Public review for

Data Analytics for Cybersecurity Enhancement of Transformer Protection

Martiya Zare Jahromi, Amir Abiri Jahromi, Deepa Kundur, Scott Sanner, Marthe Kassouf

This paper presents machine learning approaches to detect cyberattacks on electric grid transformers. Specifically, the authors compare the effectiveness of multiple anomaly detection techniques to detect malicious tampering to transformers' differential protective relays. The anomaly detection techniques use different types of autoencoder neural networks, where each focuses on a particular type of transformer fault. The authors train and validate their models using the D6 benchmark test system. The results show that their LSTM model outperforms the others on one-phase-to-ground and two-phase-to-ground faults. However, for three-phase-to-ground fault, the other models tend to detect anomalies better. The authors also evaluate their models' performance on unseen anomalies and find that a Linear AutoEnconder model performs best. The reviewers thought the authors addressed an important topic in security for smart grid systems, and rigorously validated their results using a standard benchmark test system.

Public review written by
David Irwin
University of Massachusetts Amherst, USA

Data Analytics for Cybersecurity Enhancement of Transformer Protection

MARTIYA ZARE JAHROMI, University of Toronto, Canada
AMIR ABIRI JAHROMI, University of Toronto, Canada
DEEPA KUNDUR, University of Toronto, Canada
SCOTT SANNER, University of Toronto, Canada
MARTHE KASSOUF, Hydro-Quebec Research Institute (IREQ), Canada

Electric power substations are experiencing an accelerated pace of digital transformation including the deployment of LAN-based IEC 61850 communication protocols that facilitate accessibility to substation data while also increasing remote access points and exposure to complex cyberattacks. In this environment, machine learning algorithms will play a vital role in cyberattack detection and mitigation and natural questions arise as to the most effective models in the context of smart grid substations. This paper compares the performance of three autoencoder-based anomaly detection systems including linear, fully connected, and convolutional autoencoders, as well as long short-term memory (LSTM) neural network for cybersecurity enhancement of transformer protection. The simulation results indicated that the LSTM model outperforms the other models for detecting cyberattacks targeting asymmetrical fault data. The linear autoencoder, fully connected autoencoder and 1D CNN further outperform the LSTM model for detecting cyberattacks targeting the symmetrical fault data.

CCS Concepts: • **Computer systems organization → Embedded and cyber-physical systems**; • **Computing methodologies → Anomaly detection**.

Additional Key Words and Phrases: cybersecurity, data analytics, machine learning, transformer protective relays

1 INTRODUCTION

The rapid integration of standard and interoperable information and communication technologies (ICT) in substations [12, 33] has accelerated the frequency and complexity of electric utility cyberattacks [13]. Attacks against electric power substations such as that on the Ukrainian grid in 2015 have caused significant societal and economic damage including loss of life [10, 28, 30]. As such, the North American Electric Reliability Corporation (NERC) has taken initial steps towards safeguarding cyber-assets by mandating the critical infrastructure protection (CIP) standards [1].

The emergence of standardized and interoperable communication protocols such as IEC 61850 and industrial internet of things (IIOT)-based applications renders traditional security-by-obscurity and perimeter defense security strategies obsolete [2]. Yet, these transformations facilitate accessibility to high fidelity substation data to lay a powerful groundwork for developing machine learning-based data analytics for cybersecurity enhancement [20]. Cybersecurity of substations has been analyzed in the literature from two perspectives; 1) cybersecurity risk assessment/impact analysis [7, 18, 26] and 2) cyberattack detection, mitigation and prevention [5, 15–17, 29]. Most of the approaches solely focus on information technology (IT) data. For instance, some of these approaches attempt to detect cyberattacks by examining the intruders' footprints on the communication packets. This is while the cyberattack signatures on the operational technology (OT) data have been commonly neglected. This trend is expected to rapidly change in the coming years by the introduction of novel cyberattack detection systems that rely on both information technology (IT) and operational technology (OT).

Anomaly-based techniques have three main advantages over misuse-based techniques. First, anomaly-based techniques can adaptively learn the time varying dynamics and operating points of power systems to establish comprehensive baselines for system behaviors. Second, anomaly-based detection techniques only require training on normal (non-attack) data, which is available in abundance compared to cyberattack data making it possible to easily obtain the necessary training sets for model optimization. Third, anomaly-based techniques are capable of detecting unencountered zero-day cyberattacks. The primary disadvantage of anomaly-based techniques is the potential for high false detection rates because previously unseen system behaviors can be categorized as anomalies [4, 11, 34].

Machine learning-based anomaly detection systems have been extensively examined for cyberattack detection in smart grids. An artificial intelligence-based approach has been proposed in [23] to identify compromised meters. An intrusion detection system has been proposed in [3] for wide area measurements. An unsupervised anomaly detection system has been proposed in [21] to differentiate cyberattacks from disturbances and faults in smart grids. A machine learning-based method has been proposed in [6] to detect cyberattacks against state estimation. In [32], the margin setting algorithm has been employed to defend smart grids against false data injection attacks.

