [1–3], due to certain incentives for the attacker. Unidirectional flow of information in legacy power networks (i.e., from RTUs to PCC) makes it more important to study a particular type of malicious user behavior that attempts to target the integrity of the measurement data by inserting a deceptive bias value into the SE. Such malicious activity goes mostly undetected by bad-data detection (BDD) systems in the legacy PCC. We term this kind of attack a covert cyber deception (CCD) assault, but it is also known as a false data injection (FDI) attack, a cyber stealthy deception (CSD) attack, and so on [4]. Identification and removal of the susceptibilities or anomalies injected by a CCD assault are critically important because of their negative impacts on the safety and reliability of SGs. Methods reported in the literature to mitigate the effects of CCD assaults on SGs can be broadly divided into two classes: (1) protection-based defense; and (2) detection-based defense. M. Ozay et al. [5] utilized a variety of machine learning-based schemes to detect the CCD assault in SG. Sandeep et al. [6] proposed joint-transformation-based scheme to detect CCD assault. They utilize the Kullback–Leibler distance to find out the difference between probability distributions obtained from measurement variations. However, they did not employ feature selection (FS)-based techniques in their work to tackle the dimensionality issue with increasing power system sizes. Esmalifalak et al. [7] used the PCA-based feature extraction (FE) technique to tackle the dimensionality issue in the state estimation-measurement feature (SE-MF) dataset and then employed a statistical method-based anomaly detection (AD) mechanism to detect the CCD assault. Unlike the existing schemes, in this paper, we focus on the selection of the discriminating features from the SE-MF dataset utilizing genetic algorithm (GA) to tackle the curse of dimensionality [8]. The optimal features selected from the SE-MF dataset are then used as input by the two proposed Euclidean distance (ED)-based AD schemes for the detection of a CCD assault. Contrary to the FE-based approach, the proposed FS-based method does not alter the original representation of the data.

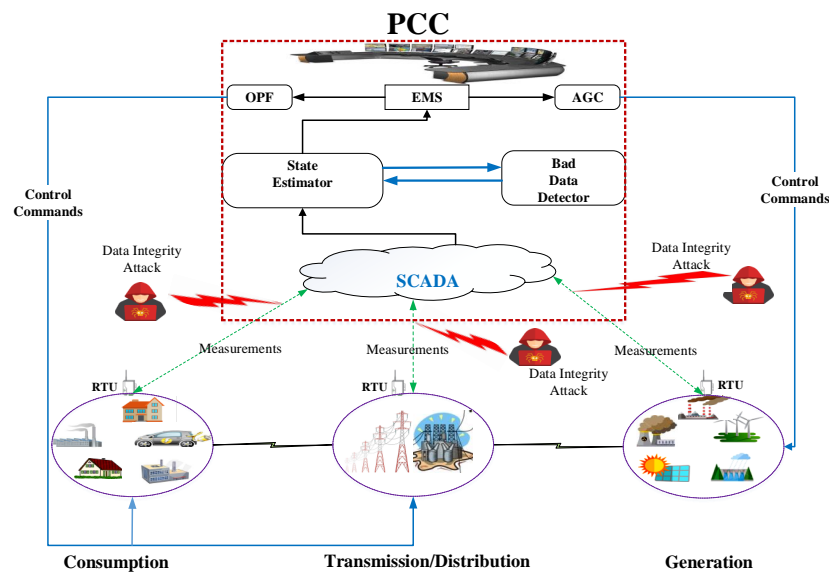


Figure 1. Covert cyber deception assault in a smart grid communications network. PCC, power control center; OPF, optimal power flow; AGC, auto-generation control; RTU, remote terminal unit.

1.1. Motivation

Normal data are consistent with physical laws, like Kirchhoff's current and voltage laws, whereas the compromised data that are affected by a CCD assault are inconsistent with these laws. Therefore, normal and compromised data will have different distributions and will, therefore, tend to form different clusters. These clusters would be distinguishable in a feature space of suitable dimensions. This fundamental distinction inspires the distance-based anomaly detection schemes for the detection of CCD assaults. Unsupervised anomaly detection (AD) techniques are persuasive at differentiating between data that have different underlying distributions, particularly when the data are not labeled. Similarly, supervised anomaly detection techniques can be employed to detect the anomalies in the labeled data. Thousands of sensors or RTUs are employed in power grids, spanning over a vast geographical area. Practically, defense mechanism can be designed to protect a limited set of critical RTUs and corresponding measurement features (MFs). Therefore, in the context of SG cyber-security, the selection of distinctive features from the SE-MF dataset becomes a promising strategy to detect the CCD assault and tackle the curse of dimensionality [8].

1.2. Related Works

The benefits and risks involved in utilizing communications technologies alongside legacy electrical power systems have been widely reviewed [9–13]. Intrusion into a communications network by a pernicious user who is aiming to target the integrity of the data can have a catastrophic impact on the secure and reliable operation of an SG [14–16]. Therefore, in the context of the security of SGs, understanding the nature of the assault and identification of compromised data has been the focus of research in electric power systems. The conventional state estimator in a PCC utilizes BDD to single out and disassociate the bad data for state estimation. However, Liu et al. [17] demonstrated that a smart attacker who has information on the network topology can realize the construction of a set of falsified data that can dodge legacy BDD. This type of attack is known as an unobservable (or covert) cyber assault. Many schemes considering the construction of intrusion assaults against state estimation, and the subsequent defense measures against them, have been discussed in the literature [4,13–21]. Li et al. [18] proposed a decentralized conjunctive rule-based majority voting algorithm to detect compromised or assaulted phase measurement units. Huang et al. [19] proposed cumulative sum

hypothesis test-based bad-data detection in a state estimator. Xie et al. [22] demonstrated that a data integrity assault can methodically result in a considerable economic loss in real-time market operations. Similarly, Esmalifalak and colleagues [23] studied the fiscal impact of a false data injection attack on electric power market operations. An encryption-based security mechanism integrated into power system devices was proposed [24] to improve the security of the power system against FDI attacks. Methods reported in the literature to mitigate the effects of CCD assaults on SGs can be broadly divided into two classes: (1) protection-based defense; and (2) detection-based defense. Thousands of sensors or RTUs are employed in power grids, spanning over a vast geographical area. Practically, defense mechanism can be designed to protect a limited set of critical RTUs and corresponding measurement features (MFs). Therefore in the context of SG cyber-security and computational complexity, the selection of distinctive features becomes a promising strategy to detect the CCD assault in real time [5]. Sandeep et al. [6] proposed a joint-transformation-based scheme to detect CCD assault. They utilized the Kullback–Leibler distance to find out the difference between probability distributions obtained from measurement variations. However, they did not employ feature selection (FS)-based techniques in their work to tackle the dimensionality issue with increasing power system sizes. Esmalifalak et al. [7] used the PCA-based FE technique to tackle the dimensionality issue in the SE-MF dataset and then employed a statistical method-based AD mechanism to detect the CCD assault.

In summary, existing works on CCD assault detection in SGs only have generally considered feature extraction or transformation [5–7] in the context of cybersecurity and the curse of dimensionality. To the best of our knowledge, the selection of distinguishing features from the SE-MF dataset in the context of SG security is still an open problem.

1.3. Contributions

In this paper, we focus on the selection of the discriminating features from the SE-MF dataset utilizing the genetic algorithm (GA) to tackle the curse of dimensionality [8]. The optimal features selected from the SE-MF dataset are then used as input by the two proposed Euclidean distance (ED)-based AD schemes for the detection of a CCD assault. Contrary to the FE-based approach, the proposed FS-based method does not alter the original representation of the data. The main contributions of this paper can be summarized as follows:

- We study intelligently-crafted CCD assaults on the SE-MF dataset, and we investigate how such an assault goes undetected in legacy systems that use bad-data detectors.
- To tackle the increasing computational complexity with the growing sizes of power systems, we use GA for the selection of independent and discriminating features from the SE-MF dataset. The selection of discriminative features leads to lower computational costs, a shorter time delay and improved accuracy.
- First, we propose an ED-based AD scheme to detect the presence of outliers in the unlabeled SE-MF dataset. Next, we extend the first scheme to propose a detection mechanism for the labeled SE-MF dataset. In both schemes, the optimal features selected through the GA are employed as input.
- We use the IEEE standard 14-bus, 39-bus, 57-bus and 118-bus test systems to evaluate the efficiency of the proposed schemes. The performance evaluation shows that the proposed schemes provide better accuracy, in comparison to existing AD-based schemes.

1.4. Paper Organization

The remainder of this paper is organized as follows. In Section 2, we present the system model and explain the behavior of a CCD assault in SG networks. In Section 3, we first describe the GA-based FS mechanism and then describe the two proposed ED-based AD schemes to detect CCD assaults. Simulation results are presented in Section 4. We conclude the paper in Section 5. Table 1 lists the abbreviations used throughout the paper, and Table 2 lists and CCD assault notations and concepts.

Table 1. Nomenclature.

