Review of Anomaly Detection Systems in Industrial Control Systems Using Deep Feature Learning Approach

Raogo Kabore, Adlès Kouassi, Rodrigue N'goran, Olivier Asseu, Yvon Kermarrec, Philippe Lenca

Scientific Research Publishing

Engineering, 2021, 13, 30-44
https://www.scirp.org/journal/eng
ISSN Online: 1947-394X
ISSN Print: 1947-3931

Review of Anomaly Detection Systems in Industrial Control Systems Using Deep Feature Learning Approach

Raogo Kabore[1], Adlès Kouassi[1], Rodrigue N'goran[1], Olivier Asseu[1], Yvon Kermarrec[2], Philippe Lenca[2]

[1]Lastic, ESATIC, Abidjan, Côte d'Ivoire
[2]Lab-STICC, IMT-Atlantique, Brest, France
Email: raogo.kabore@esatic.edu.ci, oasseu@yahoo.fr, yvon.kermarrec@imt-atlantique.fr, philippe.lenca@imt-atlantique.fr

How to cite this paper: Kabore, R., Kouassi, A., N'goran, R., Asseu, O., Kermarrec, Y. and Lenca, P. (2021) Review of Anomaly Detection Systems in Industrial Control Systems Using Deep Feature Learning Approach. *Engineering*, **13**, 30-44.
https://doi.org/10.4236/eng.2021.131003

Received: December 15, 2020
Accepted: January 10, 2021
Published: January 13, 2021

Abstract

Industrial Control Systems (ICS) or SCADA networks are increasingly targeted by cyber-attacks as their architectures shifted from proprietary hardware, software and protocols to standard and open sources ones. Furthermore, these systems which used to be isolated are now interconnected to corporate networks and to the Internet. Among the countermeasures to mitigate the threats, anomaly detection systems play an important role as they can help detect even unknown attacks. Deep learning which has gained a great attention in the last few years due to excellent results in image, video and natural language processing is being used for anomaly detection in information security, particularly in SCADA networks. The salient features of the data from SCADA networks are learnt as hierarchical representation using deep architectures, and those learnt features are used to classify the data into normal or anomalous ones. This article is a review of various architectures such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Stacked Autoencoder (SAE), Long Short Term Memory (LSTM), or a combination of those architectures, for anomaly detection purpose in SCADA networks.

Keywords

ICS, SCADA, Unsupervised Feature Learning, Deep Learning, Anomaly Detection

1. Introduction

Industrial Control Systems (ICS) are used to monitor and control industrial sys-

tems. ICS are used to be isolated from enterprise networks, making attacks against them difficult. Moreover, these systems were using proprietary hardware, software and protocols. But as technology evolves, today's ICS use Commercial-Off-The-Shelf (COTS) software and hardware as well as open protocols such as Ethernet and TCP/IP. Things have worsened nowadays with the interconnection of ICS to enterprise network and to the Internet. A successful attack against industrial control system could have severe impact ranging from economy to loss of human lives [1] [2]. ICS attacks have already targeted water treatment systems, power grids or nuclear power plants [3] [4]. Some of the most famous attacks use Duqu, Flame [5], and the Stuxnet viruses [6]. Although many countermeasures are deployed to secure ICS networks, Intrusion and anomaly detection systems are important complementary security measures used to protect them.

In recent years, Deep Learning [7] became a hot topic among researchers with successes in domains such as natural language processing (NLP), image and video classification.

Various works are attempting to use deep learning for networks anomaly detection [8] [9] [10].

One of the most important features of deep learning is the use of unsupervised methods to autonomously learn hierarchical features in deep learning models [7] [11] [12] [13] [14].

In fact, the data most salient features are unsupervisingly learnt using the automatic learning capability of deep architectures, and those learnt features are used in a classifier to discriminate anomalous data from normal ones.

In this paper we are making a review of SCADA networks anomaly detection systems which are using deep feature learning approach.

After some highlights on the concept of the unsupervised feature learning in the next section, the third section is dedicated to the review of different anomaly detection systems in SCADA networks using deep unsupervised feature learning. In section four, we draw a conclusion of the review.

2. Unsupervised Feature Learning

Feature learning consists in modeling the behavior of data from a subset of features by deriving new features from the original ones [15]. In standard machine learning, feature learning from data is a complex task as it requires experts of the domain to handcraft the original features in order to feed the machine learning algorithms with the best features. The data learning process could be supervised or unsupervised. The supervised learning also needs the intervention of human to correctly label the data, which is costly and error prone. To take advantage the huge amount of unlabeled data, deep learning algorithms can automatically learn important features from data in an unsupervised manner [7] [16]. Unsupervised feature learning main goal is to map the original features' set into a different representation more suited for a given machine learning task [17]. Deep

architectures help in building complex non-linear functions to better fit real world complex data [18]. Unsupervised feature learning can be done by using clustering on data using algorithms such as K-means [19], or by training stacked auto-encoders or convolutional networks [20].