Despite the considerable potential of machine learning-based anomaly detection systems, they have received less attention in the literature compared to analytical approaches for cybersecurity enhancement of substations due to the lack of high fidelity data in traditional substations. A 1-dimensional convolutional based autoencoder has been employed in [24] to identify cyberattacks against distance protective relays. A fully connected autoencoder has been employed in [19] to enhance the cybersecurity of the transformer differential protection. In [4], data analytics comprising long short-term memory neural network and ridge based regression classifier have been used to identify the root causes of the transmission protection mal-operation. Yet, different autoencoder-based anomaly detection systems for cybersecurity enhancement of protective relays have not been compared previously in the literature.

Authors' addresses: Martiya Zare Jahromi, University of Toronto, Canada; Amir Abiri Jahromi, University of Toronto, Canada; Deepa Kundur, University of Toronto, Canada; Scott Sanner, University of Toronto, Canada; Marthe Kassouf, Hydro-Quebec Research Institute (IREQ), Canada.

This paper compares the performances of different autoencoder-based anomaly detection systems as well as LSTM for cybersecurity enhancement of transformer protection. Specifically, we employ a variety of autoencoder-based anomaly detection systems as well as the LSTM neural network for cybersecurity enhancement of transformer protection using OT data. The performance of different autoencoder-based anomaly detection systems and the LSTM neural network for identifying different types of cyberattacks are measured and compared.

2 THE FALSE DATA INJECTION ATTACK AGAINST TRANSFORMER PROTECTION

Cyberattackers may target confidentiality, integrity, or availability (C-I-A) of data. Confidentiality aims to prevent users/devices from accessing unauthorized data. Integrity is about validity and correctness of data. Availability deals with the accessibility of data within a reasonable amount of time to an authorized user/device. Availability and integrity of data are paramount for OT systems like protective relays because they rely on real time data to identify abnormal conditions such as faults to actuate circuit breakers. Prominent examples of cyberattacks on the availability and integrity of data include distributed denial of service (DDoS) and false data injection (FDI) attacks, respectively.

The main protection of transformers are typically differential protective relays. Differential relaying monitors the currents entering and leaving a transformer calculating the geometrical sum (called the differential current) of the current phasors at all the terminals of the transformer. If the differential current takes a value of zero or a very small value due to measuring inaccuracies, the differential relay will stay inactive because the system is considered to either be operating normally or have an external fault. If the differential current takes a large value, it indicates an internal fault, within the scope of the relay, and the differential relay will trip the circuit breakers of the transformer [8, 22].

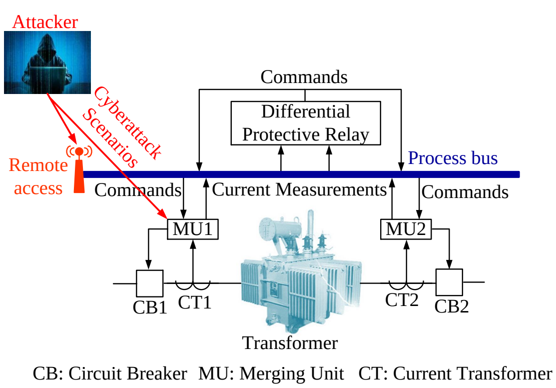

Fig. 1. Transformer differential protective relay.

In this paper, we consider IEC 61850-based substation automation as illustrated in Fig. 1. Here, the merging units (MU) collect analog measurements from the current transformers CT1 and CT2 and send them to the differential protective relay using SV packets through the process bus after performing the analog-to-digital conversion. Moreover, the merging units receive the commands from the differential protective relay in the form of GOOSE packets through the process bus and send the trip signals to the circuit breakers after performing the digital-to-analog conversion.

An FDI attack is considered here where the intruder manipulates the magnitude and phase angle of the current measurements, yielding different elements of the differential protective relay to issue false tripping commands to the transformer circuit breakers. Multiple scenarios can be considered for the execution of the FDI attack against transformer protection including: 1) the installation of malicious firmware on the merging units through a supply chain attack or by physical access to the merging units, and 2) the injection of false data to the process bus through remote rogue connections using stolen legitimate substation operator credentials.

3 MACHINE LEARNING-BASED ANOMALY DETECTION SYSTEMS

The main objective of the proposed anomaly detection systems is to detect the malicious tampering of current measurements by an attacker to illegitimately trigger different elements of the differential protective relay of a transformer. The differential protective relay of a transformer is designed to detect different types of faults including three-phase-to-ground, two-phase-to-ground, and single-phase-to-ground. Therefore, we design and train a separate anomaly detection system for each type of fault. Each of the anomaly detection systems becomes operational by the activation of the corresponding element of the differential protective relay. We further investigate the possibility of considering a universal architecture for anomaly detection systems for different types of faults by comparing the architectures obtained for each type of fault.