Abbreviation	Term	Abbreviation	Term
AD	anomaly detection	GA	genetic algorithm
AGC	auto-generation control	MF	measurement features
BDD	bad-data detection	NCA	neighborhood component analysis
CCD	covert cyber deception	OPF	optimal power flow
CSD	cyber stealthy deception	PCA	principle component analysis
EC	evolutionary computation	PCC	power control center
ED	Euclidean distance	PSO	particle swarm optimization
EMS	energy management system	ROC	receiver operating characteristic
FDI	false data injection	RTU	remote terminal unit
FE	feature extraction	SE	state estimation
FS	feature selection	SG	smart grid

Table 2. Average number of features selected by GA.

Standard IEEE Bus System	States	Features	Selected Features
14-bus	13	53	23
39-bus	38	130	61
57-bus	56	216	111
118-bus	117	489	233

2. Covert Cyber Deception Assault

State estimation at the PCC is the essential instrument for ensuring the reliable and sustainable functioning of electrical power networks [25]. As illustrated in Figure 1, the measurement data collected from RTUs via communications networks are used by the state estimator to determine the system states over time. The problem with state estimation is how to approximate power system state variables based on the measurement data.

2.1. Legacy Bad-Data Detectors in PCCs

The measurement data and the state variables are related through the following alternating AC power flow observation model:

$$Z_{meter} = h(\delta) + e, \quad (1)$$

where $h(\delta)$ is a non-linear relationship between measurement data, Z_{meter} , and the state vector δ ; $e = [e_1, e_2, \dots, e_m]^T$ is the Gaussian measurement noise vector with standard deviation σ . Using a linear or direct current (DC) power flow model, the observation model in (1) becomes further simplified with a small sacrifice of accuracy, as follows [26,27]:

$$Z_{meter} = H\delta + e. \quad (2)$$

In a DC power flow problem, the Jacobian matrix H can be approximated as follows:

$$H = \left. \frac{h(\delta)}{h\delta} \right|_{\delta=0}, \quad (3)$$

where H is composed of topology and impedance data only. One objective of (2) is to determine the estimated state, $\hat{\delta}$, that is the best fit for the meter measurements. In other words, we can say that the best estimated value can minimize estimation weighted least square (WLS) error, $(Z_{meter} - H\hat{\delta})^T \Omega (Z_{meter} - H\hat{\delta})$. By applying the WLS statistical estimation criteria, the estimated voltage phase angle is given as follows:

$$\hat{\delta} = (H^T \Omega H)^{-1} H^T \Omega Z_{meter} = W Z_{meter}, \quad (4)$$

where $W = (H^T \Omega H)^{-1} H^T \Omega$ and Ω is a diagonal matrix where diagonal elements are $\Omega_{ii} = \sigma_i^{-2}$, with σ_i^{-2} being the variances of meter errors. Noise in the wireless medium, faulty meters or malicious user behavior (like CCD assaults) can be potential sources for abnormal data in estimated measurements. Current power systems use a residual-based detector for BDD to protect state estimation [28]. The difference between observed meter measurements Z_{meter} and estimated measurements \hat{Z} is the residual, R , and it is expressed as follows:

$$R = Z_{meter} - Z_{estimated} = (I - M)Z_{meter}. \quad (5)$$

The expected value and the co-variance of the residual are:

$$\begin{aligned} E(R) &= 0, \\ \text{cov}(R) &= (I - A)R. \end{aligned} \quad (6)$$

The BDD present in the PCC performs the l_2 -norm test [28] to compare the results with a predefined threshold. The hypothesis of not being attacked is accepted if we have:

$$\max_i |R_i| \leq \lambda, \quad (7)$$

where R_i is the component of residual vector R and λ is the threshold.

2.2. The Covert Cyber Deception Assault

Familiar with the topology of H matrix, an attacker can initiate an assault by altering the value of the meter measurements. Let $Z_{assault} = Z_{meter} + a$, where $a \in \mathbb{R}^{m \times 1}$ denotes the malicious data injected into the meter measurement data vector. If the malicious user constructs vector a as follows:

$$a = Hc, \quad (8)$$

where $c \in \mathbb{R}^{m \times 1}$ is any arbitrary non-zero vector, the legacy BDD cannot detect such an assault. The reason is as follows. Let $\hat{\delta}_{assault}$ denote the estimate of state variables using assaulted meter measurements $Z_{assault}$, i.e.,

$$\hat{\delta}_{assault} = WZ_{assault} + Wa = \hat{\delta} + WHc = \hat{\delta} + c. \quad (9)$$

Now, the l_2 norm for the assaulted measurement $Z_{assault}$ residual is as follows:

$$\begin{aligned} \|R_{assault}\|_2 &= \|Z_{assault} - H\hat{\delta}_{assault}\|_2 \\ &= \|(Z + a) - H(\hat{\delta} + c)\|_2 \\ &= \|(Z - H\hat{\delta}) + (a - Hc)\|_2 = \|(Z - H\hat{\delta})\|_2 \\ &= \|R\|_2. \end{aligned} \quad (10)$$

The residual calculated with assaulted measurements is the same as it is for normal measurements. Hence, $Z_{assault}$ will be able to deceive the BDD statistical test presented in (8) and will change the system states, resulting in crucial operational failures [4,17]. This sort of assault is termed an unobservable (or covert) attack [4]. Under these assumptions, the observation model in the presence of the CCD assault can be described as follows:

$$Z_{assault} = H\delta + a + e, \quad (11)$$

where a is the non-zero assault vector.

3. CCD Assault Detection Using Euclidean Distance-Based Anomaly Detection Schemes

In this section, we discuss a two-tier mechanism for the detection of CCD assaults in the SE-MF dataset. The curse of dimensionality [8] increases the computational complexity when measurement features grow with increased sizes of the power systems. Moreover, all of the SE-MF dataset attributes would not be equally supportive in leading to plainly distinguishable clusters in the feature space; this can have a negative impact on the performance of a detection method. Therefore, first, we use the FS-based approach to select an optimal subset of features that would result in more tightly-packed and distinctly-separable clusters of vectors of chosen features in the resulting subspace. Through FS, we also can reduce the measurement and storage requirements at the PCC, as well as the training and prediction times [29]. The selected optimal features are then employed as input to the two ED-based AD schemes (EDADS-1 and EDADS-2) to detect CCD assaults in SG communications networks. When the class labels are not given in the SE-MF dataset, we utilize the proposed EDADS-1 to identify the outliers as potential assaults. On the other hand, when the data are supplemented with class labels, we propose EDADS-2 to distinguish between the normal and compromised data. In the following subsections, we explain the GA and the proposed detection schemes.

3.1. Dimensionality Reduction Using Genetic Algorithm-Based Feature Selection

The goal of FS is to choose a subset of features (from a given set of features) that yields minimum classification error. Works reported on dimensionality reduction affirm that FS techniques retain data characteristics for interpretability. Furthermore, overfitting due to fewer redundant data is reduced and modeling accuracy is improved with FS methods [29–34]. The interrelationship between numerous dimensionality reduction approaches (encompassing feature subset selection) and FE with different flavors of PCA techniques was studied. Janecek et al. [30] analytically tested the effects of these methods on classification accuracy with two different types of datasets (email data and drug discovery data). The results revealed that feature transformation using PCA is highly sensitive to the type of data. Generally, FS methods can be divided into three categories: filters, wrappers, and embedded/hybrid methods. Wrapper methods are advantageous over filter approaches due to them giving better performance since they use the target classifiers such as K-nearest neighbors and support vector machine for feature selection. For a large dataset, however, wrapper methods are computationally expensive. The filter approach is known to be more computationally efficient than wrapper methods and performs well with large dataset [29,35]. In this paper, considering the delay-sensitive nature and increasing sizes of power systems, we use a filter-based FS mechanism that is independent of any learning algorithm or classifier. Working as a preprocessor in the paper, the filter-based FS will select features by considering their scores in different statistical tests for correlation with the outcome variable. We use GA to select the subset of features from the SE-MF dataset that is the best at discriminating compromised data from normal data. GA has been widely used for FS purposes in the machine learning and is considered suitable for large combinatorial problems [30,31,36]. However, recently, particle swarm optimization (PSO) and other evolutionary and metaheuristic algorithms have gained the attention of researchers due to their lesser complexity and simplicity. PSO may be a promising method for FS and an interesting topic for future works in SG security. The GA emulates biological evolution and Darwinian selection [36]. The evolution mechanism of living beings is believed to follow natural selection, i.e., living species that are better suited to their environment thrive, whereas species that are at a disadvantage in their environment go extinct. Following the same principle, GA improves a given solution by incrementally choosing better possible solutions, while eliminating worse solutions.