3. Review of Unsupervised Feature Learning in SCADA Anomaly Detection Systems

3.1. LSTM/Bloom Filter Anomaly Detector

In order to detected anomalies due to data/command injection, reconnaissance or Denial-of-Service (DoS) attacks on a gas pipeline SCADA system, [8] propose an anomaly detection approach consisting of two detectors (**Figure 1**). The first one is a packet-level anomaly detector which checks a packet signature in its database. The database stores network patterns and communication pattern signature as they are stable in a SCADA system. If the Bloom filter does not contain the signature the analyzed package, the packet is considered anomalous. The next detector receives normal packet that pass the Bloom filter for another detection level, which uses its power of information memorization for number of time steps to predict the behavior of the next time step.

Because of the limited memory and computing resources of some of SCADA components, using a fast and light-weighted anomaly detector as a Bloom filter is of high importance. The LSTM Anomaly Detector (**Figure 2**) which takes the input of time-series learns their important features in order to predict the next data point by being trained to minimize a softmax function suited for multi-class classification [7] [21].

The evaluation of the combined anomaly detection framework on a gas pipeline SCADA dataset [22] gives an accuracy of 92%, which is higher compared to other approaches. However, the time required to train the LSTM model of 35 min during 50 epochs is rather high.

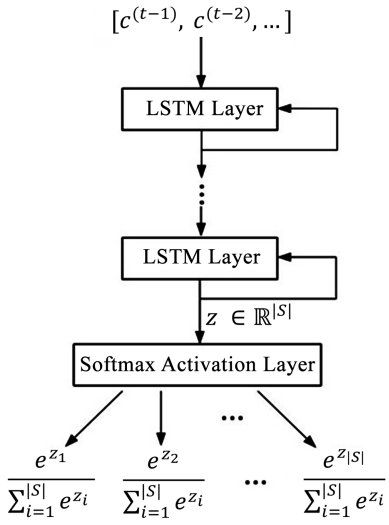

Figure 1. Stacked LSTM-based softmax model [8].

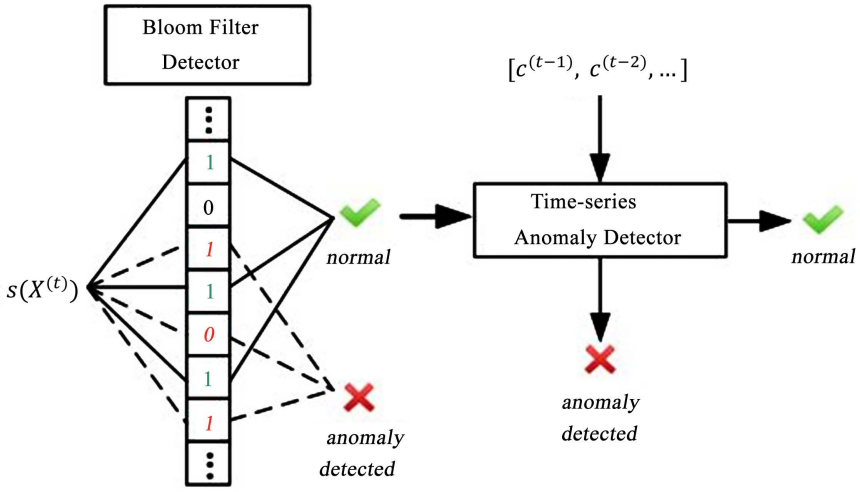

Figure 2. Combined framework for package and time-series level anomaly detection [8].

3.2. Stacked Auto-Encoder Based Anomaly Detection

Because of network bandwidth and data increase, [23] proposes a deep packet inspection in order to learn the necessary feature that would allow DoS, Probe, R2L and U2R attacks detection. The authors used a Deep Neural Networks (DNN) approach which architecture is a stacked auto-encoders for the feature learning, to which a softmax layer is added for the classification (**Figure 3**). The stacked auto-encoder has two hidden layers, one with 20 nodes and the second with 10 nodes. The dimension of the learnt features is 10 compared to the 41 original features of the NSL-KDD dataset dataset. The overall process encompasses four steps *i.e.* a feature learning step with the stacked auto-encoder, a first fine-tuning step with a supervised training of the softmax. The input of this first fine-tuning step is the compressed representation of the data. The following step is a second fine-tuning with a back-propagation training applied to the whole network layers after the first fine-tuning step. The second fine-tuning step aims at refining the features of the intermediate layers to make them more relevant for the intrusion detection task by adjusting the network weights to minimize the loss function.

Finally, the last step of the process is a classification and testing step where a test dataset is presented to the fine-tuned network to assess the efficiency of the model. Recall, accuracy, precision, and f-mesure metrics are used to evaluate the proposed approach against standard techniques like k-means, DBN, SOM, AdaBoost.

Experimental results show that despite good detection accuracy for DoS and Probe attacks (97.6% and 86.34% respectively), R2L and U2R attacks give poor results (12.98% and 39.62% respectively). The poor performance of the latter two categories of attacks is due to the lack of sufficient amount of data related to R2L and U2R (0.04% and 0.79% respectively). 9% to 10% training data samples for R2L and U2R categories of attacks as with the probe attacks would have given better detection results.