The choice of machine learning models for anomaly detection depends on the nature and dimensionality of the input data. The input data to the anomaly detection system for differential protective relays is composed of two time series of three-phase current measurements. This results in high dimensionality of the input data and complicates feature extraction for machine learning. Moreover, the evolving and clandestine nature of cyberattacks limit the possibility of effective modeling of anomalous behaviour of cyberattacks in contrast to normal behaviour in substations for which there is significantly more data and more predictable characteristics. In this environment, semi-supervised and unsupervised machine learning approaches are in a superior position for cyberattack detection in contrast to supervised machine learning approaches. Advances in semisupervised and unsupervised machine learning has made it possible to solve classification problems, including anomaly detection, with high-dimensional data sets that can suffer from complex structure, sparsity or overfitting [25]. The autoencoder-based anomaly detection systems and LSTM neural network learn to compress the input data into a smaller latent space, then reconstruct the input data from the latent space with a low reconstruction error. Since we train the autoencoders and LSTM neural network with benign or attack-free current measurement sequences, we expect to observe high reconstruction error when feeding malicious current measurement sequences as input [31]. We define the reconstruction error

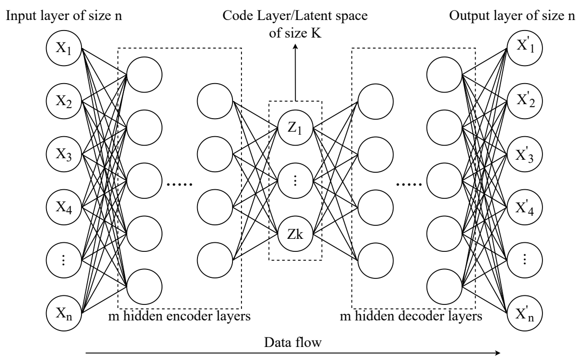

Fig. 2. Autoencoder structure

for a data sequence X_i as given in (1). A data sequence is considered anomalous if the reconstruction error is above a predefined threshold as given in (2).

$$MSE_i = ||X_i - M(X_i)||_2^2 \tag{1}$$

$$MSE_i > \epsilon \rightarrow \quad anomalous \quad data \quad sequence \tag{2}$$

where MSE_i is the Mean Squared Error, X_i denotes the input data sequence, M denotes the autoencoder or LSTM model, $M(X_i)$ denotes the output data sequence, and ϵ denotes the threshold considered on the reconstruction error for anomaly detection.

Linear Autoencoders. A linear autoencoder consists of an input layer, a code layer with a size smaller than input/output layers, and an output layer. In a linear autoencoder, all the activation functions in each layer are linear. The linear autoencoder model is similar to dimensionality reduction in Principal Component Analysis (PCA).

Fully Connected Autoencoders. In fully connected neural networks, all the neurons in each layer are connected to all the neurons in the subsequent layer. From a technical perspective, a fully connected autoencoder consists of two parts; an encoder and a decoder, as illustrated in Fig. 2. An encoder consists of an input layer, a variable number of hidden layers, and a code (embedding) layer. The code layer connects the encoder and decoder and its size is smaller than input and output layers. The decoder consists of the same number of hidden layers as the encoder and an output layer.

1D Convolutional Neural Networks. Unlike fully connected networks, neurons of each layer in CNN are not connected to all the neurons in the following layer and parameter sharing exists that reduces storage. 1D-CNN is a good candidate for anomaly detection because it is capable of detecting localized anomalies due to its window-based nature. Different types of layers are used in CNN autoencoders including convolutional, pooling, and upsampling [25, 27]. The convolution layer works based on convolution operation as given in (3).

$$S(i) = (X * f)(i) = \sum_{j=0}^{n} X(j) f(i-j) \tag{3}$$

where X denotes the input of the operation, S denotes the output, f denotes the convolution filter and n denotes the length of the convolution filter.

In convolution layers, various filters are applied in parallel to the input to produce a set of linear activations. Each linear activation is followed by a non-linear activation function. To reconstruct the original input in the decoder, upsampling and convolution layers are combined. This combination is also known as transposed convolution or deconvolution. Fig 3 shows an example of an upsampling operation.

Long Short-Term Memory Networks. Other types of networks, called recurrent networks, consist of neurons that have self-connections or connections to neurons from previous layers. This recurrency provides the ability for the network to retain what happened in the past (short-term memory). The new state h_t is expressed as:

$$h_t = f_w(h_{t-1}, x_t) \tag{4}$$

where x_t is the input vector at time step t, h_{t-1} denotes the old state, and f_w is a function with parameters w.