The quality of each solution is calculated using a fitness value function based on the objective function. The m -dimensional set of SE-MF vector data in \mathbb{R}^n is given as input to GA as follows:

$$\begin{aligned} & \left\{ X_1^{(m)}, X_2^{(m)}, \dots, X_k^{(m)} \right\} \\ \text{where } X_i^{(m)} &= \begin{bmatrix} x_1 & x_2 & \dots & x_{m-1} & x_m \end{bmatrix}, \\ & \forall i \in \{1, 2, \dots, k\}. \end{aligned} \quad (12)$$

GA yields a set of n -dimensional vectors in subspace \mathbb{R}^n , described as:

$$\begin{aligned} & \left\{ X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)} \right\} \\ \text{where } X_i^{(n)} &= \begin{bmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \end{bmatrix}, \\ & \forall i \in \{1, 2, \dots, k\}. \end{aligned} \quad (13)$$

It is notable that the GA reduces the dimensionality of each vector in the set without affecting the cardinality of the set of vectors in Equation (14), i.e., $n \ll m$. The selected dimensions are chosen to optimize the fitness function. Hence, $n \ll m$ denotes an instance of the feature vector in the subspace that optimizes the fitness function. Fitness function F , which is adopted in this paper, is given as follows:

$$F = \frac{\bar{C}}{\bar{S}}. \quad (14)$$

In (14), \bar{C} is the mean compactness of classes and is expressed as follows:

$$\bar{C} = \frac{1}{L} \sum_i^L C_i', \quad (15)$$

where the mean separability, denoted by \bar{S} in (14), is the separation between any two classes in an L -class problem, obtained as follows:

$$\bar{S} = \frac{2}{L(L-1)} \sum_{i \neq j}^L S_{ij}. \quad (16)$$

In this paper, we are dealing with a binary classification problem. Therefore, $L = 2$, i.e., normal and compromised SE-MF measurement data. GA finds a feature subspace that would minimize the ratio of the mean values of inter-class separability and intra-class compactness, defined as follows and illustrated in Figure 2.

- Inter-class separability: measures how well separated two different clusters are from each other.
- Intra-class compactness: indicates how well clustered the sample vectors are for a given class.

To measure the compactness of a given class, GA calculates the mean or centroid, $\mu^{(i)}$, of class i as follows:

$$\mu = \frac{1}{N} (X_1 + X_2 + \dots + X_N). \quad (17)$$

Here, N is the total number of samples of class i . After that, the compactness of class i is determined by finding the mean value of the Euclidean norm, as follows:

$$C_i = \frac{1}{N} \sum_{j=1}^N \|X_j - \mu^{(i)}\|. \quad (18)$$

The Euclidean distance between the centroids of two classes describes the separability between the two classes i and j . It is determined as follows:

$$S_{ij} = \left\| \mu^{(i)} - \mu^{(j)} \right\|. \quad (19)$$

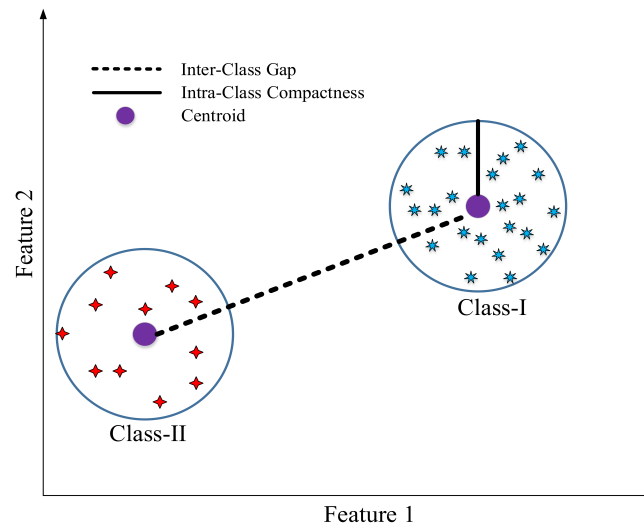


Figure 2. The concept of compactness and separability.

GA encodes the SE-MF data into chromosomes, which goes through crossovers and mutations. Thus, new generations of chromosomes are yielded, which substitute for their parents, provided they are healthier, i.e., their fitness or objective function value is smaller. This process is iterated over many generations until there is no further improvement in the fitness function [29]. A binary encoding scheme is used to represent the features or attributes of the SE-MF dataset as chromosomes. A chromosome is simply a string of binary ones and zeroes, where one indicates that a certain SE-MF feature is selected and zero means it is rejected. The index of each one and zero in the chromosome corresponds to a distinct SE-MF attribute. In the beginning, the GA randomly selects different subsets of the SE-MF. In other words, a primary population of chromosomes (a string of ones and zeroes) initiates the algorithm. A new population of chromosomes is created by subjecting the primary (parent) chromosomes to crossover and mutations. Two parent chromosomes exchange information or swap fragments at randomly-chosen crossover points during the crossover process. However, during the mutation process, the bits are flipped at randomly-selected positions in a chromosome. Then, based on their respective fitness function value, chromosomes are ranked in the evaluation process. Finally, the chromosomes that minimize the proposed fitness function are selected to produce new chromosomes. This process is repeated for many generations until there is no further decrease in the value of the proposed fitness or objective function.

3.2. Euclidean Distance-Based Anomaly Detection Scheme 1

For a large number of given data points, the datasets that vary significantly from the average of the data are called outliers or anomalies. Anomaly detection is a class of machine learning applications, and it has many areas of utilization, such as data cleaning, diagnosis, fraud detection, intrusion detection, and so on. Different types of anomaly detection techniques have been proposed in the literature, such as model-based, distance-based and statistical-based methods [37]. Considering the scenario where the class labels are not provided in the SE-MF dataset, we propose an ED-based anomaly detection scheme (EDADS-1), depicted in Figure 3. The measurement samples are periodically collected at the PCC via the RTUs installed in different locations of the electrical power network. The historical dataset is

formulated at the beginning. During the FS process, the optimal features of the SE-MF dataset are selected through GA. Then, the optimal features dataset is divided into training and testing subsets. It is worth mentioning that the training and testing subsets carry both normal and compromised data. In Step 1, centroid vector $C_d = \{\mu_1, \mu_2, \dots, \mu_n\}$ is calculated by finding the mean, μ_r , of all the features of the training data subset, where $\mu_r = \frac{\sum_{i=1}^a X_{Ri}^{(r)}}{a}, \forall r \in \{1, 2, \dots, n\}$. In Step 2, the Euclidean distance between each sample and the centroid is calculated to form a distance vector, $D_R = \{D_{R1}, D_{R2}, \dots, D_{Ra}\}$, where $D_{Rp} = \sqrt{\sum_{i=1}^n \{(\mu_i - X_{Rp}^{(i)})^2\}}, \forall p \in \{1, 2, \dots, a\}$. In Step 3, average distance $D_{R_avg} = \frac{\sum_{i=1}^a D_{Rp}}{a}$ is calculated. The average distance can be considered a virtual boundary around the normal dataset. In the testing phase, the Euclidean distance of each sample is calculated from its centroid to form distance vector $D_T = \{D_{T1}, D_{T2}, \dots, D_{Tb}\}$, where $D_{Tq} = \sqrt{\sum_{i=1}^n \{(\mu_i - X_{Tq}^{(i)})^2\}}, \forall q \in \{1, 2, \dots, b\}$. Finally, the test is performed to identify the new test point as being normal (if $D_{Tq} < D_{R_avg}$) or an outlier (if $D_{Tq} > D_{R_avg}$).

The proposed EDADS-1 algorithm has appreciably low computational complexity grounded in the fact that calculation of the distance of one sample is required to identify the new data point as being normal or a potential CCD assault. Additionally, it has the ability to identify the outliers in the SE-MF dataset as a potential assault even though the class labels are provided. Moreover, with the increased historical SE-MF dataset, the average distance D_{R_avg} becomes closer to the actual value, and the detection is more accurate.

