2022 8th IEEE International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService)

Comparing human and algorithmic anomaly detection for HVAC systems applications

1st Reuben Borrison
Corporate Research
ABB AG
Ladenburg, Germany
reuben.borrison@de.abb.com

2nd Marc Syndicus
Energy Efficiency and Sustainable Building
RWTH Aachen University
Aachen, Germany
syndicus@e3d.rwth-aachen.de

3rd André Orth
Institute for Automation and Applied Informatics
KIT Karlsruhe Institute of Technology
Karlsruhe, Germany
andre.orth@kit.edu

4th Romana Markovic
Building Science Group fbta
KIT Karlsruhe Institute of Technology
Karlsruhe, Germany
romana.markovic@kit.edu

5th Marcel Dix
Corporate Research
ABB AG
Ladenburg, Germany
marcel.dix@de.abb.com

6th Antonio Liguori
Energy Efficiency and Sustainable Building
RWTH Aachen University
Aachen, Germany
liguori@e3d.rwth-aachen.de

7th Matthias Berning
Corporate Research
ABB AG
Ladenburg, Germany
matthias.berning@de.abb.com

8th Andreas Wagner
Bauphysik und Technischer Ausbau
KIT Karlsruhe Institute of Technology
Karlsruhe, Germany
andreas.wagner2@kit.edu

9th Christoph van Treeck
Energy Efficiency and Sustainable Building
RWTH Aachen University
Aachen, Germany
treeck@e3d.rwth-aachen.de

2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService) | 978-1-6654-5890-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigDataService55688.2022.00032

***Abstract*—This paper reports the first results of a comparison of human and algorithmic anomaly detection. We are interested in how human and automated anomaly detection can be combined in the most beneficial way to improve how fault detection is practiced in building maintenance. Open source datasets with sensor data were annotated by persons with low subject matter experience, and compared with a Convolutional Autoencoder Neural Network (CAE) as well as a dedicated time series algorithm (DeepAnT), and detection metrics of human and algorithmic procedures are compared. Future comparisons will include higher levels of expertise on the human side, and more sophistication/training amount on the algorithm side. We close by discussing the advantages and caveats of our approach.**

***Index Terms*—Anomaly detection, Annotation, HVAC, Predictive maintenance, Predictive fault detection and diagnosis**

I. Introduction

Commercial buildings are equipped with a large number of sensors that provide diverse information required for building control. Depending on sensors and measurement intervals, this can amount to several thousand data points per building and ultimately requires Big Data techniques and data treatment approaches. Commonly, some hardware or part of building control fails during the building life cycle, and there is a significant delay between the defect and the identification of the malfunction. As a result, the faults and defects could lead to increased energy consumption, low satisfaction of occupants with the indoor environmental quality, or even safety concerns. The timely detection of faults and defects is challenging due to several factors. First, the large number of sensors and the diverse hardware components are time consuming to inspect manually. Second, the detection of these faults with monitoring data is a challenge, since it requires the real-time analysis of the latest monitored values. This task could be performed either by building managers familiar with the subject matter or by experienced engineers with expertise in building control and monitoring. As an alternative to manual inspection, there has been significant progress in automatic fault detection and diagnosis in buildings [1]–[5].

To develop a reliable machine learning model for detecting equipment faults, it is important to have labeled data that also contains known faults from the past. Without this ground truth, it is not possible to determine how well the model performs, since performance measures, such as accuracy and precision, rely on this information.

However, data obtained from technical equipment is often provided unlabeled to the data scientist [6], so annotations are necessary to make it useful for model training and validation. This could be done, e.g., by human experts to inspect the given data for anomalous patterns, or by an automatic approach. Yet, it is not possible to decide whether a human expert approach works better than an automatic approach. Take the following example in Figure 1, which shows the multivariate time series for an electric motor. An automatic approach may falsely assume that there is an anomaly at time 1000 (and 2500), because there is a sharp drop in the signals, but a human expert would see that this is simply a motor stop.