Consider Fig. 4 as a simple example of an RNN. Using the recurrent formula in (4) at each time step, we can process a sequence of vectors $X = \{x_1, x_2, ..., x_n\}$ using the same function f and weights w at every time step.

To address the vanishing gradient problem in RNNs, Long Short Term Memory (LSTM) networks have been designed. An LSTM network is a good candidate for anomaly detection because it is capable of detecting non-local, long term anomalies. Fig. 5 shows an LSTM unit. The horizontal line on top of the unit is responsible for passing the cell state which facilitates the long-term memory for relevant components of the data. LSTM unit consists of three gates. Forget gate is responsible for removing the parts of the cell state that are no longer needed. Input gate adds the information needed to the cell state. Output gate produces the output. It is possible to stack up arrays of LSTM units to enable more complex LSTM networks.

4 TRAINING, VALIDATION AND OPTIMIZATION OF THE ARCHITECTURE OF THE ANOMALY DETECTION SYSTEMS

In this section, we provide information about the test system and the training data set. We explain the approach employed for optimizing the architectures of the proposed machine learning-based anomaly detection systems. In addition, a set of metrics are presented for measuring the performance of the anomaly detection systems.

4.1 Test System

The IEEE power system relaying committee (PSRC) D6 benchmark test system is considered for generating the training data sets [14]. This test system connects a power plant with four 250 MVA generator units to a 230 kV transmission network through two parallel 500 kV transmission lines. The 230 kV transmission network is modeled as an infinite bus. Differential protective relays protect the power plant transformers as illustrated in Fig. 6.

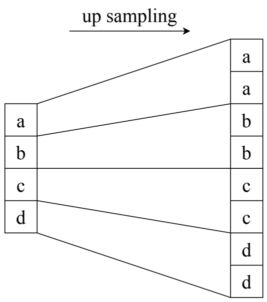

Fig. 3. Upsampling with a factor of 2.

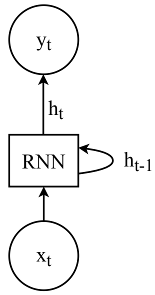

Fig. 4. Recurrent Neural Network

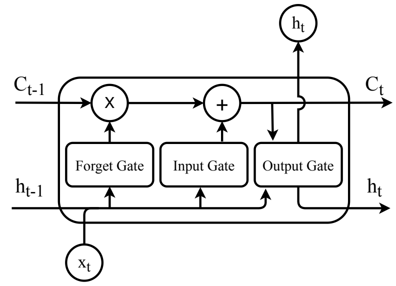

Fig. 5. LSTM Unit

4.2 Training Data Set

OPAL-RT HYPERSIM is employed to implement and simulate the PSRC D6 test system and generate the training data sets. The simulations are performed for a duration of 1.5 seconds with the fault start varying randomly between t=1 s to t=1.02 s to ensure the fault occurs at different parts of the current waveforms. Note that the period of one cycle is approximately 0.0167 s in a 60 Hz power system. Moreover, the generation levels are changed between 350 MW and 360 MW in 2 MW step size in each simulation to generate data sets under different operating conditions. The simulations are performed for different types of faults including three-phase-to-ground, two-phase-to-ground and single-phase-to-ground faults. The fault impedance is assumed to be zero. In total, 20,736 simulations are performed to generate training data sets for each type of fault. The anomaly detection systems are trained with 80% of the 20,736 simulations for each type of fault. The validation and test data sets each comprises 10% of the 20,736 simulations.

The three-phase current measurements are collected from CT1 and CT2 at the sampling rate of 4800 samples per second in compliance with IEC 61850-9-2 standard for SV packet specifications [9]. An important parameter for training of the machine learning-based anomaly detection system is the input data length. A sliding window of 10 ms, *i.e.*, 48 samples of current measurements per phase, is fed to the anomaly detection systems as input. As such, the input data to the anomaly detection systems contain $6 \times 48 = 288$ samples in total. In order to obtain the input data, we extracted a 20 ms window from each 1.5 s simulation containing 47 samples before the starting point of the fault and 47 samples after the starting point of the fault. Next, we slide the 10 ms input window of the anomaly detection systems over the 20 ms window of data sample by sample. This amounts to 48 windows of input data per simulation with 10 ms duration.

Table 1. List of hyperparameters for each model.

Model	Parameter	Set of values
Linear AE	Learning rate Code Size	{0.01, 0.001 } {30, 40, 50}
Fully Connected AE	Learning rate Depth Code Size	{0.01, 0.001 } {1, 2, 3, 4, 5 } {30, 40, 50}
1D CNN	Learning rate Depth Convolution filter size Filter count	{0.01, 0.001 } {1, 2, 3, 4, 5 } {1, 2, 3, ..., 9} {4, 8, 12, ..., 44 }
LSTM	Learning rate Depth LSTM Units count	{0.01, 0.001 } {1, 2, 3, 4, 5} {5, 10, 15, ..., 50}

4.3 Optimizing the Architecture of the Machine Learning-Based Anomaly Detection Systems

We used the grid search method for hyperparameter tuning and optimizing the architecture. In this method, we consider values in Table 1 for each hyperparameter. Different possible combinations of hyperparameters are then tested using grid search. The best hyperparameter values are selected based on the validation error observed in the grid search.