3.3. Euclidean Distance-Based Anomaly Detection Scheme 2

Using the Euclidean distance-based method, we can detect a CCD assault with improved accuracy when the labels (normal versus compromised) are given in the SE-MF dataset. In this subsection, we explain the second proposed scheme, EDADS-2, shown in Figure 4. In the beginning, the GA is applied to select the optimal features subset from the SE-MF historical dataset. The resulting optimal features data subset consisting of compromised and normal data is divided into the training dataset, $X_R = \{X_{R1}^{(n)}, X_{R2}^{(n)}, \dots, X_{Ra}^{(n)}\}$, where $X_{Rp}^{(n)} \in X, \forall p \in \{1, 2, 3, \dots, a\}$; and the testing dataset, $X_T = \{X_{T1}^{(n)}, X_{T2}^{(n)}, \dots, X_{Tb}^{(n)}\}$, where $X_{Tq}^{(n)} \in X, \forall q \in \{1, 2, 3, \dots, b\}$. The labels are used to select the normal set, $X_{RN} = \{X_{RN1}^{(n)}, X_{RN2}^{(n)}, \dots, X_{RNf}^{(n)}\}$, where $X_{RNs} \in X, \forall s \in \{1, 2, 3, \dots, f\}$, and f is the number of training samples designated as normal. In Step 1, the centroid of the training data subset designated as normal, with the help of labels, is calculated to form vector $C_d = \{\mu_1, \mu_2, \dots, \mu_n\}$, where $\mu_r = \frac{\sum_{i=1}^f X_{RNI}^{(r)}}{f}, \forall r \in \{1, 2, \dots, n\}$. It is pertinent to mention here that for EDADS-2, the centroid is calculated taking into account the normal training samples only. On the other hand, for the proposed EDADS-1, the centroid is calculated including both normal and compromised samples. In Step 2, the Euclidean distance between each sample and the centroid is calculated to form distance vector $D_R = \{D_{RN1}, D_{RN2}, \dots, D_{RNf}\}$ where $D_{RNs} = \sqrt{\sum_{i=1}^n \{(\mu_i - X_{RNs}^{(i)})^2\}}, \forall s \in \{1, 2, \dots, f\}$. In Step 3, the elements of distance vector D_R , are sorted into descending vector $S_{DRN} = \{S_{DRN1}, S_{DRN2}, \dots, S_{DRNf}\}, S_{DRN} \subseteq D_{RN}$ to choose the first 10% of its elements, which are at the farthest distance from the centroid, to make the vector $\mu_{DRN} = \{S_{DRN1}, S_{DRN2}, \dots, S_{DRN(0.1 \times f)}\}$. Employing vector μ_{DRN} , average distance $D_{R_avg} = \sum_{i=1}^{0.1 \times f} \frac{S_{DRNi}}{0.1 \times f}$ of each sample from the centroid is calculated. Finally, when a new test sample arrives, it is identified as being normal (if $D_{Tq} \leq D_{R_avg}$) or an outlier (if $D_{Tq} > D_{R_avg}$). EDADS-2 also has low computational complexity and may identify a new test sample as an outlier or normal in a very

small amount time because it only has to calculate the distance of one sample to compare it with the average distance.

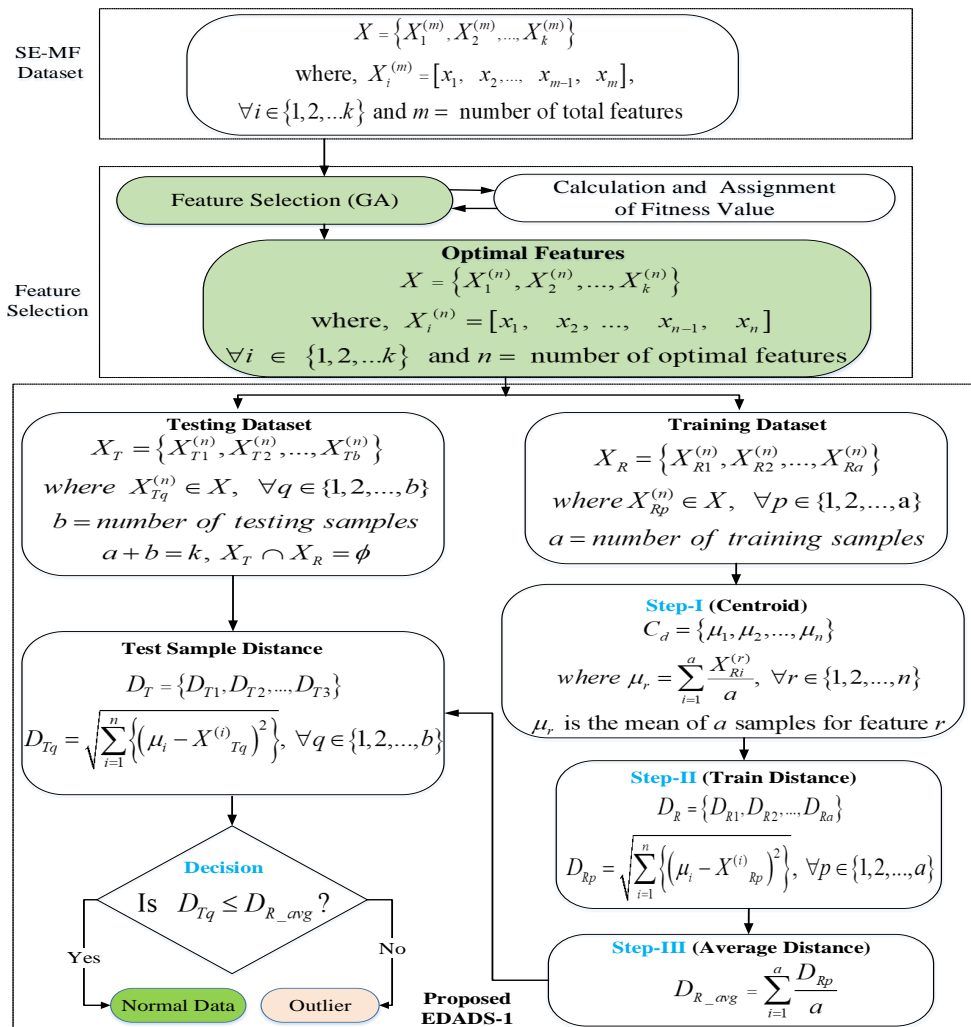


Figure 3. Flowchart of the Euclidean distance-based anomaly detection Scheme 1, EDADS-1.

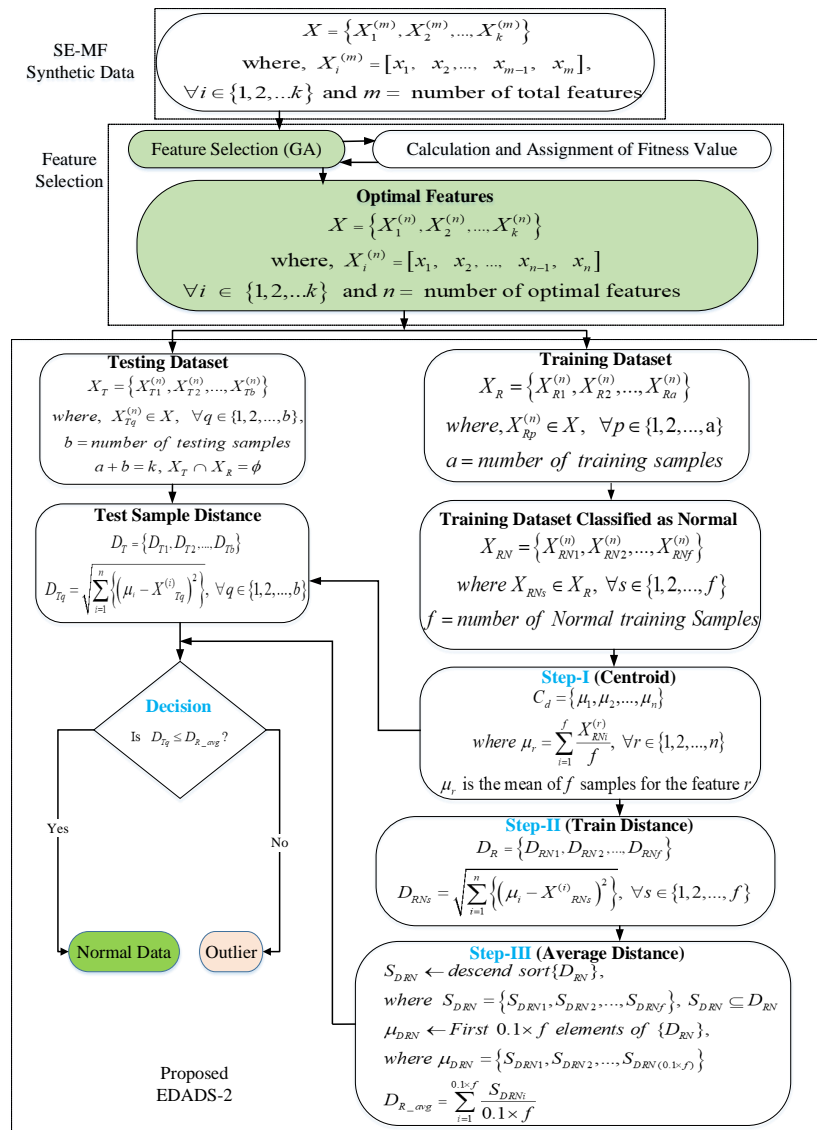


Figure 4. Flowchart of the Euclidean distance-based anomaly detection Scheme 2, EDADS-2.