Consider another example in Figure 2, which shows the multivariate time series for an air handling unit (AHU) of heating, ventilation, and air conditioning system (HVAC). The

DOI 10.1109/BigDataService55688.2022.00032

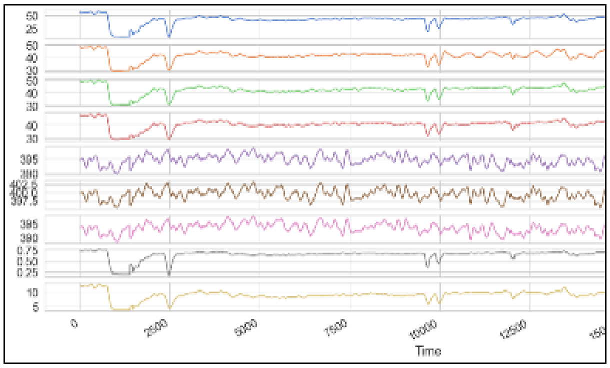

Fig. 1. Example data for an electrical motor. The time series represent normal motor operation without a fault.

left plot represents an unfaulted situation, but the right plot contains a technical fault. Obviously, it is not possible for a human expert to tell that there is a fault just by looking at these plots, but an automatic algorithm would be able to detect the situations as unfaulted and faulted [7].

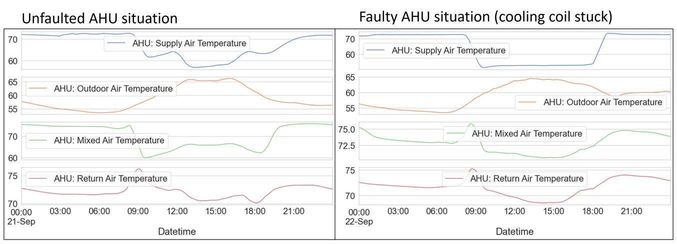

Fig. 2. Example data for an AHU. Left: the time series that represent a normal operation, right: the time series that contains a fault. Source: [8]

Hence, a combination of both approaches may be needed: An automatic approach that can efficiently find meaningful patterns in big data that would be hard to find manually, combined with human expertise that can extract the implicit knowledge that can not be found mathematically.

In this paper, we report on the first comparisons of human annotators (low levels of expertise) with algorithms that did not receive extensive training. Specifically, we used an autoencoder neural network as suggested by [9] and unsupervised anomaly detection (DeepAnT) as described by [10]. The reason for using these approaches over other available approaches is discussed in section II. We try to identify strategies, biases, and performance differences between these two approaches with the goal of an improved, cooperative anomaly detection procedure combining the respective strengths of humans and algorithms [11].

II. State of the Art

ML-based anomaly detection, in general, is a widely studied topic in academia and industry [12], [13]. Even in the area of buildings, it is a widely studied topic for different assets/topics such as HVAC [2], energy consumption [4] etc. [14] describes the problem of unavailability of labeled data in industrial datasets and mentions its significance for supervised machine learning algorithms. In this section, we provide an overview of the popular approaches for labeling anomalous data. First, we highlight different techniques for labeling/annotating the anomalous data in different domains. Then, we look at algorithmic based labeling techniques for unlabeled anomalous time series data in buildings. Lastly, we summarize human based labeling for time series data in buildings, in general and also for HVAC data.

[14] mentions crowd-sourcing and interactive learning (active learning and semi-supervised learning) as popular approaches for labeling of industrial data but also questions the accuracy and reliability of these approaches. As the labeling can be a laborious job, [15] did a systematic review on different interactive labeling approaches for different datasets. The approaches were classified as oracle, where users are asked about the correctness of a label, and teacher, where users can provide detailed explanations in interactive labeling processes. The results of the systematic survey showed that the majority of literature looked at approaches for interactive labeling of images and text data while very few papers looked at interactive labeling for tabular, 3D images, video, audio etc. Active learning and reinforcement learning was often cited in the papers as the labeling approach. Transfer learning approaches have also been in use for autolabeling of text, audio and image data [16]. [17] uses semi-supervised approaches for auto labeling time series data. The approach uses a small set of examples to label a large amount of datasets. Crowd-sourcing seems to contribute to the success of artificial intelligence, and it has been used to gather labels and solve different problems focusing on text, images, and audio data [18], [19].