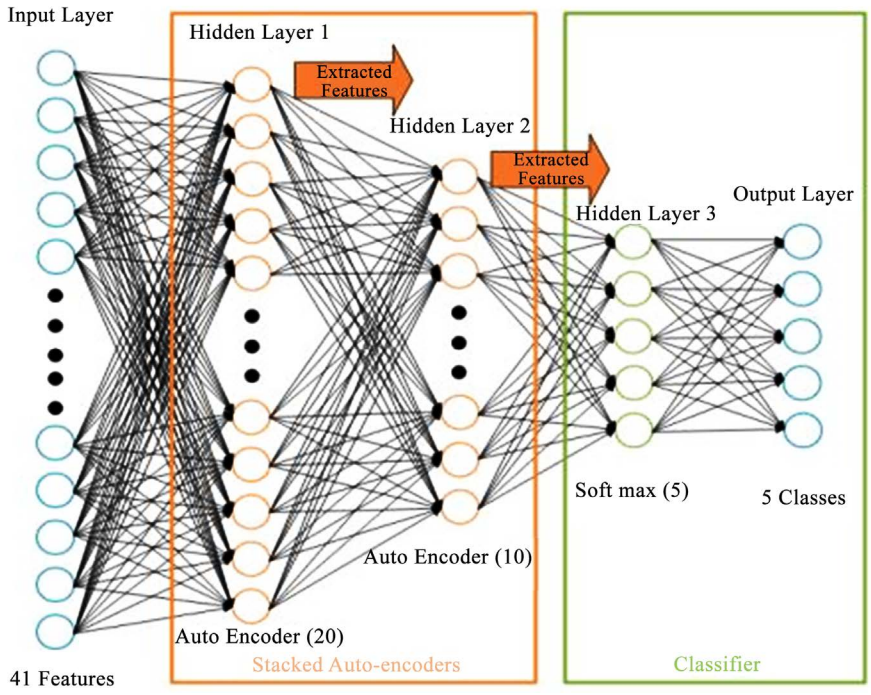

Figure 3. DNN architecture [23].

However, the approach proposed by [23] gives promising results in feature learning and good detection rate for some classes of attacks detection. It uses the NSL-KDD dataset for the experiment. Those datasets may not reflect modern networks traffic complexity nor integrate new complex attacks.

3.3. Stacked Auto-Encoder for Anomaly Detection in Smart Grids

The cyber-physical integration, exposes smart grids to large attack surface with potential severe consequences. Among the countermeasures against such attacks, Intrusion/Anomaly Detection Systems play a key role [24]. Machine learning approaches are used to develop data-driven anomaly detection systems. However, human handcrafted features for machine learning anomaly detectors are costly and ineffective in smart grids [25] [26]. This situation led [27] to use a stacked auto-encoder approach for a better feature learning for anomaly detection (**Figure 4**). The approach has two main phases: The model is first trained off-line and then follows online monitoring step. During the first phase, historical data are first collected for training purpose on different system operating conditions.

Then, the stacked auto-encoder is used to learn and deliver strong and high-level features. Finally in the off-line training phase, all the building blocks are stacked and a classifier is appended to them. The obtained architecture is then supervisingly trained using back-propagation. Next, to the training process is the acquisition of measurements from SCADA in the transmission system. These measurements are fed to the deep neural network, and the results of the classification are used for applications such as situational awareness.

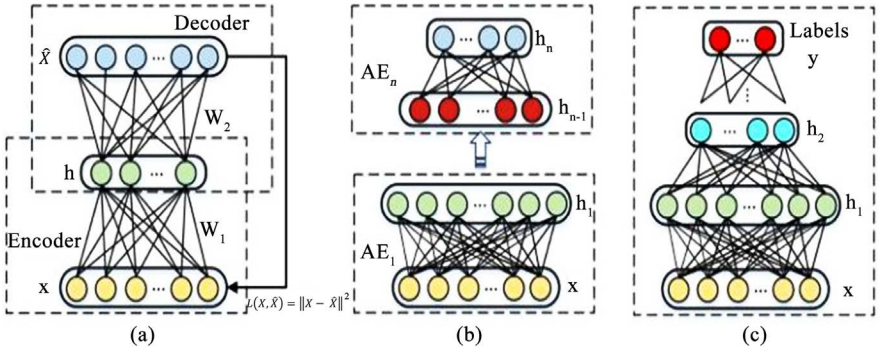

Figure 4. Stacked autoencoders training process [27].

A testbed simulating a power grid is used to evaluate the proposed approach (**Figure 5**). The proposed approach that use unsupervised feature learning achieves over 96% in accuracy doing slightly better than the supervised approaches used in the study. Furthermore, it provides an adaptive and automatic intrusion detection for smart grid environments

3.4. CNN/LSTM Anomaly Detection in SCADA

The Secure Water Treatment testbed (SWaT) dataset contains up to 36 different cyber-attacks. To evaluate the use of unsupervised feature learning for intrusion detection in such system, [28] proposes two models using either LSTM (**Figure 6**) or 1D CNN as feature learner (**Figure 7**).