4. Experimental Results

In this section, we evaluate the performance of the proposed EDADS-1 and EDADS-2. We performed the simulations using MATLAB 2017b. The proposed schemes were evaluated through experiments using the standard 14-bus, 39-bus, 57-bus and 118-bus IEEE test systems. Experiments' results have been averaged over 20 iterations for each case bus system. Figure 5 illustrates the IEEE 39-bus system [38], also known as the New England 10-machine system. Because of space limitations, figures for the other IEEE bus systems employed for testing in this work are not included. To simulate the operation of the power network, we used the Matpower 6.0 toolbox [39] to generate the configuration of these test systems (especially the Jacobian matrix). We employed the AC power flow model and used DC power flow analysis to approximate the state vectors and measurement dataset. In a B -bus system, state variable vector $\delta \in \mathbb{R}^n$ is composed of $(B - 1)$ bus voltage phase angles, and the meter measurement vector consists of active power injections into the buses and branch active power flows. To conduct a fair comparison with a real-world power network scenario, we used the stochastic loads with uniform load distributions similar to [7], i.e., in the range $[0.9 \times B_0 - 1.1 \times B_0]$, where B_0 is the base load. In these simulations, the active power measurement features, including the

active power injections into the buses and active power flows on the branches, are the input to the GA for the selection of optimal features.

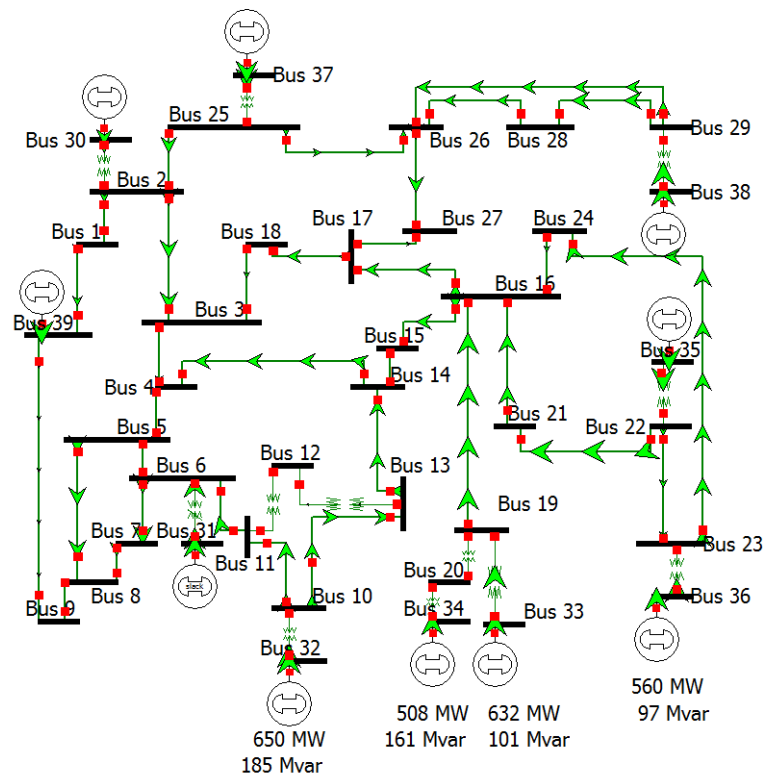


Figure 5. Standard IEEE 39-bus system (New England 10-Machine System [38]).

4.1. GA-Based Feature Selection

In the paper, we set the number of chromosomes in each population as 50 and the maximum number of generations as 80, respectively, for GA-based feature selection. Stochastic universal sampling (SUS) is used as the selection operator, and uniform crossover is employed. The crossover rate is 0.63 and the mutation rate 0.018. Because the GA randomly selects different subsets of the SE-MF dataset to create a primary population of chromosomes, we iterate the GA 30 times and choose only those features that were selected more than 70% in 30 iterations. Table 3 lists the average number of selected features from the application of the GA to the SE-MF dataset for various IEEE standard systems.

4.2. 3D Representation of the Proposed EDADS-1 and EDADS-2

In this subsection, a 3D pictorial representation of the proposed schemes (EDADS-1 and EDADS-2) is illustrated with Figures 6 and 7, respectively. To represent the workings of the proposed schemes, we use three features of the standard IEEE 57- and 118-bus systems. In Figure 6, the centroid (indicated in the figure) is calculated using training data consisting of normal and compromised optimal features. The average of the distances of all the samples from the centroid defines a virtual boundary, as shown in Figure 6. The data points lying outside of the virtual boundary are termed anomalous data or outliers. However, in Figure 7, the centroid (indicated in the figure) is calculated on the basis of data labeled as normal only. The distances of all the samples from the centroid are sorted in descending order, and then, the first 10 percent distances (farthest from the centroid) are chosen to calculate their average distance from the centroid. The average distance defines a virtual boundary as shown in Figure 7. The samples lying out of the boundary are considered as outliers. Basic performance metrics used in this work, i.e., accuracy, F_1 score and receiver operating characteristic (ROC) curves, are shown in the following subsections.

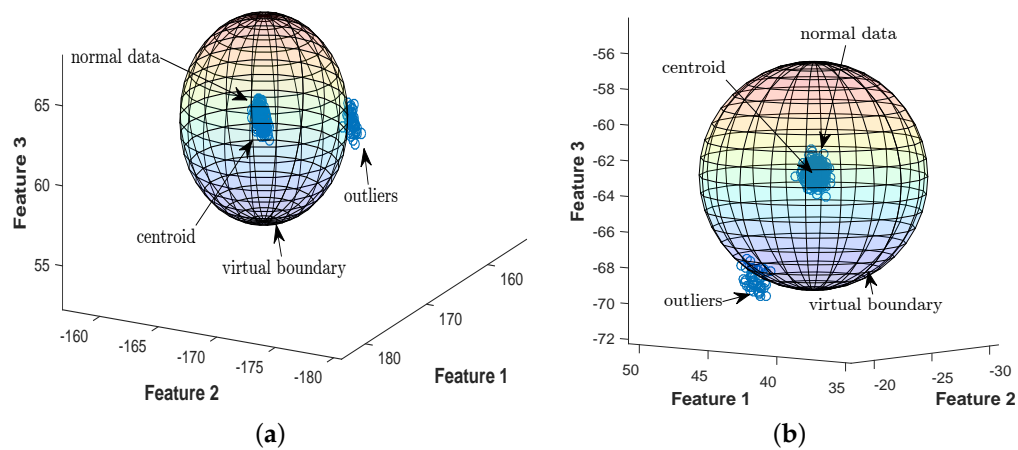


Figure 6. Three-dimensional representation of the proposed EDADS-1 with three features for standard IEEE 14- and 39-bus systems. (a) IEEE 14-bus system; (b) IEEE 39-bus system.

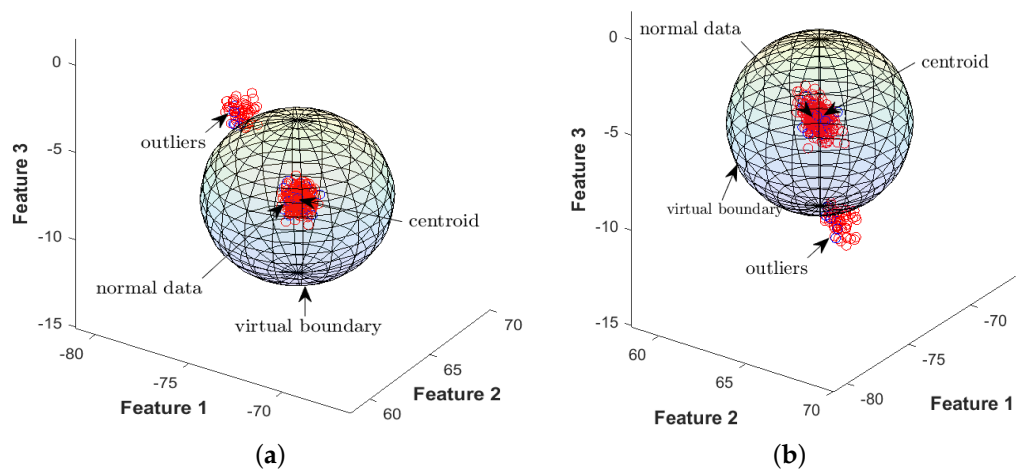


Figure 7. Three-dimensional representation of the proposed EDADS-2 with three features for standard IEEE 57- and 118-bus systems. (a) IEEE 57-bus system; (b) IEEE 118-bus system.

4.3. Receiver Operating Characteristic Curves

Figures 8 and 9 illustrate the ROC curves for the proposed EDADS-1 and EDADS-2, respectively, employing the standard IEEE 14-, 39-, 57- and 118-bus systems for testing. The ROC curve is obtained by plotting the false positive rate (FPR) versus the true positive rate (TPR). FPR is defined as the probability that normal data are identified as compromised. It is used as a measure of specificity in our detection scheme. The sensitivity of our scheme is defined as the probability that compromised data are identified as assaulted. TPR is used as a measure of sensitivity. From Figures 8 and 9, we can see that the area under the curve is closer to 1one in all cases. This means that the detection accuracy of the proposed schemes is near one, which validates its good performance. In the next subsections, we elaborate on the accuracy and F_1 score for the proposed schemes.

