Robust Feature Extraction and Ensemble Classification Against Cyber-Physical Attacks in the Smart Grid

Chengming Hu, Jun Yan, and Chun Wang
Concordia Institute for Information Systems Engineering (CIISE), Concordia University
Montréal, Québec, Canada
h_chengm@encs.concordia.ca; jun.yan@concordia.ca; chun.wang@concordia.ca

***Abstract*—Intrusion detection systems (IDS) are crucial in threats monitoring for the cyber-physical security of electrical power and energy systems in the smart grid with increasing machine-to-machine communication. However, the multi-sourced, voluminous, correlated, and often noise-contained data, which record various concurring cyber and physical events, are posing significant challenges to the accurate distinction by IDS among events of inadvertent and malignant natures. To tackle such challenges, this paper proposes a robust end-to-end framework based on Stacked Denoising Autoencoder (SDAE) and Ensemble Machine Learning to extract new noise and attack-informed feature sets from cyber-physical system data and incorporate different sources of information for reliable event classification. The proposed framework first leverages SDAE to create lower-dimensional features that allow reconstruction of a noise-free input from noise-corrupted perturbations. By combining attack and noisy inputs, we extracted new, automatically-engineered features that can preserve and present information on normal, fault, and attack events against different synthetic but realistic noises for better classification. Considering the heterogeneous nature of the inputs, which are composed of PMU measurements, system logs, and IDS alerts, we further introduced ensemble learning-based multi-classifier classification with the Extreme Gradient Boosting (XGBoost) technique to classify the samples based on the SDAE-extracted features. Normalization and oversampling were also both performed to improve the uniformity and balance of the data. On a realistic dataset of 37 sub-types of normal, fault, and attack collected from co-simulations on a hardware-in-the-loop (HIL) testbed security testbed, the results have shown that the proposed SDAE+XGBoost solution achieves over 90% classification accuracy with the SDAE features and ensemble classifiers, an effective 8% increase over the state-of-the-art.**

I. INTRODUCTION

A. Background

The recent efforts in grid modernization is leading towards a cyber-physical smart grid with increasing and stringent requirements on resiliency and security [1]. From the perspective of cybersecurity, the smart grid exposes various physical vulnerabilities and potential benefits to exploit. In order to improve the smart grid security, advanced IDS is proposed to which can apply machine learning techniques to classify various events based on improved models of systems and events [2]–[4].

However, with the increasing communication and complexity in the real cyber-physical environment [5], the voluminous, heterogeneous and noise-corrupted data are generated which represent normal, fault and attack events in the smart grids. The voluminous data are posing challenges to the effective event classification, since the redundant information covers the useful knowledge and hidden patterns in the distinction by IDS. One typical method to handle these challenges is feature extraction which can extract new relatively low-dimensional feature space that represents original feature information. The new feature set can often be used for some supervised learning-based tasks, such as classification or regression. Due to high-dimensional feature space in the fast-expanding system, traditional feature extraction methods remain challenging to learn new robust feature set from cyber-physical system data; as a result, the highly-representative features cannot preserve original information on normal, fault and attack events, which leads the severe misclassification in the IDS. The limitation of such current consequences is calling for robust feature extraction methods that can produce more discriminative features for further accurate event classification. Besides, considering the multi-sourced nature of the data that contain PMU measurements and system logs, there is also great necessary to introduce more advanced classifiers that can better distinguish among normal operations, inadvertent faults and adversarial attacks.

B. State-of-the-Art

The cyber-physical smart grid has suffered the physical sabotages on power lines, transformers, among others; cyber-espionage and cyber-attack have also been reported in real world incidences [2]. These challenges have been driving the research in cyber-physical attacks and defences of the smart grid from information [6], power [7], control [8], and many other security-related research communities, leading to a unified cyber-physical security perspective.

