Unsupervised anomaly detection of the gas turbine operation via convolutional auto-encoder

Geunbae Lee
Graduate School of Computer & Information Technology
Korea University
Seoul, Korea
escavecha@korea.ac.kr

Myungkyo Jung
Graduate School of Computer & Information Technology
Korea University
Seoul, Korea
mkyo@korea.ac.kr

Myoungwoo Song
Graduate School of Computer & Information Technology
Korea University
Seoul, Korea
supwrman@korea.ac.kr

Jaegul Choo
Graduate School of Artificial Intelligence
KAIST
Daejeon, Korea
jchoo@kaist.ac.kr

***Abstract*—This paper proposes a combination of convolutional neural network and auto-encoder (CAE) for unsupervised anomaly detection of industrial gas turbines. Autonomous monitoring systems protect the gas turbines, with the settings unchanged in their lifetime. Those systems can not detect any abnormal operation patterns which potentially risk the equipment after long-term exposure. Recently, machine learning and deep learning models are applied for industries to detect those anomalies under the nominal working range. However, for gas turbine protection, not much deep learning (DL) models are introduced. The proposed CAE detects irregular signals in unsupervised learning by combining a convolutional neural network (CNN) and auto-encoder (AE). CNN exponentially reduces the computational cost and decrease the amount of training data, by its extraction capabilities of essential features in spatial input data. A CAE identifies the anomalies by adapting characteristics of an AE, which identifies any errors larger than usual pre-trained, reconstructed errors. Using the Keras library, we developed an AE structure in one-dimensional convolution layer networks. We used actual plant operation data set for performance evaluation with conventional machine learning (ML) models. Compared to the isolation forest (iforest), k-means clustering (k-means), and one-class support vector machine (OCSVM), our model accurately predicts unusual signal patterns identified in the actual operation than conventional ML models.**

***Keywords*—gas turbine, anomaly detection, convolution, auto-encoder, deep learning, machine learning, unsupervised learning**

I. Introduction

The construction of natural gas plants is on the rise as alternatives to nuclear power plant shutdowns and pollutant emission reduction. According to the electricity information 2019 by international energy agency, the gross electricity generated from natural gas has increased by 3.5%, while that of coal has decreased by 6.8% from 2010 to 2017 in OECD countries[1]. The gas turbine, a vital component of the natural gas power plant, monitors real-time operation data to determine any anomalies during the operation.

1. OECD gross electricity production by source, 1974 – 2018 provisional. (www.iea.org/reports/electricity-information-2019)

Gas turbines installed in the plant configured with the protection measures, such as a vibration monitoring system (VMS) and a plant distributed control system (DCS). Both VMS and DCS intervenes in regular operation in a way that they automatically stop the operation of the measurement exceeds the preset threshold, or arbitrarily stops the process at the discretion of the user before reaching the limit. In other words, deployed protection systems will not trigger alarms for any outliers below the preset value. In general, vibration signals are easy to obtain, and detailed monitoring guidelines are provided [1, 2]. However, other data, such as exhaust gas temperature, bearing temperatures, rotating speed, and generator current, are also collected from plant DCS. These signals can be useful in detecting the anomaly signal.

ML and DL approaches have also applied in anomaly detection of gas turbines. Anomalies or outliers are defined as those data instances that deviate from the majority of the data. Anomaly detection is the task of successfully identifying then records within a given dataset [3].

Abnormal signals are difficult to find, and the irregular-labeled data set is usually scarce or does not even exist in some cases. When the recorded operation data seems unusual, but below the threshold value, the determination of the anomaly depends on the operator's experience. In this regard, any classifiers based on supervised learning is not applicable to fit in anomaly detection. In practice, iforest, k-means, and OCSVM are widely-used machine learning approaches for anomaly detection in various fields, as they are well working on unlabeled training data [4-6].

For deep learning applications, a CAE and its variant have recently introduced [7-9]. A CAE model achieved excellent performance in classification and anomaly detection of unlabeled data, as both convolution and auto-encoder can learn and extract essential features in spatial input data in an unsupervised setting.

We propose a CAE model to use to detect actual anomalies through the unsupervised learning of the operation signal records of the industrial gas turbines. To the best of our knowledge, this is the first time that the CAE model has applied

978-1-7281-6286-7/20/$31.00 ©2020 IEEE

to diagnose the data from the gas turbine components. The contribution of this paper is as below:

- We analyzed the actual VMS and DCS data collected from the industrial gas turbine in an operating plant.
- We proposed an unsupervised CAE model to identify the abnormal patterns of the gas turbine operation data.
- The proposed model achieves the highest performance in collected data compared to conventional machine learning models.