They use mean MSE as an error function and AdamOptimizer with weight decay. The weight decay as a regularisation technique prevent model overfitting and the AdamOptimizer [29] is computationally efficient and require little memory. The first Deep Neural Network (DNN) architecture is a stacked LTSM with a fully connected layer at the top for classification purpose. With the LSTM model, setting a learning rate (between 0.001 and 0.00001), and a decay value (from 0.9 to 0.99) they were able to test various depths of LSTM layers (from 64 to 2048) and sequence lengths (between 50 and 1000). The 1D CNN architecture adopted the ReLU-MaxPooling scheme. Different kernel sizes were used for the experimentations. On top of the convolutions layer, a fully connected layer is added for prediction, and dropout is used to prevent overfitting. The authors tested diverse variations of this CNN architecture, by adding a batch normalization layer or by replacing the basic CONV-RELU-POOL block with (CONV-RELU) × N-MAXPOOL architecture. They also replaced the convolutionals layers by Inception layers [30] know to provide better performance and lower computational cost. The Inception layers use sparse network connections instead of the fully connections used by convolution layers, hence the reduction of the computational overhead. The experiments were conducted on the SWaT dataset which has 36 different cyberattacks. The proposed 1 D CNN model has 89% of detection rate, which is fairly good, but need to be improved.

The comparison of the different architectures (**Figure 8**) shows that LSTMs and inception-based convolution not only converge faster, but also yield to

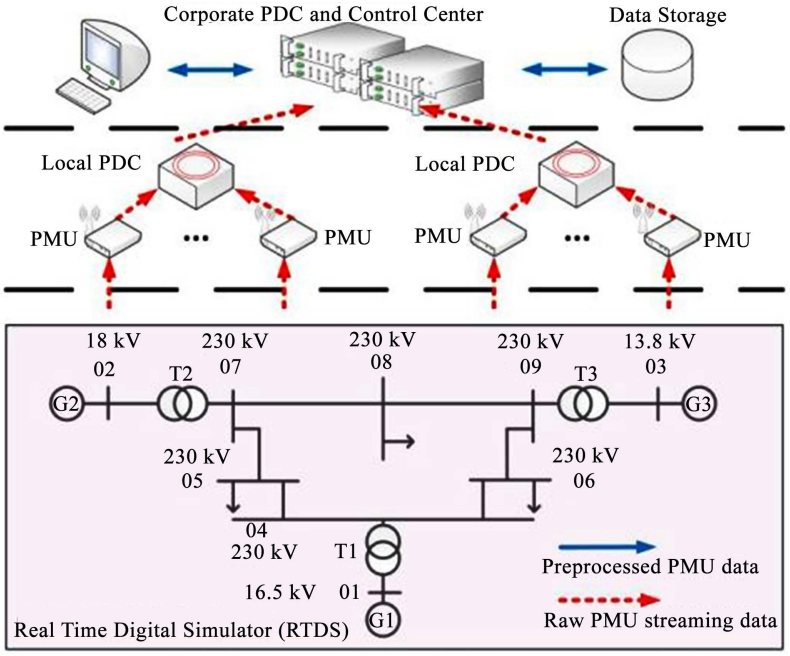

Figure 5. Smart grid benchmark testbed [27].

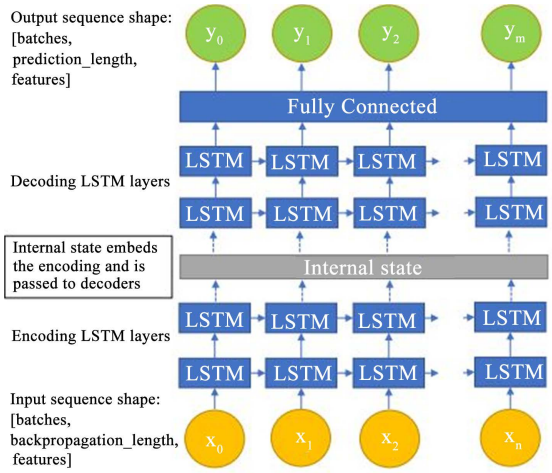

Figure 6. LSTM autoencoder model [28].

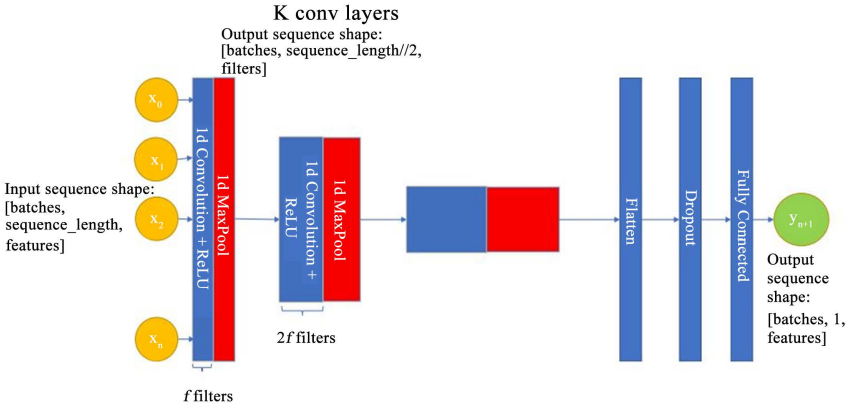

Figure 7. 1D convolutional neural network [28].