The test data set includes 2074 simulation data sets. We replaced 207 of the test data sets, *i.e.*, 10% of the test data sets, with FDI cyberattack data sets in order to create an imbalanced test data set. We use an imbalanced test data set because cyberattacks in power systems are rare events compared to normal behavior. The hyperparameters considered and tested for each model are listed in Table 1.

The FDI cyberattack data sets considered cover various situations ranging from naive scenarios where the cyberattacker only understands the principles of transformer differential protective relays to very sophisticated cyberattacks where the cyberattacker has some knowledge of power system dynamics and transformer fault signatures.

The cyberattack data are generated by OPAL-RT HYPERSIM. We considered three different scenarios for cyberattack data generation. In the first scenario, random false data are generated by OPAL-RT HYPERSIM with the appropriate magnitude to mimic a fault condition. In the second scenario, the tap setting of the current transformer in the test system are modified in OPAL-RT HYPERSIM

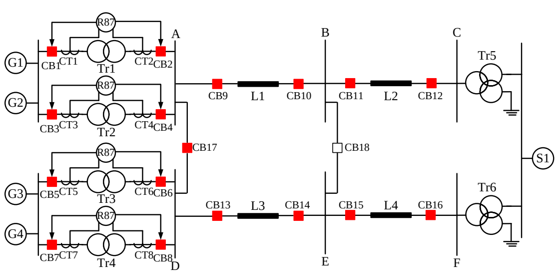

Fig. 6. The IEEE PSRC D6 benchmark test system.

Table 2. Selected Parameters for The Proposed Models. (A: One-Phase-To-Ground Faults, B: Two-Phase-To-Ground Faults, C: Three-Phase-To-Ground Faults)

Model	Parameter	A	B	C
Linear AE	Learning rate	0.001	0.001	0.001
	Code Size	40	50	50
Fully Connected AE	Learning rate	0.001	0.001	0.001
	Depth	2	3	1
	Code Size	40	40	40
1D CNN	Learning rate	0.001	0.001	0.001
	Depth	2	1	2
	Convolution filter size	4	6	4
	Filter count	48	28	48
LSTM	Learning rate	0.001	0.001	0.001
	Depth	1	1	1
	LSTM units count	40	40	20

Table 3. 1D Convolutional Network Architecture for Anomaly Detection in One-Phase-To-Ground Fault Measurements

index	layer type	Output Dimensions		parameter
		Length	Count	
1	Input	288	1	-
2	Zero Padding	292	1	pad size = 3
3	Convolution	292	32	filter size = 6
4	max pooling	146	32	window size = 2
5	Convolution	146	64	filter size = 6
6	max pooling	73	64	window size = 2
7	Convolution	73	128	filter size = 6
8	up sampling	146	128	window size = 2
9	Convolution	146	64	filter size = 6
10	up sampling	292	64	window size = 2
11	Convolution	292	32	filter size = 6
12	up sampling	292	1	window size = 2
13	cropping	288	1	size = 288

such that transformer differential protective relay receives current measurements with larger magnitude, mimicking a fault condition. In the third scenario, a fault condition is simulated in OPAL-RT HYPERSIM and used as a replay attack.

4.3.1 *Linear Autoencoder Architecture.*

We employed a fully connected autoencoder with one hidden layer, one input layer, and an output layer. All the activation functions in the model are linear.

4.3.2 *Fully Connected Autoencoder Architecture.*

We used a fully connected autoencoder with the same number of hidden layers in encoder and decoder. In the encoder part, the number of neurons in hidden layers monotonically decreases from the input to the code layer. We used the Adam optimizer for model optimization. Fig. 7 shows the fully connected autoencoder architecture obtained for anomaly detection considering the current measurements triggering the three-phase-to-ground fault element of the differential protective relay. For the sake of brevity, the autoencoder architectures obtained for different types of faults are summarized in Table 2.

4.3.3 *1D Convolutional Neural Network Architecture.*

In the 1D CNN architecture, the first layer is zero padding. In the encoder part, we use convolution and max-pooling layers. In the decoder part, there are deconvolution layers, a combination of upsampling and convolution layers. Max pooling and upsampling layers both have window sizes of 2. Filter size for convolution layers is tuned in the hyperparameter tuning step. Table 3 summarizes the details of the architecture obtained for anomaly detection. For conciseness, the hyperparameters selected for 1D CNN are summarized in Table 2.