The rest of this paper has organized as follows: In Section 2, we discuss the related work on anomaly detection. In Section 3, the proposed CAE model architecture described. Experiments and conclusions are presented in Sections 4 and 5, respectively.

II. Related Work

Detecting anomalies or outliers in data has been studied in the statistics community as early as the 19th century [10]. Thanks to the development of extensive data processing methods and platforms in the last decade, anomaly detection has become the mainstream of machine learning applications in fraud detection of financial transactions[11], fault detection of networks [5], and medical pattern recognition [3, 7].

For specific applications on gas turbines, recent studies developed from both machine learning and deep learning models. For the machine learning approach, S. Zhong [12] used iforest to detect anomalies in the unlabeled data of the aero-derivative gas turbines using exhaust gas temperatures and core speed. W. Yan and L. Yu [13] improvised a combination of stacked denoising auto-encoder (SDAE) and extreme learning machine (ELM) for combustor anomaly detection of an industrial gas turbine. In their deep learning model, an SDAE used for unsupervised feature learning, then ELM performed the classification of given features.

In power industries, statistical models and machine learning models were the mainstream of researches for anomaly detection. As deep learning models can perform feature extraction works in an automated and efficient manner, we expect more deep learning models for gas turbine applications will be available shortly.

III. The Model Architecture

Before describing the details of our experiment, we explain the main components of the proposed algorithm, followed by a description of the detailed architecture of CAE.

A. Convolutional neural network

CNNs are the network that employs a specialized kind of linear operation called convolution [14]. CNNs have applied to analyze the visual image, based on their shared-weights architecture and translation invariance characteristics [15]. Concept of CNN specified in fig.1.

1) Convolution: By choosing the filter, each input data volume reduced into a smaller scale called the feature map. Convolution is a dot product between the filter and input data, described as:

$$y_i^{l+1}(j) = K_i^l * x^l(j) + b_i^l \tag{1}$$

where K_i^l and b_i^l denote the weights and bias of the i-th layer, and $x^l(j)$ is the j-th local region of the layer l. $*$ denotes a dot product operation, and $y_i^{l+1}(j)$ is the output of convolution operation, respectively [16].

2) Nonlinearity: Nonlinearity is an activation function that calculates a weighted sum of inputs with a bias. Nonlinearity acts as a threshold to the given information. In this paper, a rectified linear unit (ReLU) used for the activation function of input x as [17]:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{2}$$

3) Max pooling: Max pooling reduces the dimension of feature maps by taking the biggest value among every window. By reducing the number of parameters, the feature dimension becomes smaller and manageable.

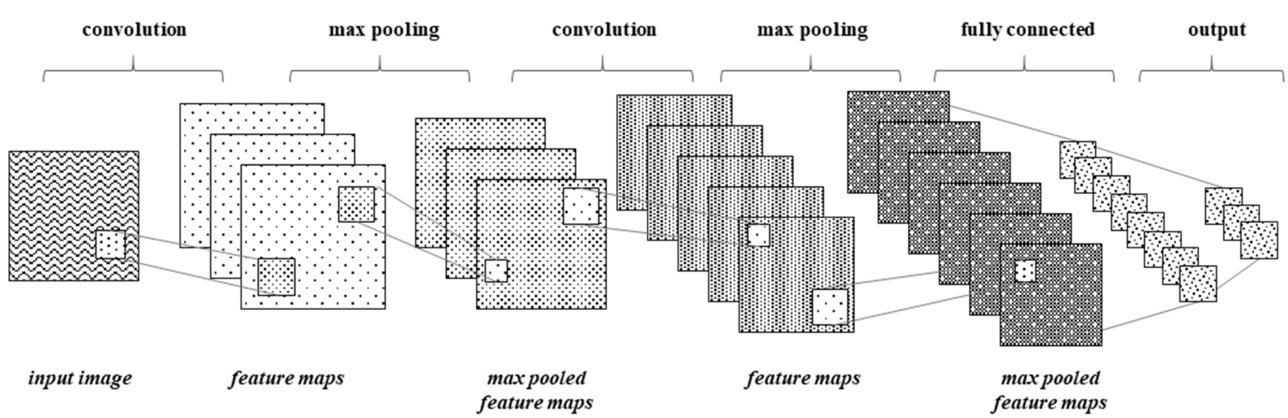

Fig.1. the typical concept of the convolutional neural networks

4) Fully connected layers: A fully connected layer or a dense layer is a multilayer perceptron which uses a softmax activation function at the end of the layer. This layer uses high-level features derived from the convolutional and max-pooling layers for the classification of input data.