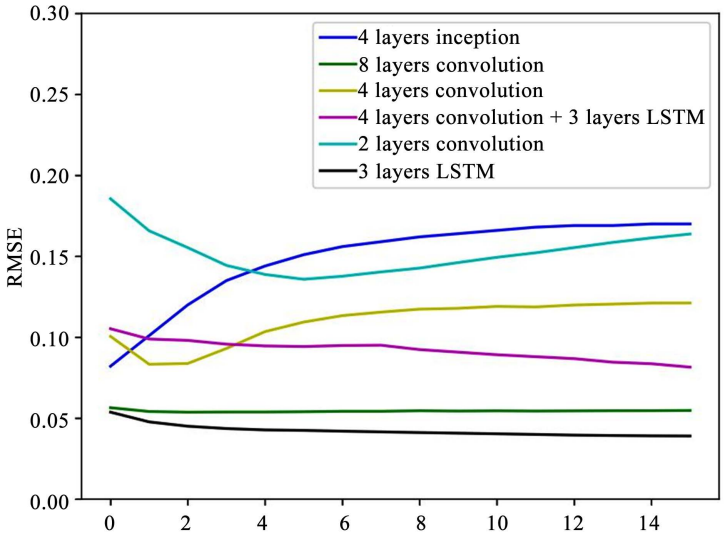

Figure 8. Test errors [28].

lower training error rate. The anomaly detection method gives high Area Under Curve (AUC), *i.e.* 0.967 for the eight layers convolutional network. The training and testing times of CNN are lower compared to LSTM network. The CNN networks performed well for anomaly detection compared to LSTM ones. The proposed CNN has a detection rates reaching 85% with a 100% precision.

3.5. Conditional Deep Belief Networks for False Data Injection in Smart Grid

As a countermeasure for False Data Injection (FDI) attack for electricity theft in smart grids, [31] proposes a detection mechanism which is formed of a State Vector Estimator (SVE) and a Deep-Learning Based Identification (DLBI) scheme. When the FDI attack bypass the SVE engine, the Deep Learning-Base Identification (DLBI) tries to detect the tampered data. The proposed Deep Neural Network is a Conditional Deep Belief Network (CDBN) that integrates the standard Deep Belief Network (DBN) with Conditional Gaussian-Bernoulli RBM (CGBRBM) (**Figure 9**). CGBRBM uses real value data and can model the impact of previously observed data on the current behavior feature learning. The use of CDBN allows the analysis of temporal attacks patterns [32]. On the other hand, using CGBRBM on the first hidden layer and regular RBM for the other hidden layer reduces the training and execution time of CDBN architectures.

An unsupervised approach is used to train the proposed CDBN and a fully connected layer is added on top of the model with a binary output node which has a sigmoid activation function. The whole deep neural network structure is then fine-tuned with back-propagation supervised training with labeled data. The proposed CDBN efficiently reveal the high-dimensional temporal behavior features of the unobservable FDI attacks that bypass the SVE mechanism with a high accuracy rate over 94% even in the presence of occasional operation faults, meaning that unknown attacks could be detected.

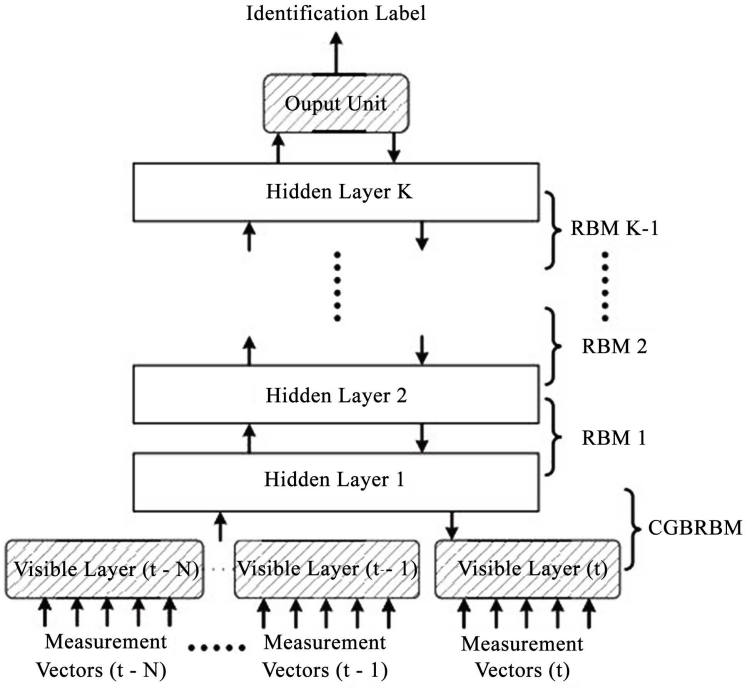

Figure 9. CDBN architecture [31].

3.6. Anomaly Detection© Using RBM-Based Deep Autoencoder

Wind turbines usually operate in harsh and variable environment, making various parts subject to failure. This situation can lead to unavailability and even destruction, causing important maintenance costs. As a remedy of this situation, the authors [33] present a deep auto-encoder (DAE) approach to detect early anomalies as well as provide fault analysis of wind turbines parts. The data associated to each wind turbine component is extracted in order to build the DAE model which is composed of stacked RBM [7]. The use of a DAE based on RBM building blocks is because of the power of RBM in highly capturing the variational potential of input data [34].