In response to these challenges, machine learning techniques have been applied to address the voluminous data and develop accurate event classifiers and detectors in the IDS. The common paths mining-based detection [9] can achieve overall 90.4% accuracy after feature selection based on expertise while capturing data logs to build such common paths is difficult for real systems. With the dataset [10] that will be discussed in the later section, the Oak Ridge National Laboratory randomly sampled the dataset at 1% to reduce the size and conducted a comparative study among

978-1-7281-3406-2/19/$31.00 ©2019 IEEE

different classifiers [11], including OneR, Nearest Neighbor, Random Forest, Support Vector Machine, Naïve Bayes, JRip, Adaboost. The best accuracy was around 95% achieved by Adaboost classifier based on information gain-selected features. However, the results were obtained on a randomly sampled subset with only 1% of the entire data for computational efficiency, which may not reflect the accurate performance overall in more comprehensive scenarios. The similar problem may also exist in the algorithms Linear Weighted Cuckoo Search Optimization (LWCSO) feature selection and Probabilistic Kernel Model (PKM) classification [12] that were combined to achieve 98.9% accuracy with the limited data (5,069 samples overall) in the detection of abnormal events. Wilson *et al.* [13] applied a multilayer perceptron (MLP) classifier based on the stacked autoencoder-extracted features. The accuracy against each type of events could be over 96.79% and 97.34% after converting 128 original features to 64 and 32 new extracted features, respectively. However, their framework did not consider the potential imbalance of training samples in practical scenarios, which could degrade the classification performance.

Inspired by the recent progress of feature learning and ensemble learning techniques in real-world applications, this paper introduces a robust framework based on stacked denoising autoencoder (SDAE) [14] and ensemble learning-based classifier using extreme gradient boosting (XGBoost) [15]. The unsupervised learning-based SDAE can automatically learn highly-representative feature sets by reconstruction of noise-free inputs from noise-contained data. We further apply XGBoost-based event classifier that combine multiple Classification and Regression Tree (CART) [16] to classify samples based on the SDAE-extracted features. Considering the uniformity and balance of the raw data, data normalization and oversampling also need to be adopted as data preprocessing procedure.

The rest of paper is organized as follows: Section II describes an overview of basic autoencoder (AE) and stacked denoising autoencoder (SDAE) as robust feature extraction method, and XGBoost algorithm as ensemble learning-based classifier. Section III introduces the power system benchmark and the experiment setup, and the simulation result is given at the end. Section IV discusses potential improvements on the architecture of proposed feature extraction, and the conclusion is drawn in Section V.

II. Robust Feature Extraction with SDAE

Feature extraction is a widely-used technique for attack signatures and dimensionality reduction. It projects original data, high-dimensional space to a relatively low-dimensional space and this transformation can be liner or nonlinear. The objective function of feature extraction method is to minimize the difference between the original space and the new highly-representative space so that the useful information can be mapped into a low-dimensional feature space. Classification tasks can be effectively performed with new extracted feature space. In this section, we introduce autoencoder (AE) and SDAE-based feature extraction methods.

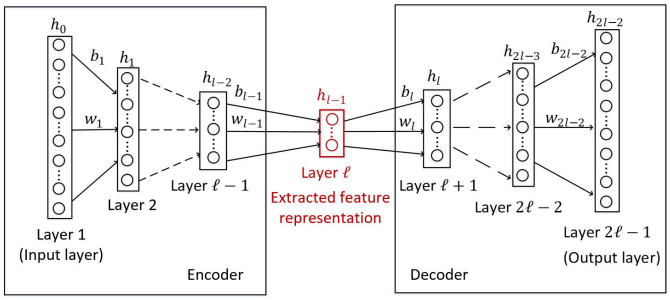

Fig. 1. The architecture of AE for feature extraction.

A. Autoencoder (AE)

AE is one of the more popular unsupervised feature learning algorithm in several IDS detection studies [13], [17], [18], which applies artificial neural network to learn deep and abstract features from original inputs, typically for feature extraction. The AE architecture is shown in Fig. 1, where original features as input h_0 in the first layer are extracted with multiple hidden layers and eventually are reconstructed at the output layer h_{2l-2}.

The training process consists of two parts: an encoder and a decoder. The encoder compresses original features into a small representation h_{l-1} by weights $W = [w_1, w_2, \dots, w_{l-1}]$ and bias $b = [b_1, b_2, \dots, b_{l-1}]$ till the middle l-th layer; the decoder reconstructs output layer from the hidden representation h_{l-1} by the other weights $W' = [w_l, w_{l+1}, \dots, w_{2l-2}]$ and bias $b' = [b_l, b_{l+1}, \dots, b_{2l-2}]$. Given a dataset x_n with n features, the feature vector h_i at the layer $i+1$ can be calculated which is as follows:

$$h_i = f(w_i h_{i-1} + b_i), \quad i = 1, 2, ..., 2l-2 \tag{1}$$

where $f(.)$ is the activation function of the layer h_i and the network output h_{2l-2} can also be represented as $\hat{x}_i$.