B. Auto-encoder

AE is an unsupervised model that trained to generate the target values equal to the input data through a combination of encoder and decoder network. These networks have a bottleneck hidden layer of a few neurons in the middle, forcing them to generate valid representations that compress the input data into a lower-dimensional code called a latent vector, which used by the decoder to reproduce the original information. The process is described as:

$$\underset{D,E}{\operatorname{argmin}} \parallel X - D(E(X)) \parallel \tag{3}$$

Where X is the input data, E is an encoder network, and D is a decoder network. A typical auto-encoder architecture consists of three main components, as described in fig.2.

1) Encoder network: An encoder network comprised of a series of layers with a decreasing number of nodes and ultimately reduces input data into a latent vector. This process is also called dimensionality reduction, and the convolution layers used for encoding.

2) Latent vector: The latent vector represents the lowest level space in which the inputs are reduced, with essential information preserved with the strong correlation between input features.

3) Decoder network: A decoder network acts as the mirror image of the encoder network. The number of nodes in every layer increases and reconstructs the latent vector to output as a similar input via transposed convolution. As a particular portion of the information lost during reconstruction, the output data always have lower quality than the input data.

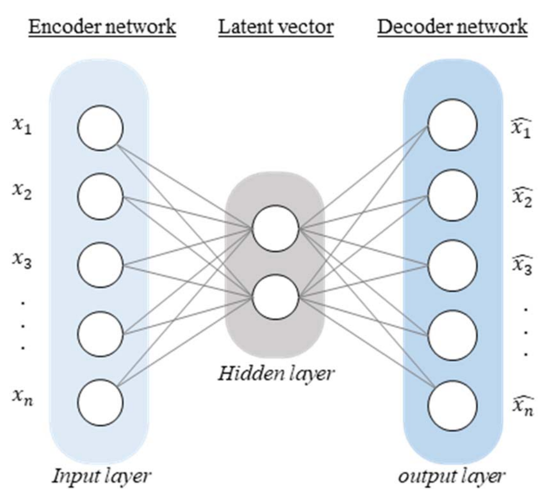

Fig.2. the typical configuration of the auto-encoder

C. The model architecture of CAE

1) Motivation: A deeper network can exponentially reduce the computational cost and decrease the number of parameters by increasing the number of nonlinearity. Empirically, higher depth architecture tends to express a useful prior over the space of functions the model learns [18]. The CAE adapts stacks of layers of convolution and max-pooling to reduce the overall dimension of input data while critical information extracted. A CAE trained on unlabeled data set without any anomalies will identify the anomalies when feeding the test set with the abnormal data. This characteristic is originated from an AE which identifies any errors larger than usual pre-trained, reconstructed errors.

2) Advantage of a CAE: A CAE architecture is useful in the classification of unlabeled image data, compared to CNN and k-means [19]. CAE models also proved useful in processing one-dimensional(1-D) data, such as sensor signals in sequence. X. Liu applied a combination of 1-D CNN and denoising CAE to improve the anti-noise diagnosis of bearing fault signals[16], and C.Fan proved an ensemble of autoencoders in 1-D CNN layers capable of detecting anomalies in the building energy consumption data [20].

In our CAE model architecture, we applied 1-D convolutions as illustrated in Fig.3 and Table 1. Table 1 shows the configuration summary of the model using the TensorFlow library (tensorflow.org).