Two major steps are involved in the DAE training process *i.e.* pre-training and fine-tuning. The former is a layer-wise pre-training of each composing RBM, while the latter allows the initialization of the deep auto-encoder. During the pre-training phase, the long-term normal operating unlabeled SCADA data is used.

Following the pre-training phase which initializes the weights and bias of the DAE is the fine-tuning step. The back-propagation algorithm uses the normal operation labeled data for a supervised learning. The SCADA data fed to the DAE is encoded, then decoded, and a reconstruction error is calculated (**Figure 10**).

A SCADA data samples obtained from wind turbine normal operation is used to train the DAE. The training process allows the DAE to extract the internal relationship between the input and the output, and setup the model parameters. An index of part health condition is obtained by the reconstruction error of the

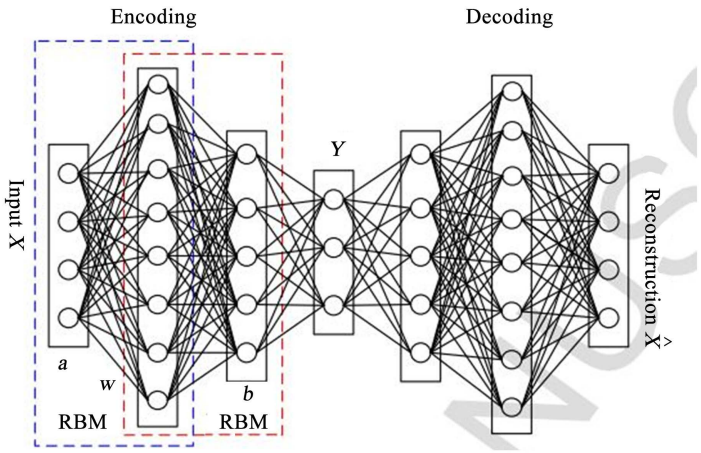

Figure 10. Structure of DAE network [33].

DAE. A dynamic adaptive threshold is then used for a better monitoring. The anomaly detection and fault location analysis is performed by combining the reconstruction error, the adaptive threshold and the input-output residual.

The proposed DAE model allows early fault detection as well as avoids false alarms. Moreover, it allows the determination of the location of the faulty component.

3.7. Gas Turbine Combustors Monitoring with Stacked Denoising Auto-Encoder and Extreme Learning Machine

In order to monitor gas turbine combustors' health and detect abnormal behaviors and incipient faults earlier, [35] proposes a deep neural network approach. The proposed model is a Stacked Denoising Auto-encoder (SDAE) [20], to which an Extreme Learning Machine (ELM) [13] is added (**Figure 11**). The SDAE used for the unsupervised learning of features allow more robust feature learning, even though the input data is noisy. The feature learned from the SDAE is fed to the ELM module for classification purpose.

Unlike in other feedforward neural networks, in ELM, don't need to be trained. Unveiling the connections between hidden and output nodes is ELM training method, which is fast [36]. The only ELM design parameter is the number of hidden neurons. To test the proposed approach, the authors have used seven months of one turbine data containing normal and abnormal data. In order to demonstrate the effectiveness of unsupervised feature learning for combustor anomaly detection, the authors compare classification performance between using the learned features and handcrafted features (**Figure 12**). The results show that the deep learned features give significant better classification performance than the handcrafted features (detection rate of 99% and 96% for deep learned features and the handcrafted features respectively).

4. Summary of Studied Approaches

Table 1 shows a summary of the different approaches.

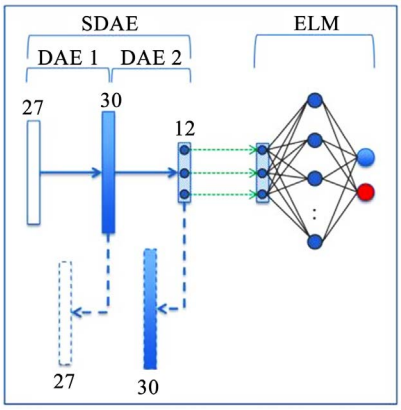

Figure 11. Structure of unsupervised feature learning for combustor anomaly detection [35].

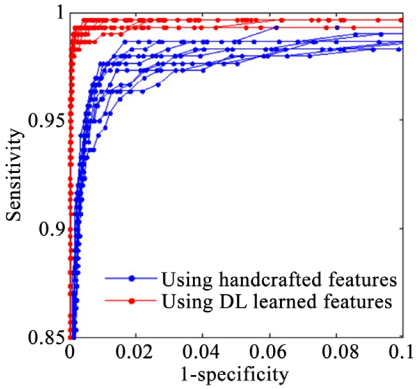

Figure 12. ROCs comparison [33].

Table 1. Summary of deep learning unsupervised feature learning in ICS.