To optimize AE, the reconstruction error between the input and output should be minimized as follows:

$$\min \frac{1}{n} \sum_{i=1}^{n} l(x_i, \hat{x}_i) \tag{2}$$

where $l(x_i, \hat{x}_i) = |x_i - \hat{x}_i|^2$ is the reconstruction error of the i-th feature.

The objective function is optimized by tuning all weights and bias in the back propagation way. Once the network parameters are determined, the middle hidden layer h_{l-1} can represent the original feature information and be employed as new extracted feature sets.

B. Stacked Denoising Autoencoder (SDAE)

Considering the increasing communication and complexity in the real environment, samples are often disturbed with stochastic noises while basic AE is hard to reconstruct the noise-free inputs from noise-corrupted perturbations. Besides, AE could cause the obvious solution by identity mapping or similarly uninteresting ones that trivially maximizes mutual information as well [14]. To find out a solution, Vincent *et*

al. [14] proposed denoising autoencoder (DAE) as the variant of basic AE to discover more robust features by additionally introducing the denoising criterion in the reconstruction.

Instead of noise-free inputs x_n in the basic AE, DAE disturbs the original inputs by the means of a stochastic mapping $\tilde{x}_n \sim q_D(\tilde{x}_n|x_n)$. The new extracted feature vector is determined by applying activation function to the corrupted inputs rather than the original inputs. Hence the objective function is to minimize the reconstruction error between the original noise-free inputs and the reconstruction outputs from noise-contained inputs.

As a neural network-based feature extraction method, DAE can be handily stacked to generate different levels of new feature representations of the original data, by iteratively adding new hidden layers after the previous one(s) has been trained and fixed, with the aim of discovering highly-nonlinear and complex patterns in the data [19]–[21]. In this paper, we introduce DAE as the basic architecture and stack multiple DAEs to form our deep model SDAE.

The overall architecture and training process of SDAE is shown in the Fig. 2, where several DAEs are treated as individual blocks stacked in the deep architecture. The first DAE is trained to reconstruct the raw data x_0 from the disturbed input $\tilde{x}_0$ including random Gaussian noise n_0, where $\tilde{x}_0 = x_0 + n_0, n_0 \sim N(0,1)$. We measure the difference between reconstruction result $\hat{x}_0$ and the raw data x_0 as the reconstruction error e_0 at the first training iteration, and the parameters $w_0, w_0^{'}, b_0, b_0^{'}$ are continuously tuned with the optimization of reconstruction error in the back propagation way [22]. It is notable that there are no noises in the testing process and DAE directly extracts features from the original feature x_0.

Once the first DAE is completely built, the hidden layer x_1 is its new extracted feature vector which will be then combined with random Gaussian noise n_1 for the training of the second DAE. The overall training and testing process is similar with that of the first DAE, and eventually the extracted features x_2 can be determined with optimization of the network parameters. Repeat the above process, each successor DAE reconstructs the extracted features of predecessor DAE from its disturbed data with Gaussian noise in the training, and the low-dimensional feature space can be generated as the training and testing process is completed. After building all individual DAEs, all highly-representative hidden layers are stacked to form the SDAE model whose extracted features can be conducted in several supervised learning algorithms.

C. Ensemble Learning-Based Classifier

Ensemble learning is a machine learning algorithm that trains multiple learners and combines their decisions to solve problems such as classification, prediction, among others. An ensemble consists of several learners called weak or base learner, and their combination can form a strong learner for the task. The base learners can be of homogeneous or heterogeneous types. By combine multiple base learners systematically, the ensemble learner is able to integrate the advantages and diversities among base learners and minimize