TABLE I. CONFIGURATION OF THE LAYERS IN CAE MODEL

No	Layer	Output shape	No. of parameters
0	Input	(none, 32, 5)	0
1	1-D Conv.	(none, 32, 100)	2,600
2	Max pool.	(none, 16, 100)	-
3	1-D Conv.	(none, 16, 50)	25,050
4	Max pool.	(none, 8, 50)	-
5	1-D Conv.	(none, 8, 25)	6,275
6	Max pool.	(none, 4, 25)	-
7	Flatten	(none, 100)	-
8	encoder	(none, 2)	202
9	Dense	(none, 100)	300
10	Reshape	(none, 4, 25)	-
11	Upsample	(none, 8, 25)	
12	1-D Conv.	(none, 8, 50)	6,300
13	Upsample	(none, 16, 50)	
14	1-D Conv.	(none, 16, 100)	25,100
15	Upsample	(none, 32, 100)	-
16	1-D Conv.	(none, 32, 5)	2505

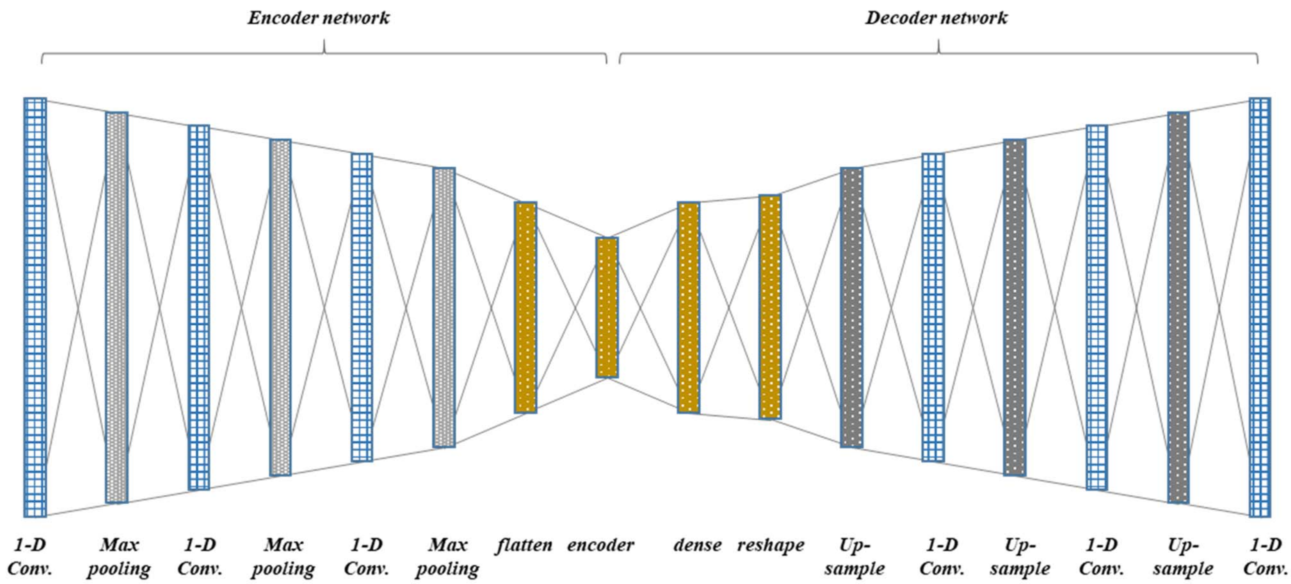

Fig.3. a proposed CAE model structure

IV. Experiments

We described information on the data set used in table 2. A deep learning model automates feature engineering works in machine learning models. Still, data cleansing and hyperparameters tuning depend on the user's discretion.

TABLE II. A CONFIGURATION OF DATA SET

No	Attributes in columns	Signal collected from
1	Vibration in µm	No.1 rotor bearing (x-axis)
2	Vibration in µm	No.1 rotor bearing (y-axis)
3	Vibration in µm	No.2 rotor bearing (x-axis)
4	Vibration in µm	No.2 rotor bearing (y-axis)
5	Vibration in µm	No.3 rotor bearing (x-axis)
6	Vibration in µm	No.3 rotor bearing (y-axis)
7	Vibration in µm	No.4 rotor bearing (x-axis)
8	Vibration in µm	No.4 rotor bearing (y-axis)
9	Pressure in mmH_2O	Exhaust gas duct
10	Pressure in mmH_2O	Compressor inlet air manifold
11	Temperature in ˚C	Compressor outlet air
12	Temperature in ˚C	Thrust bearing drain oil
13	Temperature in ˚C	No.2 row disc cavity
	Total # of samples	33.11M
Total # of anomalies in samples		44,549 (0.14%)

A. Data set

1) The data explanation: Data set sampled from an industrial gas turbine in operation. The data sets consist of sensor signals from VMS and DCS over 30 days collected simultaneously. However, the specific date and time information is not provided. In VMS, 8 signals collected from the vibration transmitters attached to the bearing housing of the turbine rotor. In DCS, signals provided from the 3 temperature and 2 pressure transmitters attached across to the combusters and rotor casing. The recording interval of the sample is at every second. Total data volume is approximately 33.11 million, which consisted of signal labels in 13 columns and sample values in 2,547,211 rows.