Method	Feature learner	Classifier	Anomaly	Accuracy
SVE + CDBN [31]	CDBN RBM	FCNN	FDI	> 94%
Stacked LSTM + Bloom Filter [8]	Stacked LSTM	Softmax	- Data injection - Command injection - Reconnaissance - DoS	92%
Stacked AE + Softmax [23]	SAE	Softmax	- DoS - Probe - R2L - U2R	97.6% 86.34% 12.98% 39.62%
CNN/LSTM + FCNN [28]	CNN/LSTM	FCNN	36 attacks	92% (F1-score)
DAE-RBM [33]	DAE	DAE Residuals	- Operating anomaly detection - Fault analysis	N/A
SAE + MLP [27]	SAE	MLP	- Data injection - Comman injection - Relay setting modification	96%
SDAE + ELM [35]	SDAE	ELM	Operating faults	99%

The deep architectures are formed with stacked autoencoders, convolutional neural networks, long short term memories or deep belief networks, or by combining these architectures. Those deep architectures are used to learn the SCADA networks features and softmax, fully connected neutral network, multilayer perceptron or extreme learning machine are used for the classification. For each approach we highlight the feature learning architecture, the classifier used to discriminate the data, the types of the attacks detected and the results in terms of accuracy.

5. Conclusions

Deep Learning approaches are more and more used for anomaly detection in SCADA systems. The unsupervised feature learning capability that makes it possible to learn important features from available SCADA network large data in order to deliver high anomaly detection rate contributes to the rising interest in deep learning approaches. Multiple architectures such as CNN, LSTM, DBN, SAE, SDAE or a combination of them are used to learn the SCADA data features, and classifiers such as Softmax layer, Fully connected neural network; ELM, DAE or MLP are used for the classification. In most situations, deep learning approaches outperform standard approaches, but their Achille's heel remains the high training time required for their training.

Interesting research direction took by the scientific community to overcome the high training time shortcoming is the use of distributed deep learning approaches for anomaly detection in Industrial Control Systems. In a future work, we will propose a distributed deep learning approach for anomaly detection in Industrial Control Systems.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Valdes, A. and Cheung, S. (2009) Intrusion Monitoring in Process Control Systems. 2009 42*nd Hawaii International Conference on System Sciences*, Waikoloa, 5-8 January 2009, 1-7.

[2] Zhu, B., Joseph, A. and Sastry, S. (2011) A Taxonomy of Cyber-Attacks on SCADA Systems. 2011 *International Conference on Internet of Things and* 4*th International Conference on Cyber*, *Physical and Social Computing*, Dalian, 19-22 October 2011, 380-388. https://doi.org/10.1109/iThings/CPSCom.2011.34

[3] Slay, J. and Miller, M. (2007) Lessons Learned from the Maroochy Water Breach. In: *International Conference on Critical Infrastructure Protection*, Springer, Boston, 73-82. https://doi.org/10.1007/978-0-387-75462-8_6

[4] Case, D.U. (2016) Analysis of the Cyber-Attack on the Ukrainian Power Grid. Electricity Information Sharing and Analysis Center (E-ISAC), Washington DC, 388.

[5] Miller, B. and Rowe, D. (2012) A Survey SCADA of and Critical Infrastructure Incidents. *Proceedings of the* 1*st Annual Conference on Research in Information*

Technology, Calgary, 10-13 October 2012, 51-56.
https://doi.org/10.1145/2380790.2380805

[6] Falliere, N., Murchu, L.O. and Chien, E. (2011) W32. Stuxnet Dossier. *Security Response*, **5**, 29.

[7] Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. (2016) Deep Learning (Vol. 1, No. 2). MIT Press, Cambridge.

[8] Feng, C., Li, T. and Chana, D. (2017) Multi-Level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks. 2017 47*th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (*DSN*), Denver, 26-29 June 2017, 261-272.
https://doi.org/10.1109/DSN.2017.34

[9] Javaid, A., Niyaz, Q., Sun, W. and Alam, M. (2016) A Deep Learning Approach for Network Intrusion Detection System. *Proceedings of the* 9*th EAI International Conference on Bio-Inspired Information and Communications Technologies* (*Formerly BIONETICS*), New York City, 3-5 December 2015, 21-26.
https://doi.org/10.4108/eai.3-12-2015.2262516

[10] Kim, J., Kim, J., Thu, H.L.T. and Kim, H. (2016) Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection. 2016 *International Conference on Platform Technology and Service*, Jeju, 15-17 February 2016, 1-5.
https://doi.org/10.1109/PlatCon.2016.7456805

[11] Ranzato, M.A., Boureau, Y.L. and Cun, Y.L. (2008) Sparse Feature Learning for Deep Belief Networks. In: Platt, .C., Koller, D., Singer, Y. and Roweis, S.T., Eds., *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, 1185-1192.

[12] Kwon, D., Kim, H., Kim, J., Suh, S.C., Kim, I. and Kim, K.J. (2019) A Survey of Deep Learning-Based Network Anomaly Detection. *Cluster Computing*, **22**, S949-S961.
https://doi.org/10.1007/s10586-017-1117-8

[13] Linda, O., Vollmer, T. and Manic, M. (2009) Neural Network Based Intrusion Detection System for Critical Infrastructures. 2009 *International Joint Conference on Neural Networks*, Atlanta, 14-19 June 2009, 1827-1834.
https://doi.org/10.1109/IJCNN.2009.5178592

[14] Mohammadpour, L., Ling, T.C., Liew, C.S. and Chong, C.Y. (2018) A Convolutional Neural Network for Network Intrusion Detection System. *Proceedings of the Asia-Pacific Advanced Network*, Auckland, 5-8 August 2018, 50-55.