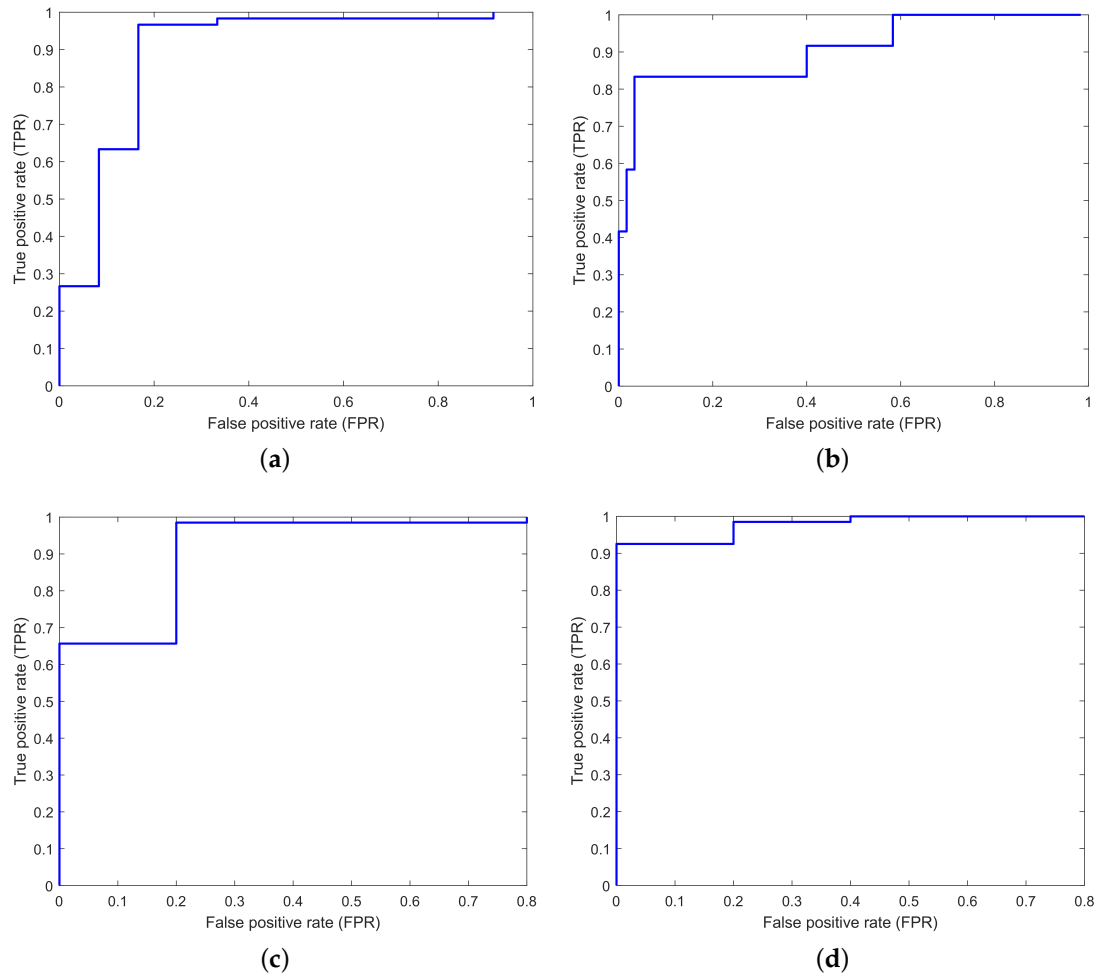


Figure 8. The ROC curves of the proposed EDADS-1 in standard IEEE 14-, 39-, 57- and 118-bus systems. (a) IEEE 14-bus system; (b) IEEE 39-bus system; (c) IEEE 57-bus system; (d) IEEE 118-bus system.

4.4. Accuracy

Calculating the accuracy is a standard way to evaluate the anomaly detection algorithms. It is a single-number summary of the performance of the proposed algorithm and can be calculated as follows:

$$Accuracy = \left(\frac{\sum TP + \sum TN}{TotalPopulation} \right), \quad (20)$$

where true positive (TP) corresponds to the samples that the proposed algorithm detects as positive samples and that are, in fact, positive. Similarly, true negatives (TNs) are the points that the proposed algorithm detects as negative samples and that are, in fact, negative.

Figure 10 shows the accuracy of the proposed schemes (EDADS-1 and EDADS-2) for various IEEE standard bus systems according to a varying number of training samples. The efficiency of the learning algorithm can be improved by increasing the amount of learning data. We compare the performance of the proposed scheme with that of the statistical model-based anomaly detection method [7] in which the FE technique for dimensionality reduction was utilized and further that of the neighborhood component analysis (NCA) technique, respectively. NCA is a supervised learning method for classifying multivariate data into distinct classes according to a given distance metric over the data. The results show that the proposed FS-based schemes (EDADS-1 and EDADS-2) have higher CCD assault detection accuracy. EDADS-2 exhibits slightly higher performance than EDADS-1 since it only employs the normal features for training, utilizing the labeled data. Hence, the average of all

the distances of test samples is close to the value that is required to accurately separate the normal class from the compromised one. It is also shown from Figure 10 that accuracy in detection increases with the increasing number of training data samples. In addition, the NCA-based FS technique has low detection accuracy as compared to the proposed schemes. However, the NCA-based FS scheme performs better than the statistical-based method [7].

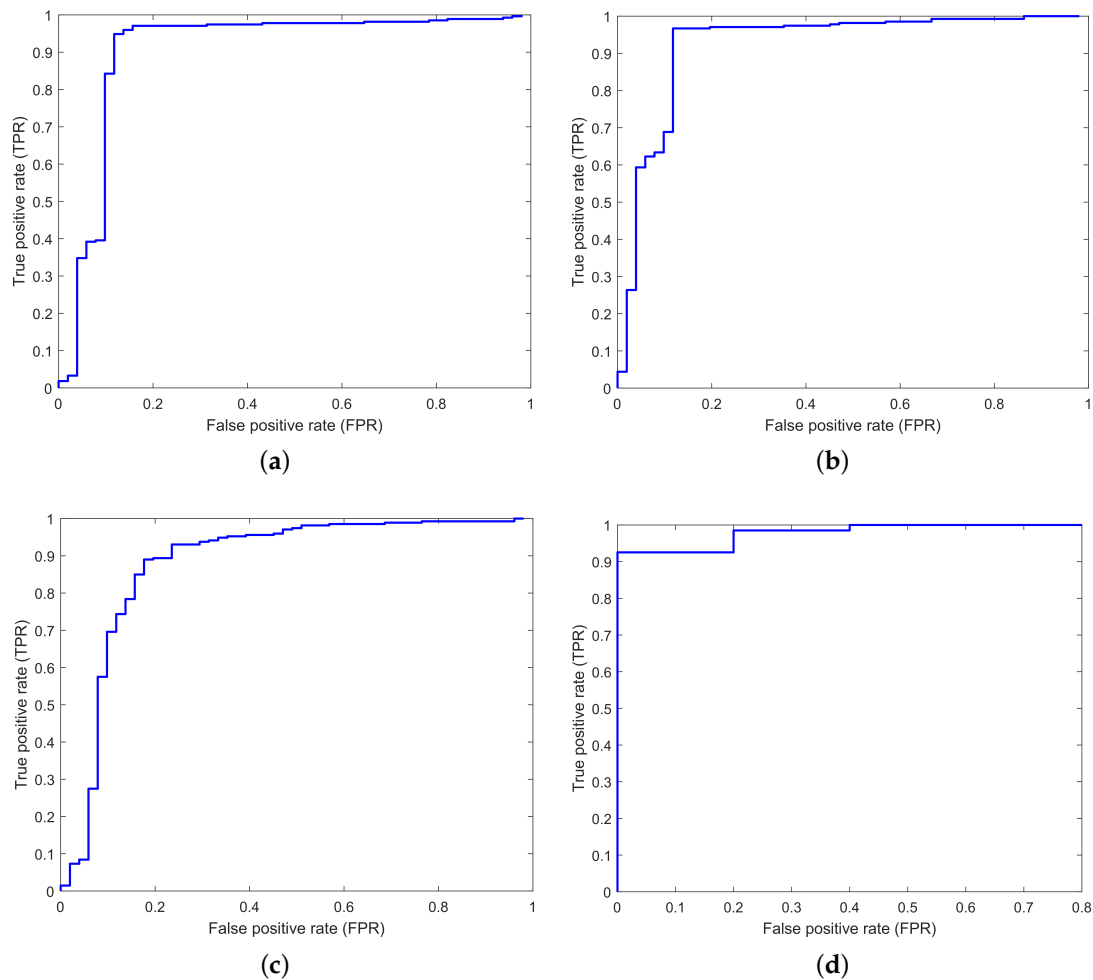


Figure 9. The ROC curves of the proposed EDADS-2 in standard IEEE 14-, 39-, 57- and 118-bus systems. (a) IEEE 14-bus system; (b) IEEE 39-bus system; (c) IEEE 57-bus system; (d) IEEE 118-bus system.

