Point and contextual anomaly detection in building load profiles of a university campus

1st Long Wang
3rd Meng Zhang
4th Till Riedel
5th Michael Beigl
TECO, Karlsruhe Institute of Technology
Karlsruhe, Germany
{wanglong,mzhang,riedel,beigl}@teco.edu

2nd Marian Turowski
6th Ralf Mikut
7th Veit Hagenmeyer
Institute for Automation and Applied Informatics (IAI)
Karlsruhe Institute of Technology
Karlsruhe, Germany
{marian.turowski,ralf.mikut,veit.hagenmeyer}@kit.edu

***Abstract*—The increasing use of smart meters enables the monitoring and diagnostics of the underlying systems. The application of data analytics methods can help to automate monitoring and diagnostics such that human intervention is limited to the situations where and when it is necessary. In a smart grid, diagnostics can relate to faulty smart meters and unusual consumption, corresponding to point anomalies and contextual anomalies. A Deep Neural Network Regression, an Autoencoder with reconstruction, and the encoder of the Autoencoder are proposed for automated anomaly detection. The three models are evaluated on real-world building load profiles of a university campus containing such anomalies. The results demonstrate that the proposed models have superior detection accuracies over benchmarks and differ in the discrimination between anomalies and normal electrical load profiles. At the same time, the models correctly identify different anomalous electrical load profiles that were wrongly labeled as normal.**

***Index Terms*—Anomaly Detection, Building Load Profiles, Load Analysis, Deep Neural Network, Autoencoder**

I. Introduction

Energy use in buildings accounts for about one-third of the world's final energy consumption [1]. Emissions caused by lighting and HVAC (heating, ventilation and air-conditioning) systems in buildings play a key role in contributing to global warming and climate change. At the same time, faulty equipment in buildings, the energy wasting behavior of their users, and inappropriate control strategies constitute a substantial energy saving potential. For this reason, energy management in the managed facilities is increasingly important for the responsible facility management (FM) departments.

In the field of building energy management, the analysis of building operations, equipment status, and equipment failures is a common problem [2]. With the increasing deployment of smart meters and other sensors, FM departments focus more strongly on monitoring the managed facilities. Using data analytic methods, one of their goals is an automated monitoring system that prompts human intervention only where and when required. In this context, two typical anomalies, serving as use cases for such a monitoring in a smart grid, are the identification of faulty smart meters and unusual consumption.

Generally, any kind of forecasting model [3] could serve as a basis for detecting these anomalies by considering large deviations from the forecasting model as anomalies. This paper examines three specific solutions for detecting the two aforementioned anomalies that we manually label in our data set. A Deep Neural Network Regression (DNNR), an Autoencoder with reconstruction (AER), and the encoder of the Autoencoder (EAE) are applied to a real-world building load profiles containing such anomalies. The proposed models are compared with each other and two benchmarks, a naive model and a Support Vector Regression (SVR), regarding their detection accuracy in both use cases. The ultimate goal is to provide the FM department with an automated machine learning-based solution that facilitates the monitoring and diagnostics of the managed resources in a smart grid.

The remainder of this paper is organized as follows: Section II presents related work. Section III gives an overview of the used data and preprocessing. Section IV defines the anomalies considered in this work. Section V describes the applied anomaly detection models and Section VI presents their evaluation. Finally, Section VII gives a conclusion.

II. Related work

Despite various other promising approaches regarding anomaly detection such as noise clustering [4] and Super State Hidden Markov Models [5], this section reviews anomaly detection literature that

- focuses on smart meter failures or unusual energy consumption data of a building as well as
- employs a Support Vector Regression (SVR), a Deep Neural Network Regression (DNNR), or an Autoencoder (AE).

In [6], electrical energy consumption data are investigated with a SVR and a linear regression regarding both usual load modeling and anomaly detection. Similarly, a deep semi-supervised convolutional neural network is proposed and compared with a Fully Connected AE (FCAE), a Convolutional AE (CAE), and a Support Vector Machine (SVM) in [7]. In [8], recorded electrical power data serve as a basis for evaluating four approaches: the Classification And Regression Tree (CART) and k-means clustering, each together with the Generalized Extreme Studentized Deviate (GESD) algorithm, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and the Artificial Neural Networks and Basic Ensembling Method (ANN BEM).