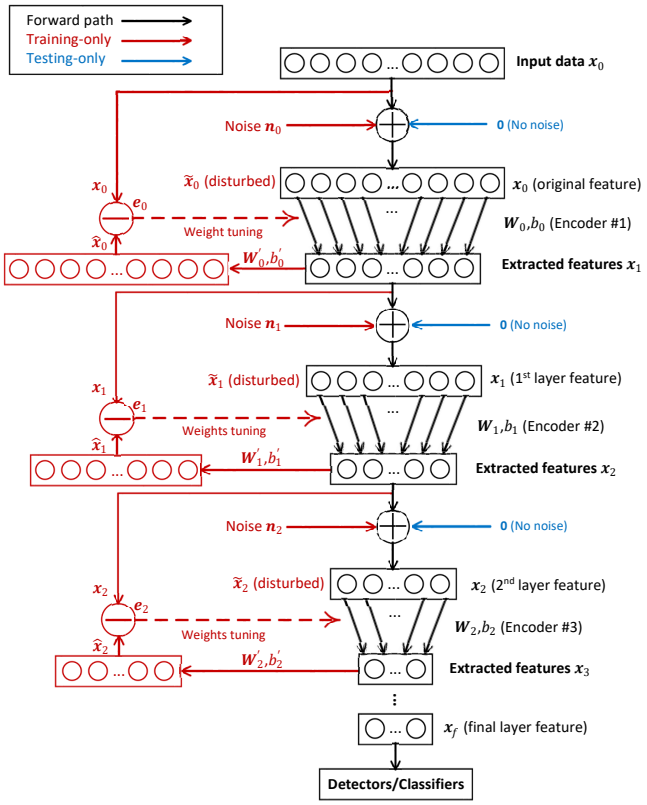

Fig. 2. The overall architecture and training process of SDAE.

the drawbacks and misclassifications as well, so that the performance of single learner can be boosted in the final decision.

In this paper, we consider XGBoost [15], [23] ensemble learning-based classifier after SDAE feature extraction, which is one of the typical ensemble learning algorithms that witness an increasing adoption in real-world applications. The Classification and Regression Tree (CART) [16] is regarded as the base learner due to the ability of learning from both discrete and continuous feature values, handling sparse data and instance weight efficiently [24] and parallel and distributed implementation [25], [26].

The objective function of XGBoost classifier includes the loss function to measure the difference between the predicted label and the target. Besides, compared to traditional gradient tree boosting, the XGBoost objective function also has the additional regularization item which represents the model complexity and removes the need to prune trees after they are built. XGBoost classifier is trained in an addictive way which discovers the optimized tree to minimize the objective function at each iteration rather than after the whole training process.

III. Simulations and Results

A. Benchmark System and Dataset

Fig. 3 shows the power system framework, which is a three-bus two-machine system with four circuit breakers [10]. Each breaker is controlled by an Intelligent Electronic Device

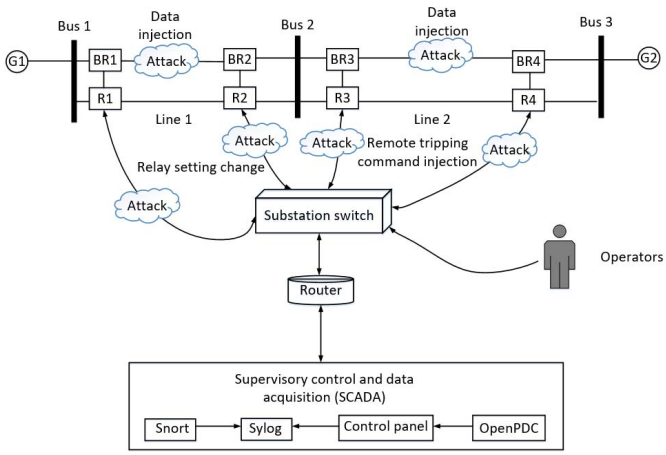

Fig. 3. The benchmark power system [10].

TABLE I
THE CYBER-PHYSICAL SMART GRID ATTACK DATASET

Class	Description	Scenarios	Scenario IDs
0	No Event	Normal operation with load demand variation	41
1	Natural Event	Single line-to-ground faults (SLG) and line maintenance	1–6,13,14
2	Attack Event	Data injection, remote tripping, command injection, relay settings change	7–12,15–30, 35–40

(IED) that applies a distance protection scheme to trip each breaker whenever faults are detected. Operators can also issue commands to the IEDs R1 through R4 to manually trip the breakers BR1 through BR4. These IEDs information back through a substation switch through a router back to the supervisory control and data acquisition systems (SCADA). The dataset contains three attack event scenarios listed below:

1) **Data injection**: This attack aims to blind the operator and causes a black out by changing values to parameters such as current, voltage, and frequency, etc.;
2) **Remote tripping command injection**: This attack sends a command to a relay which causes a breaker to open. It can only be done once an attacker has penetrated outside defenses;
3) **Relay setting change**: This attack changes the setting to disable the relay function such that relay will not trip for a valid fault or a valid command.