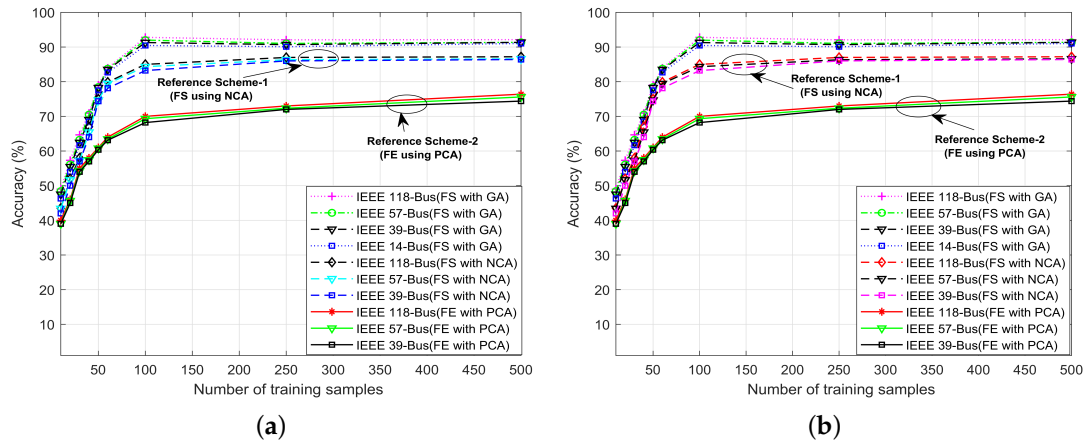


Figure 10. The CCD assault detection accuracy of the proposed EDADS-1 and EDADS-2 in standard IEEE 14-, 39-, 57- and 118-bus systems with varying number of training samples. (a) Accuracy of EDADS-1; (b) accuracy of EDADS-2.

4.5. F_1 score

Next, we utilize the F_1 score as another metric of detection accuracy. The F_1 score is considered a measure of the precise detection or classification of the subject dataset. The F_1 score is obtained as follows:

$$F_1 = 2 \left(\frac{P_r \times R_e}{P_r + R} \right), \quad (21)$$

where P_r is precision and is calculated as follows:

$$P_r = \left(\frac{\text{TruePositive}}{\text{PredictedPositive}} \right). \quad (22)$$

True positive corresponds to samples that the proposed algorithm detects as positive samples and that are, in fact, positive. Predicted positives may include both compromised and normal sample points, but the algorithm detects them all as positive. R_e is recall, calculated as follows:

$$R_e = \left(\frac{\text{TruePositive}}{\text{ActualPositive}} \right). \quad (23)$$

Figure 11 shows the F_1 score of the proposed schemes for different IEEE standard bus systems. The F_1 score of the statistical model-based anomaly detection method [7] employing FE for dimensionality reduction is included for comparison. The proposed schemes are also compared with the NCA-based scheme. It is obvious that the proposed FS-based schemes have a higher F_1 score for all test cases, whereas the FE-based scheme requires many historical samples from the SE-MF dataset for learning to achieve a higher F_1 score. The proposed EDADS-2 has a higher F_1 score than EDADS-1 due to the reason that it employs only normal data samples for training. Hence, the average of the distances of the normal training samples from their centroid is close to the value required for accurately separating the normal class from the compromised class. The performance of the NCA-based scheme is lower than that of proposed schemes; however, it performs better compared to the FE-based scheme [7].

Next, to investigate the impact of several compromised load profiles (compromised samples), we consider different numbers of compromised load profiles, i.e., 24, 30, 36, 40, 45 and 60. The SE-MF dataset load profile is comprised of 360 samples collected through sensors or RTUs at regular intervals of four minutes over 24 h. We use 75% of the data for training and the rest of the samples for testing. Figure 12 shows the F_1 score as a measure of the accuracy of the proposed FS-based proposed detection

schemes for standard IEEE 14-, 39-, 57- and 118-bus systems. Figure 12 shows that the FS-based proposed methods have an accuracy of more than 90% for all the employed test systems.

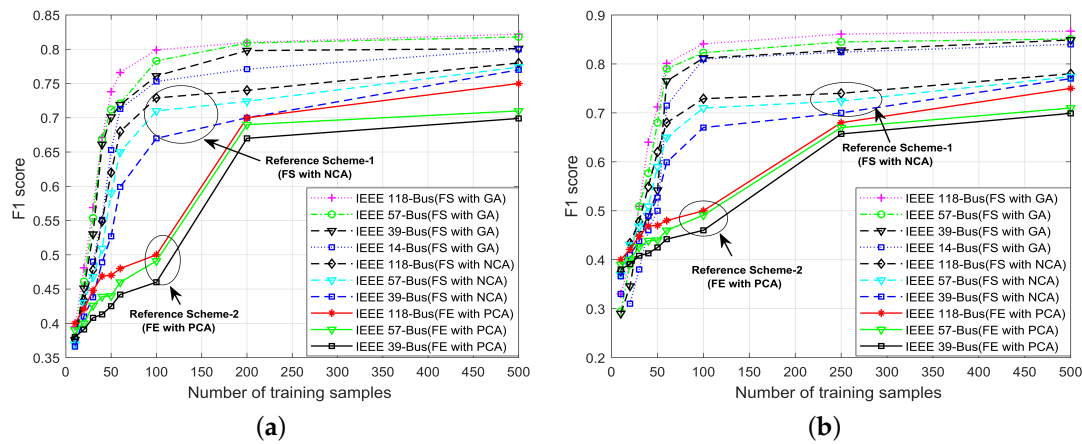


Figure 11. F_1 score of the proposed EDADS-1 and EDADS-2 in standard IEEE 14-, 39-, 57- and 118-bus systems with varying numbers of training samples. (a) F_1 score of EDADS-1; (b) F_1 score of EDADS-2.

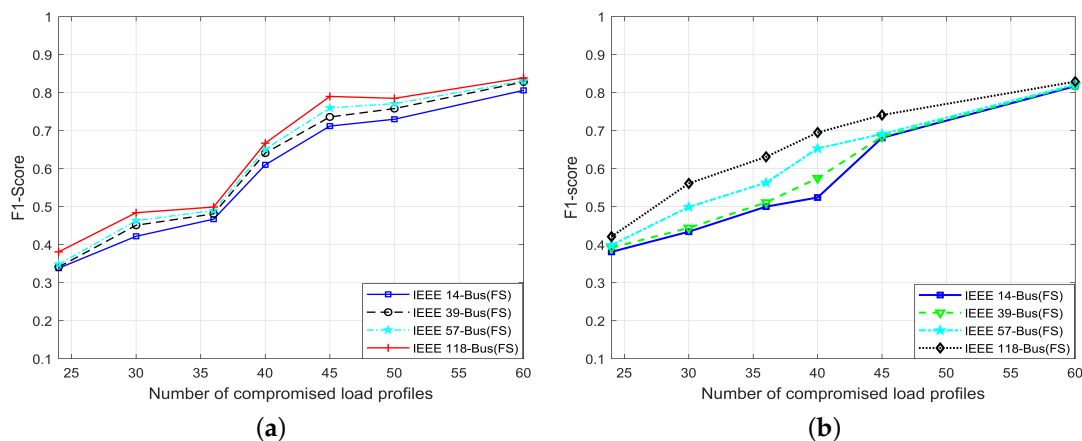


Figure 12. F_1 score of the proposed EDADS-1 and EDADS-2 in standard IEEE 14-, 39-, 57- and 118-bus systems with varying numbers of compromised load profiles. (a) EDADS-1; (b) EDADS-2.

4.6. Execution Time Comparison

In this subsection, we compare the execution time of the proposed schemes with that of the existing schemes. Table 3 shows that the proposed schemes (EDADS-1 and EDADS-2) consume less time for feature selection and anomaly detection as compared to the existing schemes. The feature selection time of the two proposed schemes is similar because both schemes utilize GA. However, the detection time of EDADS-2 is slightly less than that of EDADS-1. The reason is that EDADS-2 utilizes only normal features for training from the test data. On the other hand, EDADS-1 employs all the features from the test data for training because the data are unlabeled for this case. Table 3 shows that for NCA, the execution time is higher than the proposed schemes. Table 3 clearly shows that the FE technique [7] employing PCA requires more time. The PCA-based approach achieves the dimensionality reduction by transforming original samples with binary values into new samples with numeric values while making the execution time much longer than other schemes.

Table 3. Comparison of the CPU time for different schemes.