On the algorithmic side, there has been a lot of research in finding anomalous labels in the data. Statistical methods are commonly applied for finding anomalies, such as k-NN [20] and different improved LOF [21] techniques such as connectivity based outlier factor (COF) [22], influenced outlierness (INFLO) [23] or cluster based local outlier factor (CBLOF) [24]. Other unsupervised approaches include OCSVM [25], PCA [26] etc. These techniques have been applied for different domains and datasets. Furthermore, there are ANN based anomaly detection approaches for time series data as well, such as stacked LSTMs [27] and Autoencoder based anomaly detection [28] [29]. Several unsupervised anomaly detection approaches have been used and open sourced by major corporations such as EGADS from yahoo [30], Twitter anomaly detection [31] and skyline from etsy [32]. These algorithms, together with some others [33], [34], were used by [10] to compare their DeepAnT algorithm for labeling anomalous data. DeepAnT seems to outperform all these methods and thus has been also used in this study to compare the human annotations.

Anomaly detection in buildings and for HVACs has also been investigated. [2] mention many supervised anomaly detection approaches for different parts of HVAC but very few unsupervised approaches. A study of unsupervised anomaly detection applied to HVAC [35] compared isolation forest, OCSVM and LSTM autoencoders. Results indicated that

LSTM autoencoders outperformed the other two algorithms, but it seems that LSTM autoencoders were only trained on the normal dataset which meant that there was a requirement of labels in the dataset to train the algorithm. There are other attempts to perform unsupervised anomaly detection such as [36], but labeling with the help of human intelligence seems to be less explored in anomaly detection for buildings.

Crowd-sourcing and human-in-the-loop labeling have been useful for image and natural language processing [37], but the applicability to building data time series is unclear. General domain information is clearly available in case of images and text, but more difficult to obtain from time series data of buildings. Yet, this information may be accessible to people familiar with the topic such as engineering students or facility managers. Utilizing the information from students and algorithms in the first step could reduce the burden of acquiring ratings from seasoned domain experts and allow a comparison of human and algorithmic annotation processes.

III. Method

A. Process description

1) Human labeling: Within the human-algorithm comparison, different level of expertise on the human side will be compared to varying levels of training and/or sophistication of the algorithm. In the first step, we intend to compare naive algorithms with low level of human expertise regarding anomalies. Therefore, the annotation task was aiming at only identifying anomalies, without making assumptions about the reasons for their occurrence. For the annotation of time series data, we used the online tool TRAINSET [38], see Figure 3.

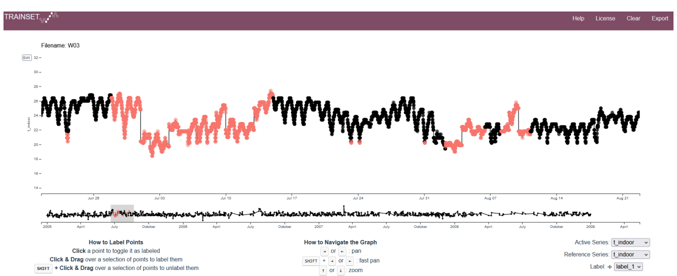

Fig. 3. Screenshot of TRAINSET web-based annotation tool [38]

Students from energy engineering and civil engineering annotated with three labels: The first label whenever the annotator suspected data points to be anomalous, the second one whenever an anomaly was found in comparison to another signal, and the third label whenever the annotator was unsure. All labels could be denoted point- or section-wise.

They received an information sheet with a brief description of the annotation task as well as the variables included in the dataset, alongside a video-based tutorial on the use of the TRAINSET software [38]. Basic knowledge about HVAC systems and indoor environmental quality (e.g., relative humidity) was the only requirement.

2) Expert review: Experts are reviewing the already labeled data, due to the fact that they have less time (experts are not selected from the author team). The data labeled by student annotators is prepared in sequences. Experts are provided with the user interface shown in Figure 4 for their review. If applicable, comparison features are shown next to the main feature as well as the type of anomaly that was chosen by the student annotators during the label process. Based on this information, the experts can select if the highlighted sequence of the main feature was correctly assumed as an anomaly.