A dataset of 37 scenarios has been generated on this system and made available online. These data belong to three classes labeled as No Events, Natural Events, and Attack Events. TableI describes the different scenarios in each class. Each sample in this dataset contains 128 features: 116 of them are from four phasor measurement units (PMUs); another 12 features are collected from the relay logs of the four PMUs, the control panel logs, and the Snort alert [10].

B. Experiment Setup

The original data contain the raw features of different ranges, such as voltage magnitude, current angle and frequency, among others, so data normalization is essential to be first performed to improve the uniformity of the data by adjusting all raw data values into [0, 1]. Considering that the original dataset is imbalanced, where the number of samples in Class 0, 1, and 2 are 4,405, 18,309, and 55,663, respectively. In this case, a classifier that labels a sample as Class 2 (Attack Event) can achieve around 71% accuracy, which is unacceptable. To address this imbalanced problem, after determining 80% of each class for the training set and the remaining 20% for testing, we oversample the training samples as the second procedure of data pre-processing. The 3,524 No Event samples and the 14,647 Natural Event samples are oversampled to the same number (44,530) of the Attack Event samples in the training set, while the testing set is not oversampled. After data pre-processing, SDAE feature extractor is trained with the corrupted inputs that are generated by adding Gaussian noise following $N(0, 0.01)$ into the oversampled training samples, while the SDAE feature extractor is tested without Gaussian noise and obtains the extracted features from original testing samples. XGBoost ensemble learning-based classifier is performed based on the SDAE extracted features, which can distinguish the normal operation, fault and attack events.

C. Simulation Results

Table II shows the classification performance comparison (per class and overall) between individual XGBoost, SAE with XGBoost, SDAE with XGBoost, individual Random Forest, SAE with Random Forest and SDAE with Random Forest models, respectively. As a balanced dataset, we use both per-class and overall accuracies to evaluate the performance of each classifier. Also, we decompose the performance into two sets of false positive rate (FPR) and false negative rate (FNR) between events vs. non-events and faults vs. attacks as follows:

- FPR_1: the fraction of normal samples (non-events) misclassified as non-normal (fault or attack);
- FNR_1: the fraction of non-normal samples (fault or attack) misclassified as normal (non-events);
- FPR_2: the fraction of fault samples misclassified as attack;
- FNR_2: the fraction of attack samples misclassified as fault;

All models based feature extraction can outperform individual XGBoost and Random Forest classifiers, whose overall accuracy are 82.24% and 80.03%, respectively. Specifically, SDAE with XGBoost model has the best overall accuracy with 90.48%, and meanwhile SDAE with XGBoost classifier is the best one in Class 0 (No Events) and Class 2 (Attack Events) with the accuracy of 86.95% and 95.75%, respectively. Besides, SDAE with XGBoost classifier can also outperform the other classifiers in the term of FPR_1, FNR_1 and FNR_2, whose values are 0.009, 0.031, and 0.124, respectively. In addition, it can be found that SAE with Random Forest model has the best accuracy in Class 1 (Natural Events) with 77.44% and the lowest FPR_2 with the value of 0.073. Table II shows that the misclassification mainly arises from the values of FPR_2 and FNR_2, which

represents that there is more challenging in the classification between Natural Events and Attack Events.

Fig. 4 shows the XGBoost classification accuracy with and without SDAE feature extraction, respectively. The training accuracy of individual XGBoost classifier is improved from 73.98% to 100% after 1,200 iterations, and meanwhile its testing accuracy increases from 63.24% to 82.24%. Compared with individual XGBoost classifier, the testing accuracy with SDAE feature extraction increases from 41.31% to 90.48%, as the training accuracy improves from 57.97% to 100% after 1,340 iterations. Besides, the above SDAE is built with two DAE layers whose reconstruction errors are shown in Fig. 5, respectively. The first DAE layer is built to extract 90 features from the original 128 features and its reconstruction error is reduced from 0.18806 to 0.00226 within 6,000 iterations, after the difference of reconstruction error between any two consecutive iterations is lower than 10^{-5} in 100 iterations. After that, the second DAE layer is generated to extract 60 features from the 90 features that is produced by the first DAE layer, and the reconstruction error decreases from -0.78157 to -3.00877 on a log-scale.