Standard IEEE Bus System	Proposed EDADS-1 (FS + Detection)	Proposed EDADS-2 (FS + Detection)	FE with PCA FE + Detection	FS with NCA
14	0.1149 (s)	0.1147 (s)	2.0134 (s)	0.3125 (s)
39	0.1316 (s)	0.1299 (s)	4.3417 (s)	0.5313 (s)
57	0.2322 (s)	0.2317 (s)	7.5121 (s)	0.8013 (s)
118	0.5179 (s)	0.5177 (s)	10.7925 (s)	1.0130 (s)

5. Conclusions

In this paper, we propose two FS-based anomaly detection schemes for the detection of CCD assaults in SG communications networks. In the proposed schemes, GA is employed for the selection of discriminative and distinguishing features from historical SE-MF datasets. The selected optimal features are used as the input for two Euclidean distance-based anomaly detection schemes (EDADS-1 for unlabeled data and EDADS-2 for labeled data) to detect anomalies/outliers in the smart-grid SE measurement samples. To validate the performance of the proposed schemes, we utilize the standard IEEE 14-bus, 39-bus, 57-bus and 118-bus systems. In addition, we utilize data that are collected from active power injections into the buses and active power flow measurements in the branches as the learning data and study the accuracy of our detection methods under CCD attack. The test results show that the proposed ED-based FS schemes have reasonably improved detection accuracy, compared to PCA-based FE and NCA-based FS schemes in the occasional operational environment. The low computational complexity of the proposed schemes enables the identification of outliers or anomalies in a short time. In the future, we will model our work by considering more diverse attack scenarios, and we will incorporate a learning mechanism to automatically update the Euclidean distances with incoming test data to improve the detection accuracy.

Author Contributions: All authors conceived of and proposed the research idea. S.A. and Y.L. designed the experiments. S.A. performed the experiments. S.-H.H. and I.K. analyzed the experimental results. S.A. wrote the paper under the supervision of S.-H.H. and I.K.

Acknowledgments: This work was supported by the 2018 Research Fund of University of Ulsan.

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

References

1. Liu, J.; Xiao, Y.; Li, S.; Liang, W.; Chen, C.P. Cyber security and privacy issues in smart grids. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 981–997. [[CrossRef](#)]
2. Mo, Y.; Kim, T.H.J.; Brancik, K.; Dickinson, D.; Lee, H.; Perrig, A.; Sinopoli, B. Cyber-physical security of a smart grid infrastructure. *Proc. IEEE* **2012**, *100*, 195–209.
3. Yan, Y.; Qian, Y.; Sharif, H.; Tipper, D. A survey on cyber security for smart grid communications. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 998–1010. [[CrossRef](#)]
4. 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. Ind. Inform.* **2017**, *13*, 411–423. [[CrossRef](#)]
5. Ozay, M.; Esnaola, I.; Vural, F.T.Y.; Kulkarni, S.R.; Poor, H.V. Machine Learning Methods for Attack Detection in the Smart Grid. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1773–1786. [[CrossRef](#)] [[PubMed](#)]
6. Singh, S.K.; Khanna, K.; Bose, R.; Panigrahi, B.K.; Joshi, A. Joint-Transformation-Based Detection of False Data Injection Attacks in Smart Grid. *IEEE Trans. Ind. Inform.* **2018**, *14*, 89–97. [[CrossRef](#)]
7. Esmalifalak, M.; Liu, L.; Nguyen, N.; Zheng, R.; Han, Z. Detecting stealthy false data injection using machine learning in smart grid. In Proceedings of the 2013 IEEE Global Communications Conference, Atlanta, GA, USA, 9–13 December 2013.
8. Bishop, C.M. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, UK, 1995.

9. Khan, F.; ur Rehman, A.; Arif, M.; Aftab, M.; Jadoon, B.K. A survey of communication technologies for smart grid connectivity. In Proceedings of the 2016 International Conference on Computing, Electronic and Electrical Engineering, Quetta, Pakistan, 11–12 April 2016; pp. 256–261.
10. Gungor, V.C.; Lambert, F.C. A survey on communication networks for electric system automation. *Comput. Netw.* **2006**, *50*, 877–897. [[CrossRef](#)]
11. Wang, W.; Xu, Y.; Khanna, M. A survey on the communication architectures in smart grid. *Comput. Netw.* **2011**, *55*, 3604–3629. [[CrossRef](#)]
12. Emmanuel, M.; Rayudu, R. Communication technologies for smart grid applications: A survey. *J. Netw. Comput. Appl.* **2016**, *74*, 133–148. [[CrossRef](#)]
13. Wang, W.; Lu, Z. Cyber security in the smart grid: Survey and challenges. *Comput. Netw.* **2013**, *57*, 1344–1371. [[CrossRef](#)]
14. Bao, H.; Lu, R.; Li, B.; Deng, R. BLITHE: Behavior rule-based insider threat detection for smart grid. *IEEE Internet Things J.* **2016**, *3*, 190–205. [[CrossRef](#)]
15. Iqbal, S.; Kiah, M.L.M.; Dhaghighi, B.; Hussain, M.; Khan, S.; Khan, M.K.; Choo, K.K.R. On cloud security attacks: A taxonomy and intrusion detection and prevention as a service. *J. Netw. Comput. Appl.* **2016**, *74*, 98–120. [[CrossRef](#)]
16. Li, B.; Lu, R.; Wang, W.; Choo, K.K.R. DDOA: A dirichlet-based detection scheme for opportunistic attacks in smart grid cyber-physical system. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 2415–2425. [[CrossRef](#)]
17. Liu, Y.; Ning, P.; Reiter, M.K. False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.* **2011**, *14*, 13. [[CrossRef](#)]
18. Li, B.; Lu, R.; Wang, W.; Choo, K.K.R. Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system. *J. Parallel Distrib. Comput.* **2017**, *103*, 32–41. [[CrossRef](#)]
19. Huang, Y.; Tang, J.; Cheng, Y.; Li, H.; Campbell, K.A.; Han, Z. Real-time detection of false data injection in smart grid networks: an adaptive CUSUM method and analysis. *IEEE Syst. J.* **2016**, *10*, 532–543. [[CrossRef](#)]
20. Ferrag, M.A.; Maglaras, L.A.; Janicke, H.; Jiang, J. A survey on privacy-preserving schemes for smart grid communications. *arXiv* **2016**, arXiv:1611.07722. [[CrossRef](#)]
21. Baumeister, T. *Literature Review on Smart Grid Cyber Security*; Collaborative Software Development Laboratory at the University of Hawaii: Honolulu, HI, USA, 2010.
22. Xie, L.; Mo, Y.; Sinopoli, B. Integrity data attacks in power market operations. *IEEE Trans. Smart Grid* **2011**, *2*, 659–666. [[CrossRef](#)]
23. Esmalifalak, M.; Han, Z.; Song, L. Effect of stealthy bad data injection on network congestion in market based power system. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference, Shanghai, China, 1–4 April 2012; pp. 2468–2472.
24. Dan, G.; Sandberg, H. Stealth attacks and protection schemes for state estimators in power systems. In Proceedings of the 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 4–6 October 2010; pp. 214–219.
25. Kundur, P.; Balu, N.J.; Lauby, M.G. *Power System Stability and Control*; McGraw-Hill: New York, NK, USA, 1994; Volume 7.
26. Gomez-Exposito, A.; Abur, A. *Power System State Estimation: Theory and Implementation*; CRC Press: Boca Raton, FL, USA, 2004.
27. Casazza, J.; Casazza, J.; Delea, F. *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*; John Wiley & Sons: New York, NK, USA, 2003; Volume 13.
28. Monticelli, A. *State Estimation in Electric Power Systems: A Generalized Approach*; Springer Science & Business Media: Berlin, Germany, 1999; Volume 507.
29. Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [[CrossRef](#)] [[PubMed](#)]
30. Min, S.H.; Lee, J.; Han, I. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Syst. Appl.* **2006**, *31*, 652–660. [[CrossRef](#)]
31. Demel, M.A.; Janecek, A.G.; Thai, K.M.; Ecker, G.F.; Gansterer, W.N. Predictive QSAR models for polyspecific drug targets: The importance of feature selection. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 91–110. [[CrossRef](#)]
32. Ma, S.; Song, X.; Huang, J. Supervised group Lasso with applications to microarray data analysis. *BMC Bioinform.* **2007**, *8*, 60. [[CrossRef](#)] [[PubMed](#)]

33. Ambroise, C.; McLachlan, G.J. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6562–6566. [[CrossRef](#)] [[PubMed](#)]
34. Jafari, P.; Azuaje, F. An assessment of recently published gene expression data analyses: Reporting experimental design and statistical factors. *BMC Med. Inform. Decis. Making* **2006**, *6*, 27. [[CrossRef](#)] [[PubMed](#)]
35. Hruschka, E.R.; Hruschka, E.R.; Ebecken, N.F. Feature selection by Bayesian networks. In *Conference of the Canadian Society for Computational Studies of Intelligence*; Springer: Berlin, Germany, 2004; pp. 370–379.
36. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271. [[CrossRef](#)]
37. Chandala, V.; Banerjee, A.; Kumar, V. *Anomaly Detection: A Survey*, *ACM Computing Surveys*; University of Minnesota: Minneapolis, MN, USA, 2009.
38. Illinois Center for a Smarter Electric Grid (ICSEG). Available online: <http://icseg.iti.illinois.edu/> (accessed on 4 february 2018).
39. Zimmerman, R.D.; Murillo-Sánchez, C.E.; Thomas, R.J. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).