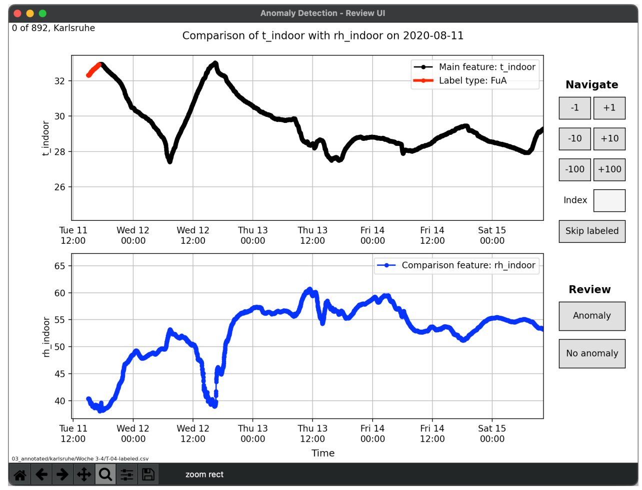

Fig. 4. Review UI for expert reviews of student labeled data

The workflow of the human anomaly detection is shown in Figure 5. The expert review part is not analyzed in this paper, as securing suitable experts presents a difficult task. Therefore it is planned to be approached in a future publication.

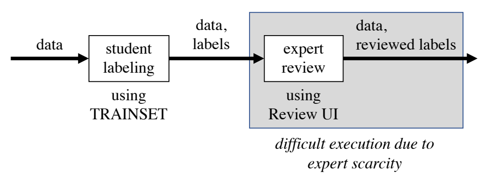

Fig. 5. Human anomaly detection process

3) CAE anomaly detection: An existing Convolutional Autoencoder Neural Network for reconstructing missing indoor environment data time series [39] was used. Following the methodology proposed in [40], each variable was re-organized in subsequence matrices. Each row had the size of the most dominant period so that an incremental step of two hours was applied to generate overlapping subsequences. To identify the most dominant periods, the auto-regressive spectral density [41] was estimated for each data stream. The final matrices were then randomly shuffled by row and equally distributed for model's training, validation, and evaluation. An anomaly score ranging between 0 and 1 was computed on the evaluation set of each matrix, based on the following formula [40]:

$$Score_i = \frac{RMSE_i - RMSE_{min}}{RMSE_{max} - RMSE_{min}} \;, \tag{1}$$

where $\mathrm{Score_i}$ and $\mathrm{RMSE_i}$ are the scores and root mean squared errors (RMSE) computed for the $i-th$ subsequence of a certain matrix, while $\mathrm{RMSE_{max}}$ and $\mathrm{RMSE_{min}}$ are maximum and minimum RMSEs of the same matrix [39]. The final scores for the ensemble evaluation set were obtained based on the mean operation. Eventually, anomalies were identified based on a confidence interval of 0.95.

4) DeepAnT anomaly detection: DeepAnT [10] is an unsupervised deep learning neural network approach to identify point and contextual anomalies in time series. It requires less data and can be trained on the dataset containing anomalies, making it an ideal choice for annotating the unlabeled dataset. The algorithm is a two step approach, first step consists of a CNN based *time series predictor* which trains on the time series windows to forecast in the future. The second step involves an *anomaly detector* which uses the euclidean distance between the predicted value and the actual value to determine the anomaly score ranging from 0 to 1. A threshold then can be set to determine the anomalous and normal data. The threshold for identifying anomalies in the data is based on a quantile ranging from 0.80 to 0.95.

B. Dataset description

We use two AHU datasets by the Lawrence Berkeley National Laboratory [8]. Both consist of sensor and control data from a physical single zone AHU, a schematic is shown in Figure 6. An AHU takes in outside air, filters it, cools/heats it and supplies it to the rooms, while a heat exchanger recovers heat before exhaust air is released. Accordingly, the data set comprises relevant temperature sensor data such as outside air temperature (OAT), supply air temperature (SAT), return air temperature (RAT) and mixed air temperature (MAT).