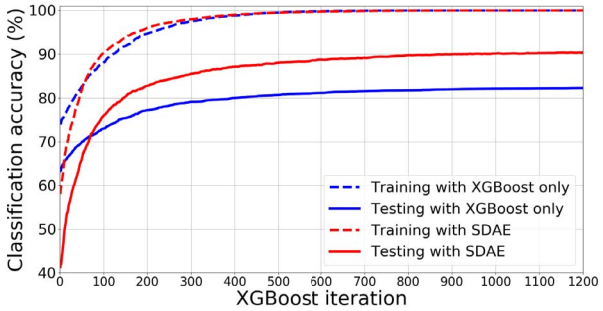

Fig. 4. Classification accuracy without and with SDAE feature extraction.

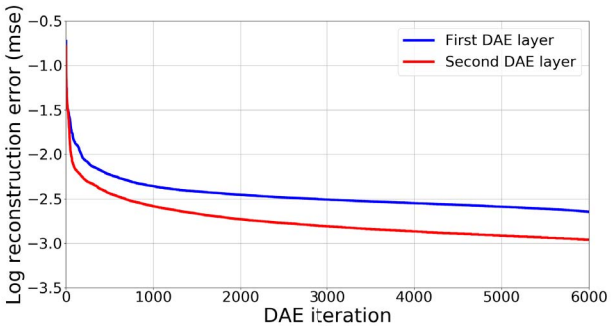

Fig. 5. Reconstruction errors of the stacked DAE layers in SDAE.

IV. Discussions

Both these existing works [11]–[13] achieved over 95% accuracy with the partial data, which may not reflect the accurate overall performance on more comprehensive scenarios, as the accuracy of individual XGBoost classifier drops from 99.01% to 82.24% with the increase of scenario instances [23]. Table III shows the accuracy comparison of different SDAE architectures which consider the numbers of extracted features from 60 to 30, and the numbers of DAE layers from single layer to three layers. The design of two DAE layers with 60 extracted features is recommended due to the best accuracy of 90.48%. To reduce the model complexity as well, we can adopt the simple SDAE architecture with single layer that consists of 40 or 30 neurons, due to the better accuracy of 89.58%.

A further look at the accuracy comparison on different SDAE architectures suggests that the choice and design of feature extractors may have a strong impact on the classification performance. Future studies can be developed on introducing more advanced feature extraction methods and finding best architectures of the proposed SDAE model, so that new noise/attack-informed features can be extracted from cyber-physical system data and the classification performance will further be improved in the real smart grid. In addition, temporal information may also be considered to improve the feature extraction and attack classification performance using recurrent neural networks, such as Long Short-Term Memory [27].

V. Conclusions

This paper proposes a robust end-to-end framework based on SDAE feature extraction and XGBoost ensemble classification. SDAE was applied for feature extraction, which could reconstruct the noise-free data from noise-corrupted perturbations. The learned low-dimensional feature space can present information on heterogeneous and voluminous data. The ensemble learning-based XGBoost classifier was trained with the SDAE-extracted features, which combines multiple CART base learners to distinguish normal, fault, and attack events in the cyber-physical smart grid. Normalization and oversampling were also applied to improve the uniformity and balance of the original data. Our proposed SDAE with XGBoost classifier reported 90.48% testing accuracy which is an effective improvement over the 82.24% accuracy achieved by individual XGBoost classifier with the original features.

Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grants RGPIN-2018-06724 and RGPIN-2016-06691, and by the Fonds de Recherche du Québec - Nature et Technologies (FRQNT) under grant 2019-NC-254971.

References

[1] S. Sridhar, A. Hahn, M. Govindarasu *et al.*, "Cyber-physical system security for the electric power grid." *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[2] H. He and J. Yan, "Cyber-physical attacks and defences in the smart grid: a survey," *IET Cyber-Physical Systems: Theory & Applications*, vol. 1, no. 1, pp. 13–27, 2016.

[3] G. Liang, J. Zhao, F. Luo, S. Weller, and Z. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, July 2017.

[4] C.-C. Sun, A. Hahn, and C.-C. Liu, "Cyber security of a power grid: State-of-the-art," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 45–56, 2018.