4.3.4 *LSTM Architecture.*

The many to many one direction LSTM network is considered with input is a sequence of 48 vectors of size 6. We considered stacks of LSTM unit arrays followed by a dense layer. The learning rate, number of LSTM layers, and the size of LSTM unit arrays are tuned in the hyperparameter tuning step and summarized in Table 2. Fig. 8 shows the LSTM architecture obtained for anomaly detection considering the current measurements triggering the two-phase-to-ground fault element of the differential protective relay. X_t represents the current sample at the time step t. Y_t represents the output of time step t. h represents the hidden state, and c represents the cell state.

The similarity of architectures obtained for anomaly detection systems for different types of faults in Table 2 indicates that a universal architecture can be possibly designed for different types of faults.

5 SIMULATION RESULTS

The performance of the anomaly detection systems are measured using precision and recall metrics, which are more appropriate for imbalanced datasets. It is worth noting that the accuracy metric is not helpful because cyberattacks are rare events. The correct selection of the threshold value plays a vital role on the performance.

5.1 Performance Analysis of Autoencoder-Based Anomaly Detection Systems

The performance of the linear autoencoder, fully connected autoencoder, 1D convolutional autoencoder, and LSTM are measured for detecting cyberattacks against different elements of the transformer differential protective relay. We use the precision-recall curve to understand the performances of the four models for different possible thresholds.

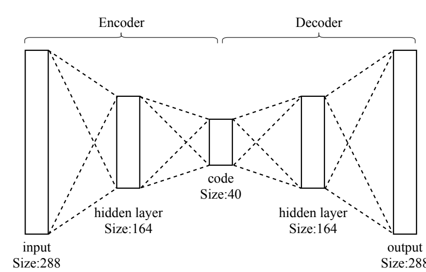

Fig. 7. Autoencoder structure for three-phase-to-ground fault.

Table 4. Performance of the Anomaly Detection Systems. (A: One-Phase-To-Ground Faults, B: Two-Phase-To-Ground Faults, C: Three-Phase-To-Ground Faults)

	Linear AE		Fully-connected AE		1D CNN		LSTM	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
A	1	0.992	1	0.995	1	0.995	1	1
B	1	0.990	1	0.981	1	0.975	1	0.993
C	1	0.991	1	0.989	1	0.986	1	0.969

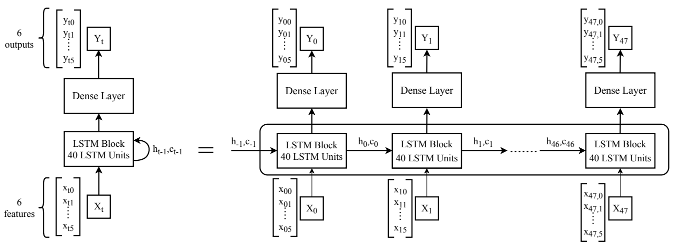

Fig. 8. LSTM Architecture for Anomaly Detection in the Two-Phase-To-Ground Fault Measurements. We feed 48 samples sequentially.

Table 5. Performance of The Anomaly Detection Systems While Considering Unseen Data with Finer Granularity.

Linear AE		Fully-connected AE		1D CNN		LSTM	
Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
1	0.84	1	0.8	1	0.61	1	0.58

The LSTM model outperforms the other three models for the one-phase-to-ground and two-phase-to-ground faults as illustrated in figures 9 and 10. However, the linear autoencoder, fully connected autoencoder and 1D CNN approximately have similar curves and outperform the LSTM model for the three-phase-to-ground fault as illustrated in Fig. 11. Table 4 summarizes the results when the threshold is selected such that the precision is equal to 1. It is worth noting that even a subtle change in the performance of anomaly detection systems for protective relays is significant because the misoperation due to cyberattacks has the potential to cause major disturbances and widespread blackouts in power systems. Given the symmetry of three-phase-to-ground faults and asymmetry of single-phase-to-ground and two-phase-to-ground faults, we conclude that LSTM performs better for asymmetrical faults and is weaker than the other models for symmetrical faults. We feel that such a trend will generalize to other more complex systems beyond the benchmark system employed in this paper because of the inherent ability of LSTM to recognize time series patterns and manage long-term memory patterns in contrast to the other models.

5.2 Impact of Data Granularity on Anomaly Detection Performance

In this case study, we investigate the impact of generation level granularity on the performance of each of the machine learning algorithms while considering the three-phase-to-ground-fault. Thus, test data sets are generated for the three-phase-to-ground-fault with finer generation level granularity compared to the training data set, *i.e.*; the generation levels are changed with 1 MW step size in each simulation. Next, we measured the performance of the linear autoencoder, fully connected autoencoder, 1D convolutional autoencoder and LSTM for detecting cyberattacks while considering cyberattack data and data of finer generation level granularity that have not been considered in the training step.