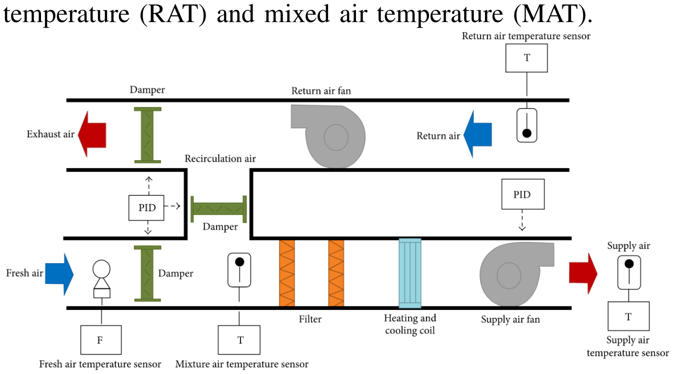

Fig. 6. Schematic of AHU [42]

Several different types of faults were injected by the LBNL team in the AHU, such as leakages or a stuck damper. Each fault lasted for a whole day and was removed again at the end of the day. These faults were performed on two different AHU: The first dataset is referred as SZVAV (11 days of data, thereof 7 fault-days), the second dataset is referred as SZCAV (15 days, all of them with faults).

IV. Results

A. CAE anomaly detection

The auto-regressive spectral density described in Section III-A was estimated for each variable of the Berkeley dataset. Ten out of 14 variables had clear daily seasonality. Since the unsupervised model was originally optimized on daily profile patterns, the anomaly detection analysis was performed only on the aforementioned data streams. Due to the limited size of the dataset, evaluation was run for all the labeled data. This might have underestimated the number of anomaly points, since the training set was also included in the evaluation.

As CAE anomaly detection performance varies with different thresholds, reconstruction errors were tested for different thresholds using ROC as shown in Figure 7 which shows the curve around the random classifier dotted line which means that the model was only as good as guessing the anomalies.

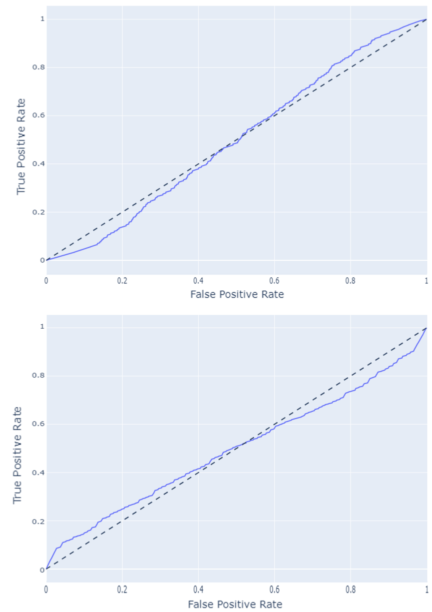

Fig. 7. ROC curves for SZCAV (top) and SZVAV (bottom) datasets used.

B. DeepAnT anomaly detection

The most important data came from temperature sensors from both the datasets which were used to determine labels for the dataset. Due to the dimension of the data obtained after feature selection the CNN architecture proposed in [10] was modified to fit the data. The algorithm was tested with a different number of window sizes and thresholds to get the Response Operator Characteristc (ROC) curve as shown in Figure 8. The Area Under Curve (AUC) for different window sizes is comparable and it suggests that there would be some differences in accuracy on varying this parameter.

C. Human labeling

Students used three categories to annotate, unlabeled data points were considered as normal. In two separate analyses, the “not sure” category was included either as anomaly (eval 1) or normal (eval 2). MCC and Cohen’s Kappa were used as comparison metrics for imbalanced classification [43]. AUC was mentioned as the most robust metric by [43] followed by MCC and Kappa. But as AUC requires calculating true positive and false positive rates, this was impossible for the labels obtained by humans. The annotation comparison for both the datasets employed MCC and Kappa scores obtained from each evaluation (Figure 9). In both cases DeepAnT performs much better, while students and autoencoders perform similar.