2) Pre-processing: The data includes sudden spike or dip patterns under the threshold alarm level. As there are no labels provided, we adapted the interquartile range (IQR) as a measure of the dispersion in the data. Any values fall below the lower quartile Q1 – 1.5 IQR, or above the upper quartile Q3 + 1.5 IQR, are considered as anomalies[21]. To identify normal and abnormal signals, we added the 14th column. In this column, any rows include anomalies are labeled to 1, while other rows without anomalies are labeled to 0. By this separation, the number of rows or columns contain any abnormal value is counted as 44,549. Those abnormal signals are consisting of 0.14% of the total data volume. For the training set, we have split the 14th column labeled to 0 until the size of the samples reaches about 80% of total data volume. The rest of the samples include anomalies are consisted of 20% of the total data volume and split for validation purposes. After the split, the 14th column has removed only for the CAE model. This point will be explained in section C. Lastly; each set has normalized to a range of 0 and 1 for further processing. All pre-processings committed on R.

B. Hyperparameters set up results

1) CAE: Layers compiled on Keras. The window size is 32. Stride set to 1. Mini batch size is 128. Kernel size is 5 for all convolutions. Max pooling size is 2. L2 regularization of 0.001 applied. The learning rate set to 0.01. For gradient descent optimizer, adaptive moment estimation (Adam) used [22]. The model is trained for 15 epochs. The test data set size is 20% of the total data volume.

2) iforest: Number of estimators is 100. Contamination value is 0.2.

3) k-means: Number of clusters is 8.

4) OCSVM: A gaussian kernel used for kernel function. Values for γ and ν are 0.3 and 0.5, respectively.

C. Performance evaluation

Technically speaking, there is no way to measure the performance of unsupervised learning models, since there are no labels available to compare the ground truth and predictions. In this regard, we used abnormal signals, as labels solely for models performance evaluation purposes. During the training process, labeled datasets are not provided for the selected CAE model. Other ML methods used labeled datasets for classification. For performance metrics, we applied the mean squared error (MSE) and an F-1 score. A brief explanation of precision and recall included as supplementary.

1) MSE: MSE calculates the average squared difference between the vectors of predicted values $\widehat{Y_i}$ and the correct values Y, as:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2 \tag{4}$$

The MSE indicates how the quality of the model has estimated the correct value, in the fraction between 0 and 1. When the estimation gets better during the learning process, the value goes closer to 0.

2) Precision: Precision means how the trained model has made more relevant estimation than irrelevant ones, which defined as:

$$p = \frac{TP}{TP+FP} \tag{5}$$

Where TP indicates the true positives, which means the sum of examples correctly classified as positive. FP indicates the false positives, which is the sum of negative examples incorrectly classified as positive. For our data set, abnormal samples are considered as positive examples.

3) Recall: Recall indicates that the trained model has returned most of the relevant results, which defined as:

$$r = \frac{TP}{TP+FN} \tag{6}$$

Where FN indicates the false negative, which is the sum of positive examples incorrectly classified as negative.

4) F-1 score: An F-1 score measures a test accuracy of binary classifications in the imbalanced data set. Precision and recall are related in a trade-off, an f-1 score is derived for balanced assessment with a single number, between precision and recall. An F-1 score is defined as a weighted average of precision and recall value of the test result, as:

$$f = \frac{2 \cdot p \cdot r}{p+r} \tag{7}$$

A higher F-1 score close to 1, in the fraction between 0 and 1, implies a higher classification performance is expected. In most classification cases, the threshold is typically set to 0.5 [23].

D. Test results

We have compared the performance metrics results of the applied method in table 3. Compared to other machine learning models, a CAE showed pair results of the identification of abnormal data. In the case of K-means and iForest, MSE is high, while an F-1 score is relatively low. An F-1 score lower or near the threshold implies that the estimation result is biased to precision or recall. The sensitivity to outliers may be improved, employing ensemble and boosting algorithms.

As mentioned in the configuration of the data set, the sparse volume of the abnormal data directly affects the performance of the algorithms. For OCSVM, a more viable volume of abnormal data set will be required to perform as expected.