Table 5 summarizes the results when the threshold is selected such that the precision is equal to 1. The comparison between the results in Table 4 and Table 5 show that the performance of all four models significantly drops when they are exposed to data captured from other generation levels that are not included in the original dataset. Yet, the linear autoencoder model outperforms the three other models.

6 CONCLUSION

This paper presented four machine learning-based anomaly detection systems including linear autoencoder, fully connected autoencoder, convolutional autoencoder, and LSTM neural network for cybersecurity enhancement of transformer differential protection for anomaly detection in transformer relays. The simulation results underscore that the LSTM model outperforms the other models for one-phase-to-ground and two-phase-to-ground faults. The linear autoencoder, fully connected autoencoder and 1D CNN further outperform the LSTM model for the three-phase-to-ground fault. The impact of input data granularity on the performance of the deep learning-based anomaly detection systems is further investigated using a sensitivity analysis. The results showed that the performance of the all four models significantly drop when they are exposed to the previously unseen system behaviors. Yet, the linear autoencoder model outperformed the three other models when it is exposed to the previously unseen system behaviors.

REFERENCES

[1] [n.d.]. North american electric reliability corporation (nerc) Critical Infrastructure Protection (CIP) Reliability Standards. http://www.nerc.com.
[2] Pascal Ackerman. 2017. *Industrial Cybersecurity: Efficiently secure critical infrastructure systems.* Packt Publishing Ltd.
[3] Uttam Adhikari, Thomas H Morris, and Shengyi Pan. 2016. Applying non-nested generalized exemplars classification for cyber-power event and intrusion detection. *IEEE Transactions on Smart Grid* 9, 5 (2016), 3928–3941.
[4] Arman Ahmed, Vignesh VG Krishnan, Seyedeh Armina Foroutan, Md Touhiduzzaman, Caroline Rublein, Anurag Srivastava, Yinghui Wu, Adam Hahn, and Sindhu Suresh. 2019. Cyber physical security analytics for anomalies in transmission protection systems. *IEEE Transactions on Industry Applications* 55, 6 (2019), 6313–6323.
[5] Amir Ameli, Ali Hooshyar, and Ehab F El-Saadany. 2018. Development of a cyber-resilient line current differential relay. *IEEE Transactions on Industrial Informatics* 15, 1 (2018), 305–318.
[6] Mohammad Ashrafuzzaman, Saikat Das, Yacine Chakhchoukh, Sajjan Shiva, and Frederick T Sheldon. 2020. Detecting stealthy false data injection attacks in the smart grid using ensemble-based machine learning. *Computers & Security* 97 (2020), 101994.
[7] Mahdi Bahrami, Mahmud Fotuhi-Firuzabad, and Hossein Farzin. 2019. Reliability evaluation of power grids considering integrity attacks against substation protective IEDs. *IEEE Transactions on Industrial Informatics* 16, 2 (2019), 1035–1044.
[8] J Lewis Blackburn and Thomas J Domin. 2006. *Protective relaying: principles and applications.* CRC press.
[9] Ch Brunner, G Lang, F Leconte, and F Steinhauser. 2004. Implementation guideline for digital interface to instrument transformers using IEC 61850–9–2. *Tech. Rep.* (2004).
[10] Defense Use Case. 2016. Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)* 388 (2016).
[11] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41, 3 (2009), 1–58.
[12] Göran N Ericsson. 2010. Cyber security and power system communication—essential parts of a smart grid infrastructure. *IEEE Transactions on Power delivery* 25, 3 (2010), 1501–1507.
[13] Colleen Glenn, Dane Sterbentz, and Aaron Wright. 2016. *Cyber threat and vulnerability analysis of the US electric sector.* Technical Report. Idaho National Lab.(INL), Idaho Falls, ID (United States).
[14] H Gras, J Mahseredjian, E Rutovic, U Karaagac, A Haddadi, O Saad, I Kocar, and A El-Akoum. 2017. A new hierarchical approach for modeling protection systems in EMT-type software. In *Intern. Conf. Power Syst. Transients.*
[15] Junho Hong and Chen-Ching Liu. 2017. Intelligent electronic devices with collaborative intrusion detection systems. *IEEE Transactions on Smart Grid* 10, 1 (2017), 271–281.