Other works do not only compare but also rather combine different methods. One such ensemble approach is the combination of an AE with Random Forest and a SVR in [9]. In [10], an autoencoder-based ensemble method is developed taking into account different architectures and training schemes in order to identify anomalies in an unsupervised manner.

To the best of our knowledge, no previous studies have compared SVR, DNNR, and AE with regard to detecting smart meter failures and unusual energy consumption in the building energy domain.

III. Data overview and preprocessing

This section gives a short overview of the available data and the preprocessing carried out. Our case study uses the real-world data presented in [11]. The data set contains readings of smart meters that collect the electricity consumption every 15 minutes over a period of about 10 years (from January 1, 2006 to May 6, 2015).

Furthermore, we make use of environmental data collected at a 200-meter high meteorological tower on the campus. Sensors measure the mean and the maximum of various meteorological parameters at different altitudes on the tower every 10 minutes. For our case study, we only consider the parameters that have a close correlation with energy consumption, i.e. outside temperature, illumination, and humidity.

Within the electrical load data, some smart meters have unusually high quarterly consumption rates, probably due to a replacement or reset. We replace such values with the last-known values. Since we have two different temporal resolutions, we also convert the electrical energy consumption data from one sample every 15 minutes to one sample every 10 minutes by linear interpolation in order to align them with the environmental data. Finally, the time series of each smart meter are segmented into days (0:00 to 24:00) to obtain the electrical load profiles for our analysis.

IV. Anomaly definition

In this section, we present the definition of point anomalies and contextual anomalies that we identify in our data set.

A. Point anomaly

If an individual data instance can be considered as anomalous with respect to the rest of data, then the instance is termed a point anomaly [12]. Since we consider electrical load profiles of one day, a point anomaly refers to up to several anomalous hours within such a day.

In our data set, we find such anomalous patterns. The smart meters sometimes report several zero values followed by a large leap. This leap is roughly equal to the accumulated value of missing preceding readings. We can reasonably infer that the smart meters fail to report readings for a certain period of time and then report the accumulated high value after their recovery. We review the data of a building (i.e. Building A) for the second five years and manually label these anomalous patterns as point anomalies.

B. Contextual anomaly

If a data instance is anomalous in a specific context but not otherwise, then it is termed a contextual anomaly [12]. In contrast to the point anomaly, the detection of a contextual anomaly requires both the electrical load profile and—as additional information—calendar and public holiday data.

The data set contains various days with unexpected patterns such as "Proximity days". For some working days, the smart meters report an electrical energy consumption that is more like that of a non-workday, for instance April 30, 2012. As a theoretically normal workday, it is followed by the Labor Day as a public holiday on May 1. We assume that people commonly enjoy a four-day weekend and return to work on May 2, thus causing a lower electricity consumption on April 30. We label these proximity days as contextual anomalies in the data of Building A for the second five years.

V. Anomaly detection models

In this section, we present the applied anomaly detection models and their corresponding anomaly scores. Two types of models are involved, i.e. regression-based models and autoencoder-based models. A naive model and a support vector regression model are used as benchmarks.

A. Naive model

Since a clear weekly pattern can be observed in the building load profiles, we assume that it is mainly caused by the people studying and working on the campus and their tendency to repeat their behavior weekly. Therefore, we use the electrical load profile of the same building in the previous week as the prediction of our naive model. The corresponding anomaly score S is defined as the Mean Squared Error (MSE) between the predicted electrical load and the measured electrical load:

$$S_{regression} = \frac{1}{n} \sum_{k=1}^{n} (x_{measured}[k] - x_{prediction}[k])^2 \quad (1)$$

B. Regression-based models

Regression-based anomaly detection models predict the electrical load and compare it to the measured electrical load. We use the MSE of the predicted and the measured electrical load as shown in Equ. (1) as the anomaly score for the regression-based models. For these models, we also make use of historical electrical load data and environmental data as input features as shown in Table I. From the many methods proposed for load profile prediction, we adopt Support Vector Regression (SVR) [13] and Deep Neural Network Regression (DNNR) [14] as regression-based models:

1) SVR: Since the load profile is non-linear, we use a Radial Basis Function (RBF) kernel for the SVR model. We conduct a grid search and cross-validation to optimize the parameters C and Gamma.