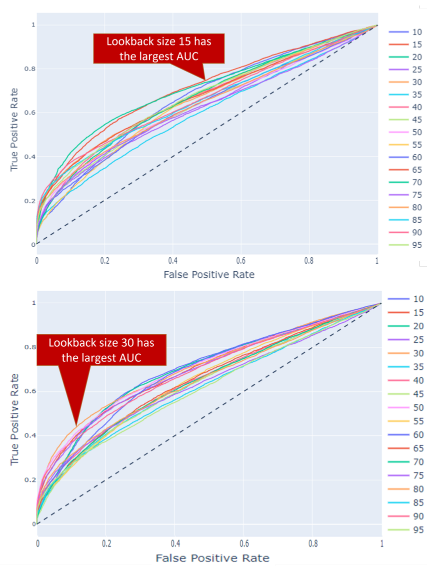

Fig. 8. Multiple ROC curves for different window sizes (look back sizes) for SZVAV (Top) and SZCAV (bottom). The look back size of 15 and 30 has the largest AUC respectively.

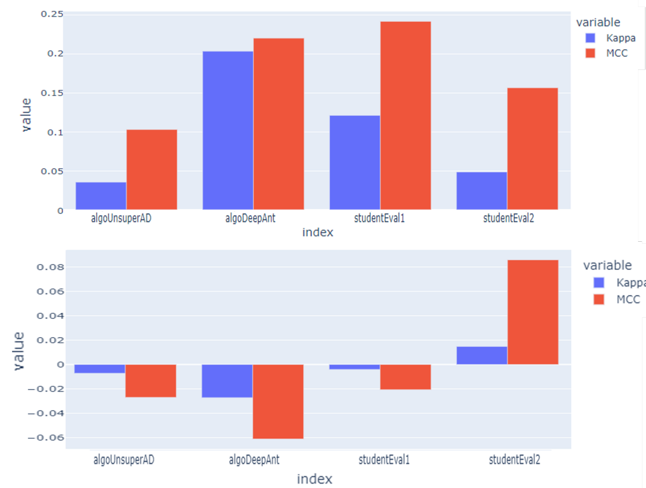

Fig. 9. MCC and Kappa scores for SZVAV (top) and SZCAV (bottom) dataset.

V. Conclusion and Future Work

1) Missing ground truth: Due to a scarcity of fully annotated test datasets, missing ground truth will remain a commonly encountered problem. The HVAC dataset used in our comparison comprises ground truth information, but we expect this to be not the case in many industrial applications. Therefore, one approach is probability- or agreement-based: Human as well as algorithmic detection capabilities may not substitute ground truth information, but can be used to denote data points in time series that resemble a prototypical anomaly. In case of the human annotator based on experience, and in case of the algorithm based on past data and learned patterns.

2) Algorithmic data screening: Future systems might work best when pre-screening and raw data analysis is done by algorithms, and critical data patterns are flagged and handed over to human experts for further inspection. This holds true for typical maintenance situations, where a computer-based fault indication needs to be verified (or falsified) by personnel inspecting and repairing a component if needed.

3) Limitations: A limitation of this study was the focus on deep learning. Our findings from recent studies [9] indicated that artificial neural networks can detect anomalies in multivariate time series data, while a follow-up study we have carried out indicated that statistical methods, including PCA, OneClassSVM, and Isolation Forest, did not provide good results. The main reason could be non-linear correlations between the multivariate time series that contain the anomalies to be found. However, there is support for time series data analysis in scikit-learn [44], [45], and a possible further research could be to evaluate its use and performance for anomaly detection. Another important step is to gather expert annotations in order to contrast these with student annotations and compare them with the algorithm results.

So far, we only compared low levels of expertise (students) with naive algorithms. Comparisons on the intermediate and expert level (of both human annotators and algorithms) are mandatory and planned next steps, as well as improvements to the human subject study design. We acknowledge that the human and algorithmic systems have commonalities, but remain distinct entities. The cognitive revolution provided new impulses for investigating human thinking with computer analogies and vice versa [46], but differences regarding the degree of knowledge about how the two systems operate still remain. Research on expert systems and anomaly detection give rise to the assumption that the combination of humans and algorithms is the most promising approach [11]. This can be used to incorporate an upstream decision-making for building control to only consult experts or facility managers if the found or still emerging anomaly is of a higher predefined threat level.

In sum, the current study reports the first level of an approach to compare and fruitfully combine human-algorithmic anomaly detection in HVAC time series, explicitly addressing the challenges of time series data [47] and following a stepwise approach which contrasts different levels of annotator knowledge and expertise with different sophistication and training amount of algorithms.