TABLE III. PERFORMANCE COMPARISON WITH OTHER METHODS

Method	**Performance metrics**	
	MSE	***F-1***
OCSVM	0.9813	0.0347
k-means	0.0606	0.1050
iforest	0.0323	0.5220
CAE	0.0056	0.8748

V. CONCLUSIONS

We propose a CAE model for anomaly detection in the operation of the industrial gas turbines. We use a combination of a CNN and AE to maximize the advantage of anomaly detection and unsupervised learning capabilities. We analyze actual multivariate data from the operating power plant. We review the compatibility of the state-of-the-art deep learning algorithm into the real world scenario. We confirm that the proposed model proved more precise and accurately predicts irregular signal patterns identified in the actual operation than conventional machine learning models.

However, deep learning algorithms have their limits. For higher performance, more computational resources and time are essential. In urgent cases, accurate but slow models will not achieve desired purposes. Moreover, even though the deep learning automates feature engineering, hyperparameter tuning still depends on researchers' knowledge and experience. Achieved performance metric values imply that further developments are still required for the reliable application of deep learning models for industrial use. We will continue to focus on research of faster and robust deep learning models.

ACKNOWLEDGMENT

The data sets described in this paper have used under permission of Doosan heavy industries and construction, Changwon, Korea.

REFERENCES

[1] C. Wu, P. Jiang, C. Ding, F. Feng, and T. Chen, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network," *Computers in Industry,* vol. 108, pp. 53-61, 2019.

[2] K. Listewnik, G. Grzeczka, M. Kłaczyński, and W. Cioch, "An on-line diagnostics application for evaluation of machine vibration based on standard ISO 10816-1," *Journal of Vibroengineering,* vol. 17, no. 8, pp. 4248-4258, 2015.

[3] K.-O. Cho and H.-J. Jang, "Comparison of different input modalities and network structures for deep learning-based seizure detection," *Scientific Reports,* vol. 10, no. 1, pp. 1-11, 2020.

[4] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, 2013, pp. 8-15.

[5] M. Wazid and A. K. Das, "An efficient hybrid anomaly detection scheme using K-means clustering for wireless sensor networks," *Wireless Personal Communications,* vol. 90, no. 4, pp. 1971-2000, 2016.

[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 6, no. 1, pp. 1-39, 2012.

[7] N. Pawlowski *et al.*, "Unsupervised lesion detection in brain CT using bayesian convolutional autoencoders," 2018.

[8] Z. Alaverdyan, J. Chai, and C. Lartizien, "Unsupervised feature learning for outlier detection with stacked convolutional autoencoders, siamese networks and wasserstein autoencoders: application to epilepsy detection," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*: Springer, 2018, pp. 210-217.

[9] C. Aytekin, X. Ni, F. Cricri, and E. Aksu, "Clustering and unsupervised anomaly detection with l 2 normalized deep auto-encoder representations," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018: IEEE, pp. 1-6.

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR),* vol. 41, no. 3, pp. 1-58, 2009.

[11] P. H. Tran, K. P. Tran, T. T. Huong, C. Heuchenne, P. HienTran, and T. M. H. Le, "Real time data-driven approaches for credit card fraud detection," in *Proceedings of the 2018 International Conference on E-Business and Applications*, 2018, pp. 6-9.

[12] S. Zhong, S. Fu, L. Lin, X. Fu, Z. Cui, and R. Wang, "A novel unsupervised anomaly detection for gas turbine using Isolation Forest," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2019: IEEE, pp. 1-6.

[13] W. Yan and L. Yu, "On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach," *arXiv preprint arXiv:1908.09238,* 2019.

[14] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396-404.

[15] H. Soma and O. Sinan, "Hands-On Machine Learning for Cybersecurity. Birmingham–Mumbai: Packt Publishing, 2018," ed.

[16] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors,* vol. 19, no. 4, p. 972, 2019.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[19] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*, 2011: Springer, pp. 52-59.

[20] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Applied energy,* vol. 211, pp. 1123-1135, 2018.

[21] Y. Dovoedo and S. Chakraborti, "Boxplot-based outlier detection for the location-scale family," *Communications in statistics-simulation and computation,* vol. 44, no. 6, pp. 1492-1513, 2015.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[23] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, "Finding the best classification threshold in imbalanced classification," *Big Data Research,* vol. 5, pp. 2-8, 2016.