TABLE I: Description of the features used for both regression-based models.

Feature	Description
Year	2006, 2007, ..., 2015
Month	1, 2, 3, ..., 12
Day	1, 2, 3, ..., 31
Weekday	1, 2, 3, ..., 7
Hour	1, 2, 3, ..., 24
Minute	0, 10, 20, 30, 40, 50
Public Holiday	0, 1
Humidity, Illumination, Outside temperature	0-1 (normalized)

TABLE II: Description of the features used for both autoencoder-based models.

Feature	Description
Year	2006, 2007, ..., 2015
Month	1, 2, 3, ..., 12
Day	1, 2, 3, ..., 31
Weekday	1, 2, 3, ..., 7
Public Holiday	0, 1
$E_a \{a = 1, ..., 144\}$	Load profile of one day
Daily average temperature	0-1 (normalized)
Daily average illumination	0-1 (normalized)
Daily average humidity	0-1 (normalized)
Daily temperature difference	0-1 (normalized)
Daily maximum illumination	0-1 (normalized)
Daily humidity difference	0-1 (normalized)

2) DNNR: Our DNNR has 10 neurons in the input layer and one neuron in the output layer that represents the electricity consumption within 10 minutes. The structure of DNNR model is a feed forward network (FFN) with four hidden layers. The number of neurons of each hidden layer is 8, 6, 4, 2. The Rectified Linear Unit (ReLU) is used as the activation function in each layer. The model is optimized by the Adam optimizer [15] that automatically adjusts the learning rate during the training of the model.

C. Autoencoder-based models

An Autoencoder is a type of feed-forward neural network that tries to minimize the difference between the output $\overline{\mathbf{x}}$ and the input $\mathbf{x}$ [16]. We apply both a complete Autoencoder as well as only its encoder for anomaly detection in this work. To distinguish them, we use the names Autoencoder with reconstruction (AER) and Encoder of the Autoencoder (EAE) respectively. The features used as input for both autoencoder-based models are shown in Table II. Each input vector has 155 dimensions, consisting of time information, the electrical load profile of one day, and environmental factors.

1) AER: Intuitively, the Autoencoder trained with normal data is likely to generate a larger $|\mathbf{x}-\overline{\mathbf{x}}|$ value when it is applied to anomalies due to the inherent difference between anomalies and normal data. The assumption is that the anomalies have more principle components than normal data. By tuning the parameters of the Autoencoder, we expect the model to output a small $|\mathbf{x}-\overline{\mathbf{x}}|$ value for normal data and to output a large $|\mathbf{x}-\overline{\mathbf{x}}|$ value for anomalies. Naturally, $|\mathbf{x}-\overline{\mathbf{x}}|$ is used as an anomaly score for the AER model:

$$S_{AER} = |\mathbf{x} - \overline{\mathbf{x}}| \tag{2}$$

To optimize this parameter, we start with 10 hidden neurons and gradually reduce the number until the reconstruction loss is large enough for anomalies and still low for normal data.

2) EAE: The EAE model is merely the encoder of the AER model. Generally, the encoder compresses the input data into a lower dimensional space. Fig. 1 illustrates a typical output of such an encoder. Clearly, there are two clusters. The green points are normal non-workdays, the blue points are normal workdays, and the red points are labeled contextual anomalies. We can observe that the anomalies lie in between these two clusters and typically in a sparse region, i.e. they do not have many close neighbors. We therefore define the anomaly score for the EAE model as follows:

$$S_{EAE} = D_c * D_k \tag{3}$$

D_c is the distance between the data point and the cluster center. D_k is the mean distance between the data point and its k nearest neighbors.

VI. Evaluation

In this section, we evaluate the three proposed anomaly detection models and the benchmark models on the electrical load profiles of building A.