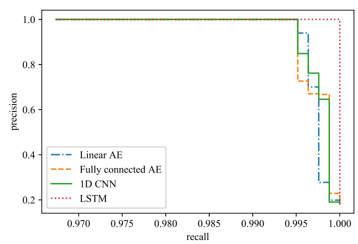

Fig. 9. Precision-recall curve for one-phase-to-ground faults

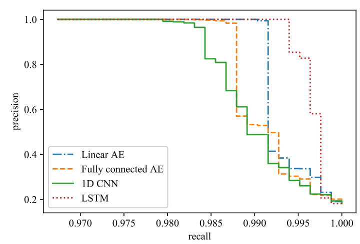

Fig. 10. Precision-recall curve for two-phase-to-ground faults

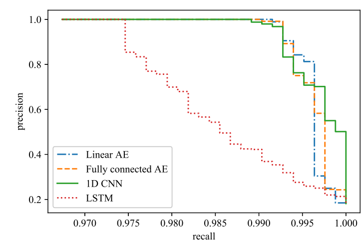

Fig. 11. Precision-recall curve for three-phase-to-ground faults

[16] Junho Hong, Chen-Ching Liu, and Manimaran Govindarasu. 2014. Integrated anomaly detection for cyber security of the substations. *IEEE Transactions on Smart Grid* 5, 4 (2014), 1643–1653.

[17] Junho Hong, Reynaldo F Nuqui, Anil Kondabathini, Dmitry Ishchenko, and Aaron Martin. 2018. Cyber attack resilient distance protection and circuit breaker control for digital substations. *IEEE Transactions on Industrial Informatics* 15, 7 (2018), 4332–4341.

[18] Amir Abiri Jahromi, Anthony Kemmeugne, Deepa Kundur, and Aboutaleb Haddadi. 2019. Cyber-physical attacks targeting communication-assisted protection schemes. *IEEE Transactions on Power Systems* 35, 1 (2019), 440–450.

[19] Martiya Zare Jahromi, Amir Abiri Jahromi, Scott Sanner, Deepa Kundur, and Marthe Kassouf. 2020. Cybersecurity Enhancement of Transformer Differential Protection Using Machine Learning. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 1–5.

[20] Hadis Karimipour, Ali Dehghantanha, Reza M Parizi, Kim-Kwang Raymond Choo, and Henry Leung. 2019. A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. *IEEE Access* 7 (2019), 80778–80788.

[21] Hadis Karimipour, Ali Dehghantanha, Reza M Parizi, Kim-Kwang Raymond Choo, and Henry Leung. 2019. A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. *IEEE Access* 7 (2019), 80778–80788.

[22] Mladen Kezunovic, Jinfeng Ren, and Saeed Lotfifard. 2016. *Design, modeling and evaluation of protective relays for power systems*. Springer.

[23] Kush Khanna, Bijaya Ketan Panigrahi, and Anupam Joshi. 2018. AI-based approach to identify compromised meters in data integrity attacks on smart grid. *IET Generation, Transmission & Distribution* 12, 5 (2018), 1052–1066.

[24] Yew Meng Khaw, Amir Abiri Jahromi, Arani Mohammadreza FM, Deepa Kundur, Scott Sanner, and Marthe Kassouf. 2019. Preventing false tripping cyberattacks against distance relays: A deep learning approach. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 1–6.

[25] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[26] Xindong Liu, Mohammad Shahidehpour, Zuyi Li, Xuan Liu, Yijia Cao, and Zhiyi Li. 2016. Power system risk assessment in cyber attacks considering the role of protection systems. *IEEE Transactions on Smart Grid* 8, 2 (2016), 572–580.

[27] AL-Hawawreh Muna, Nour Moustafa, and Elena Sitnikova. 2018. Identification of malicious activities in industrial internet of things based on deep learning models. *Journal of Information security and applications* 41 (2018), 1–11.

[28] National Academies of Sciences, Engineering, and Medicine. 2017. *Enhancing the resilience of the nation's electricity system*. National Academies Press.

[29] Upeka Kanchana Premaratne, Jagath Samarabandu, Tarlochan S Sidhu, Robert Beresh, and Jian-Cheng Tan. 2010. An intrusion detection system for IEC61850 automated substations. *IEEE Transactions on Power Delivery* 25, 4 (2010), 2376–2383.

[30] Joe Slowik. 2018. Anatomy of an attack: Detecting and defeating crashoverride. *VB2018, October* (2018).

[31] Dušan Sovilj, Paul Budnarain, Scott Sanner, Geoff Salmon, and Mohan Rao. 2020. A comparative evaluation of unsupervised deep architectures for intrusion detection in sequential data streams. *Expert Systems with Applications* 159 (2020), 113577.

[32] Yi Wang, Mahmoud M Amin, Jian Fu, and Heba B Moussa. 2017. A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids. *IEEE Access* 5 (2017), 26022–26033.

[33] Solveig Ward, Jim O'Brien, Bob Beresh, Gabriel Benmouyal, Dennis Holstein, John T Tengdin, Ken Fodero, Mark Simon, Matt Carden, Murty VVS Yalla, et al. 2007. Cyber security issues for protective relays; c1 working group members of power system relaying committee. In *2007 IEEE Power Engineering Society General Meeting*. IEEE, 1–8.

[34] Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. 2018. Machine learning and deep learning methods for cybersecurity. *Ieee access* 6 (2018), 35365–35381.