[5] B. Zhang, Y. Yu, and J. Li, "Network intrusion detection based on stacked sparse autoencoder and binary tree ensemble method," in *2018 IEEE International Conference on Communications Workshops*, 2018, pp. 1–6.

TABLE II
TESTING PERFORMANCE COMPARISON (PER CLASS AND OVERALL)

Model type	Predict / Actual	Class 0	Class 1	Class 2	Per-Class Accuracy (%)	FPR_1	FNR_1	FPR_2	FNR_2	Overall Accuracy (%)
XGBoost	Class 0	612	50	219	69.47	0.022	0.144	0.129	0.266	82.24
	Class 1	31	2,208	1,423	60.29					
	Class 2	96	966	10,071	90.46					
SAE + XGBoost	Class 0	750	25	106	85.13	0.010	0.037	0.088	0.151	89.16
	Class 1	9	2,673	980	72.99					
	Class 2	24	553	10,556	94.82					
SDAE + XGBoost	Class 0	766	21	94	**86.95**	**0.009**	**0.031**	0.081	**0.124**	**90.48**
	Class 1	7	2,756	899	75.26					
	Class 2	20	453	10,660	**95.75**					
Random Forest	Class 0	556	55	270	63.11	0.027	0.226	0.141	0.291	80.03
	Class 1	43	2,066	1,553	56.42					
	Class 2	156	1,054	9,923	89.13					
SAE + Random Forest	Class 0	725	25	131	82.29	0.012	0.057	**0.073**	0.153	89.95
	Class 1	13	2,836	813	**77.44**					
	Class 2	37	557	10,539	94.66					
SDAE + Random Forest	Class 0	738	17	126	83.77	0.011	**0.031**	0.076	0.138	90.30
	Class 1	4	2,810	848	76.73					
	Class 2	23	503	10,607	95.28					

TABLE III
COMPARISON OF ACCURACY FOR DIFFERENT SDAE ARCHITECTURES

Numbers of DAE layers \ Numbers of extracted features	60	50	40	30
1	89.48%	89.34%	**89.58%**	**89.58%**
2	**90.48%**	87.69%	87.78%	87.44%
3	**86.21%**	86.03%	86.08%	86.09%
Best accuracy	**90.48%**	89.34%	89.58%	89.58%

[6] G. Ericsson, "Cyber security and power system communicationessential parts of a smart grid infrastructure," *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1501–1507, July 2010.

[7] A. Stefanov and C.-C. Liu, "Cyber-power system security in a smart grid environment," in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE, 2012, pp. 1–3.

[8] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 76:1–76:36, Jul. 2018.

[9] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3104–3113, 2015.

[10] U. Adhikari, T. Morris, and S. Pan, "Wams cyber-physical test bed for power system, cybersecurity study, and data mining," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2744–2753, Nov 2017.

[11] R. Hink, J. Beaver, M. Buckner *et al.*, "Machine learning for power system disturbance and cyber-attack discrimination," in *2014 7th International symposium on resilient control systems (ISRCS)*, 2014, pp. 1–8.

[12] D. Sadhasivan and K. Balasubramanian, "A novel LWCSO-PKM-based feature optimization and classification of attack types in SCADA network," *Arabian Journal for Science and Engineering*, vol. 42, no. 8, pp. 3435–3449, 2017.

[13] D. Wilson, Y. Tang, J. Yan, and Z. Lu, "Deep learning-aided cyber-attack detection in power transmission systems," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2018, pp. 1–5.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[16] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[17] M. Nicolau, J. McDermott *et al.*, "A hybrid autoencoder and density estimation model for anomaly detection," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2016, pp. 717–726.

[18] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. ACM, 2014, p. 4.

[19] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[20] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.

[21] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 89–100, Feb 2018.

[22] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*. Elsevier, 1992, pp. 65–93.

[23] C. Hu, J. Yan, and C. Wang, "Advanced cyber-physical attack classification with extreme gradient boosting for smart transmission grids," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, accepted.

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

[25] H. Zhang, S. Si, and C.-J. Hsieh, "GPU-acceleration for large-scale tree boosting," *arXiv preprint arXiv:1706.08359*, 2017.

[26] G. Ke, Q. Meng, T. Finley *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[27] Y. Lin, J. Wang, and M. Cui, "Reconstruction of power system measurements based on enhanced denoising autoencoder," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, accepted.