Building A is an office building with about 9000 m^2 floor space. For Building A, electrical load profiles of 3414 days exist in total in the data set. We train the models on the data of the first five years and test the models on the data of the second five years. The second five years of data contain the days labeled as anomalies. Among the test data set, there are 1169 normal days, 19 instances of point anomalies, and 35 instances of contextual anomalies. The Area Under the Curve (AUC) is used as the metric for evaluate the models.

A. Point anomaly detection

Fig. 2 shows the ROC curves of the five models regarding the point anomaly detection. In this context, anomalies are referred to as being positive. The AER model performs the best with an AUC value of 0.95. It is followed by the DNNR model with an AUC value of 0.9. The EAE (AUC = 0.73) and SVR (AUC = 0.5) models perform worse than the naive model (AUC = 0.79), which means that they are not suitable for detecting point anomalies.

B. Contextual anomaly detection

Fig. 3 presents the ROC curves of the five models concerned with the contextual anomaly detection. The DNNR model performs the best among all five models with an AUC value that approximately equals 1. The EAE model also performs well with an AUC value of 0.99. The performance of the AER model decreases considerably when detecting contextual anomalies (AUC = 0.66) compared with detecting point anomalies (AUC = 0.95). The naive model also performs well when detecting both point anomalies (AUC = 0.79) and

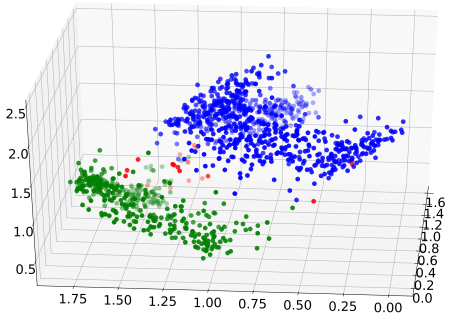


Fig. 1: Typical output of the EAE model with normal non-workdays (green points), normal workdays (blue points), labeled contextual anomalies (red points), and gradual colors for better readability.

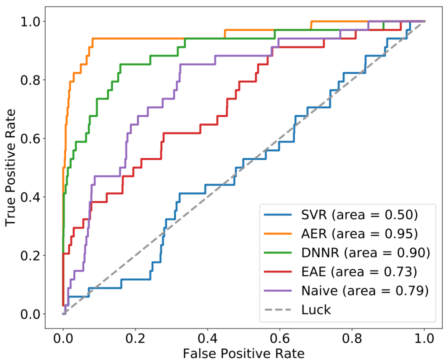


Fig. 2: Comparison of ROC curves across models regarding the detection of point anomalies.

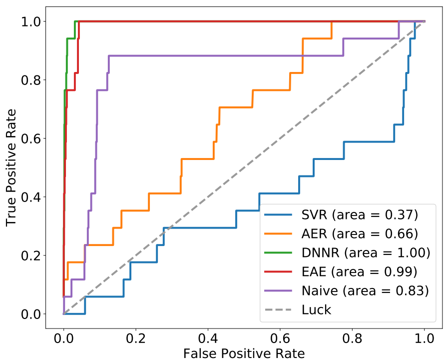
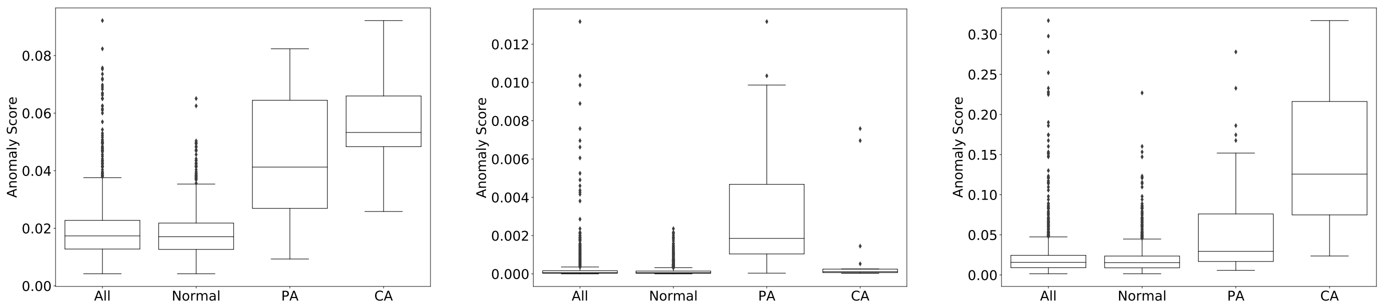


Fig. 3: Comparison of ROC curves across models regarding the detection of contextual anomalies.