[15] Kim, K., Aminanto, M.E. and Tanuwidjaja, H.C. (2018) Network Intrusion Detection Using Deep Learning: A Feature Learning Approach. Springer, Berlin.
https://doi.org/10.1007/978-981-13-1444-5

[16] Xu, Q., Zhang, C., Zhang, L. and Song, Y. (2016) The Learning Effect of Different Hidden Layers Stacked Autoencoder. 2016 8*th International Conference on Intelligent Human-Machine Systems and Cybernetics* (*IHMSC*), Vol. 2, 148-151.
https://doi.org/10.1109/IHMSC.2016.280

[17] Yousefi-Azar, M., Varadharajan, V., Hamey, L. and Tupakula, U. (2017) Autoencoder-Based Feature Learning for Cyber Security Applications. 2017 *International Joint Conference on Neural Networks* (*IJCNN*), Anchorage, 14-19 May 2017, 3854-3861. https://doi.org/10.1109/IJCNN.2017.7966342

[18] Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.A. (2008) Extracting and Composing Robust Features with Denoising Autoencoders. *Proceedings of the* 25*th International Conference on Machine Learning*, Helsinki, 5-9 July 2008, 1096-1103.
https://doi.org/10.1145/1390156.1390294

[19] Bengio, Y., Courville, A.C. and Vincent, P. (2012) Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. ArXiv.

[20] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A. and Bottou, L. (2010) Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research*, **11**, 3371-3408.

[21] Patterson, J. and Gibson, A. (2017) Deep Learning: A Practitioner's Approach. O'Reilly Media, Inc., Sebastopol.

[22] Morris, T. and Gao, W. (2014) Industrial Control System Traffic Data Sets for Intrusion Detection Research. In: *International Conference on Critical Infrastructure Protection*, Springer, Berlin, 65-78. https://doi.org/10.1007/978-3-662-45355-1_5

[23] Potluri, S. and Diedrich, C. (2017) Deep Feature Extraction for Multi-Class Intrusion Detection in Industrial Control Systems. *International Journal of Computer Theory and Engineering*, **9**, 374-379. https://doi.org/10.7763/IJCTE.2017.V9.1169

[24] Axelsson, S. (2000) Intrusion Detection Systems: A Survey and Taxonomy (Vol. 99). Technical Report.

[25] Li, L., Ota, K. and Dong, M. (2017) When Weather Matters: IoT-Based Electrical Load Forecasting for Smart Grid. *IEEE Communications Magazine*, **55**, 46-51. https://doi.org/10.1109/MCOM.2017.1700168

[26] Nie, D., Zhang, H., Adeli, E., Liu, L. and Shen, D. (2016) 3D Deep Learning for Multi-Modal Imaging-Guided Survival Time Prediction of Brain Tumor Patients. In: Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W., Eds., *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 212-220. https://doi.org/10.1007/978-3-319-46723-8_25

[27] Wilson, D., Tang, Y., Yan, J. and Lu, Z. (2018) Deep Learning-Aided Cyber-Attack Detection in Power Transmission Systems. 2018 *IEEE Power & Energy Society General Meeting* (*PESGM*), Portland, 5-10 August 2018, 1-5. https://doi.org/10.1109/PESGM.2018.8586334

[28] Kravchik, M. and Shabtai, A. (2018) Detecting Cyber-Attacks in Industrial Control Systems Using Convolutional Neural Networks. *Proceedings of the* 2018 *Workshop on Cyber-Physical Systems Security and Privacy*, Toronto, 19 October 2018, 72-83. https://doi.org/10.1145/3264888.3264896

[29] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization.

[30] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2016) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.

[31] He, Y., Mendis, G.J. and Wei, J. (2017) Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism. *IEEE Transactions on Smart Grid*, **8**, 2505-2516. https://doi.org/10.1109/TSG.2017.2703842

[32] Wei, J. and Mendis, G.J. (2016) A Deep Learning-Based Cyber-Physical Strategy to Mitigate False Data Injection Attack in Smart Grids. 2016 *Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids* (*CPSR-SG*), Vienna, 12 April 2016, 1-6. https://doi.org/10.1109/CPSRSG.2016.7684102

[33] Zhao, H., Liu, H., Hu, W. and Yan, X. (2018) Anomaly Detection and Fault Analysis of Wind Turbine Components Based on Deep Learning Network. *Renewable Energy*, **127**, 825-834. https://doi.org/10.1016/j.renene.2018.05.024

[34] Fiore, U., Palmieri, F., Castiglione, A. and De Santis, A. (2013) Network Anomaly Detection with the Restricted Boltzmann Machine. *Neurocomputing*, **122**, 13-23.

https://doi.org/10.1016/j.neucom.2012.11.050

[35] Yan, W. and Yu, L. (2019) On Accurate and Reliable Anomaly Detection for Gas Turbine Combustors: A Deep Learning Approach. ArXiv.

[36] Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006) Extreme Learning Machine: Theory and Applications. *Neurocomputing*, **70**, 489-501. https://doi.org/10.1016/j.neucom.2005.12.126