(a) DNNR (b) AER (c) EAE

Fig. 4: Comparison of anomaly score distributions across the DNNR, AER, and EAE models, where Normal covers all load profiles labeled as normal, PA are point anomalies, and CA are contextual anomalies.

contextual anomalies (AUC = 0.83). The SVR model once again has a worse performance (AUC = 0.37) than the naive model. The reasons for the limited performance of the SVR model when detecting both point anomalies and contextual anomalies are not clear and will be investigated in future research.

C. *Models' discrimination quality*

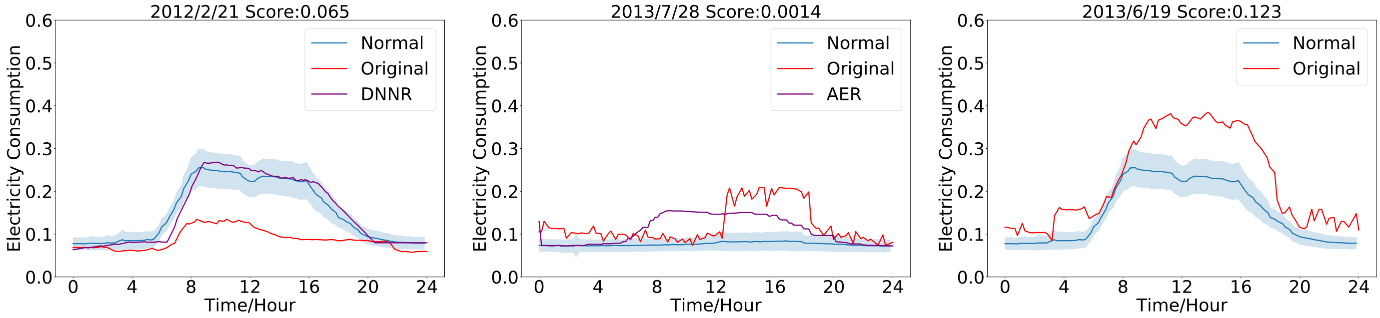
In order to evaluate the proposed models' discrimination quality, we visualize the anomaly score distributions for all models in Fig. 4. This shows the anomaly score of all load profiles, the load profiles labeled as normal, the labeled point anomalies, and the labeled contextual anomalies for each model. The anomaly score distributions further confirm the previous observation. As shown in Fig. 4b, the AER model distinguishes point anomalies from both normal data and contextual anomalies well. Meanwhile, the EAE model performs well at discriminating contextual anomalies from normal data and point anomalies as illustrated in Fig. 4c. Last but not least, the DNNR model separates both types of anomalies from normal data well but cannot distinguish both types of anomalies from each other.

D. *Case study: False positives*

As Fig. 4 indicates, some electrical load profiles labeled as normal have anomaly scores so high that they are classified as anomalies. These load profiles are referred to as False Positives (FP). Since it is desirable to avoid such wrongly labeled load profiles, we have a closer look at them in order to gain a deeper understanding of the DNNR, AER, and EAE models. For each model, we discuss one typical FP.

As shown in Fig. 5a, the measured load profile on February 21, 2012 is considerably lower than the normal electrical load profile and the electrical load predicted by the DNNR model. After checking the calendar, we find these days are Carnival Festival, which is not officially a public holiday but often taken as one by many people. Consequently, the visual inspection of the FP leads to the conclusion that the DNNR model correctly detects an anomalous electrical load profile that was wrongly labeled as normal.

This finding also applies to the FPs of the AER and the EAE models shown in Fig. 5b and Fig. 5c. However, it is worth noting that these two models identify different FPs. Naturally, the idea of combining both models arises so that the resulting combination covers different types of anomalies. The case study shows that the proposed models have the potential



(a) FP-DNNR (b) FP-AER (c) FP-EAE

Fig. 5: Case study of False Positives (FPs) of the DNNR, AER and EAE models with the mean value of the load profiles labeled as normal (blue) with a one sigma band, the measured load (red), and, if applicable, the predicted load (purple).

capability to detect new types of anomalies.

VII. Conclusion

The present paper addresses a typical challenge in the smart grid. It considers monitoring and diagnostics by focusing on smart meter failures and unusual electricity consumption in real-world data. These two anomalies occur in building load profiles of a university campus and are detected by three different machine learning models, namely a DNNR, a AER and a EAE model. We compare these models with each other and two benchmarks, a naive and a SVR model, regarding their detection accuracy for both types of anomalies.

The results show that the AER and the DNNR models well classify point anomalies. For contextual anomalies, the DNNR and the EAE models show the best results. Overall, the AER, DNNR, and EAE models provide better results than the SVR and the naive models. These three models, however, distinguish between normal data, point anomalies, and contextual anomalies differently. At the same time, the DNNR, AER, and EAE models correctly identify different kinds of anomalous electrical load profiles that were wrongly labeled as normal ones during manual labeling.

As a consequence, future work will investigate a combination of these three best performing models and evaluate an automated machine learning-based solution that facilitates the monitoring and diagnostics of the managed resources for the university's FM department. Furthermore, we expect to transfer the pre-trained anomaly detection models to similar buildings and to prove a good scalability of this approach.

Acknowledgment

We thank the Facility Management and the Institute of Meteorology and Climate Research (IMK) at KIT for providing the data and Katherine Quinlan-Flatter for proofreading this article. The present contribution is supported by the Helmholtz Association under the Joint Initiative "Energy System 2050 A Contribution of the Research Field Energy".

References

[1] D. Ürge-Vorsatz, L. F. Cabeza, S. Serrano, C. Barreneche, and K. Petrichenko, "Heating and cooling energy trends and drivers in buildings," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 85–98, 2015.

[2] M. Molina-Solana, M. Ros, M. D. Ruiz, J. Gómez-Romero, and M. J. Martin-Bautista, "Data science for building energy management: A review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 598–609, 2017.

[3] J. Á. González Ordiano, S. Waczowicz, V. Hagenmeyer, and R. Mikut, "Energy forecasting tools and services," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 2, p. e1235, 2018.

[4] S. Waczowicz, M. Reischl, V. Hagenmeyer, R. Mikut, S. Klaiber, P. Bretschneider, I. Konotop, and D. Westermann, "Demand response clustering - How do dynamic prices affect household electricity consumption?" in *Proc. IEEE PowerTech 2015*, 2015.

[5] R. Gabriel, J. Matthes, H. B. Keller, and V. Hagenmeyer, "Detection and Localization of Manipulated Smart Meters Using Super State Hidden Markov Models," in *In Proc. 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2019.

[6] C. Nordahl, M. Persson, and H. Grahn, "Detection of residents' abnormal behaviour by analysing energy consumption of individual households," in *Proc. 2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 729–738.

[7] N. L. Tasfi, W. A. Higashino, K. Grolinger, and M. A. M. Capretz, "Deep neural networks with confidence sampling for electrical anomaly detection," in *Proc. 2017 IEEE International Conference on Internet of Things (iThings), IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2018, pp. 1038–1045.

[8] A. Capozzoli, F. Lauro, and I. Khan, "Fault detection analysis using data mining techniques for a cluster of smart office buildings," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4324–4338, 2015.

[9] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. Capretz, and G. Bitsuamlak, "An ensemble learning framework for anomaly detection in building energy consumption," *Energy and Buildings*, vol. 144, pp. 191–206, 2017.

[10] 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.

[11] L. Wang, Y. Ding, T. Riedel, A. Miclaus, and M. Beigl, "Data analysis on building load profiles: A stepping stone to future campus," in *Proc. 2017 International Smart Cities Conference (ISC2 2017)*, 2017.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 15:1–15:58, 2009.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[14] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks," *Energy Conversion and Management*, vol. 50, no. 1, pp. 90–96, 2009.

[15] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[16] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 665–674.