Acknowledgment

Part of this work was funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) as per resolution of the German Parliament under the funding code 03EN1002 [48]. We gratefully acknowledge this support. The authors are responsible for the contents of this contribution.

References

[1] Y. Zhao, T. Li, X. Zhang, and C. Zhang, "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future," *Renewable and Sustainable Energy Reviews*, vol. 109, pp. 85–101, 2019.

[2] M. S. Mirnaghi and F. Haghighat, "Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review," *Energy and Buildings*, vol. 229, p. 110492, 2020.

[3] A. Rogers, F. Guo, and B. Rasmussen, "A review of fault detection and diagnosis methods for residential air conditioning systems," *Building and Environment*, vol. 161, p. 106236, 2019.

[4] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," *Applied Energy*, vol. 287, p. 116601, 2021.

[5] G. Li, Y. Hu, J. Liu, X. Fang, and J. Kang, "Review on fault detection and diagnosis feature engineering in building heating, ventilation, air conditioning and refrigeration systems," *IEEE Access*, vol. 9, pp. 2153–2187, 2021.

[6] M. Dix, B. Kloepper, J. C. Blanchon, and E. Thorud, "A formula for accelerating autonomous anomaly detection," *ABB Review 02/21*, 2021.

[7] M. Dix and R. Borrison, "Open set anomaly classification," in *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2021, pp. 361–364.

[8] J. G. et al., "Building fault detection data to aid diagnostic algorithm creation and performance testing," *Scientific Data*, vol. 7, no. 1, pp. 1–14, 2020.

[9] M. Dix, A. Chouhan, S. Ganguly, S. Pradhan, D. Saraswat, S. Agrawal, and A. Prabhune, "Anomaly detection in the time-series data of industrial plants using neural network architectures," in *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, 2021, pp. 222–228.

[10] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, 2018.

[11] E. Calikus, "Together we learn more: Algorithms and applications for user-centric anomaly detection," dissertation, Halmstad University, 2022, http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-46404.

[12] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Computers in Industry*, vol. 131, p. 103498, 2021.

[13] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 15:1–15:58, 2009. [Online]. Available: https://doi.org/10.1145/1541880.1541882

[14] T. Fredriksson, D. I. Mattos, J. Bosch, and H. H. Olsson, "Data labeling: An empirical investigation into industrial challenges and mitigation strategies," in *Product-Focused Software Process Improvement*, M. Morisio, M. Torchiano, and A. Jedlitschka, Eds. Cham: Springer International Publishing, 2020, pp. 202–216.

[15] M. Knaeble, M. Nadj, and A. Maedche, "Oracle or teacher? A systematic overview of research on interactive labeling for machine learning," in *15. Internationale Tagung Wirtschaftsinformatik (WI 2020), Potsdam, March 2020*. GITO Verlag, 2020, p. 2–16.

[16] S. Zhang, O. Jafari, and P. Nagarkar, "A survey on machine learning techniques for auto labeling of video, audio, and text data," 2021. [Online]. Available: https://arxiv.org/abs/2109.03784

[17] M. González, C. Bergmeir, I. Triguero, Y. Rodríguez, and J. M. Benítez, "Self-labeling techniques for semi-supervised time series classification: an empirical study," *Knowledge and Information Systems*, vol. 55, no. 2, pp. 493–528, 2018.

[18] M. Bernstein, G. Little, R. Miller, B. Hartmann, M. Ackerman, D. Karger, D. Crowell, and K. Panovich, "Soylent: A word processor with a crowd inside," in *UIST 2010 - 23rd ACM Symposium on User Interface Software and Technology*, vol. 58, 2010, pp. 313–322.

[19] C. Grady and M. Lease, "Crowdsourcing document relevance assessment with mechanical turk," in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, 2010, pp. 172–179. [Online]. Available: https://aclanthology.org/W10-0727

[20] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," *ACM SIGMOD Record*, vol. 29, no. 2, p. 427–438, 2000.

[21] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '00. New York, NY, USA: Association for Computing Machinery, 2000, p. 93–104. [Online]. Available: https://doi.org/10.1145/342009.335388

[22] J. Tang, Z. Chen, A. W.-C. Fu, and D. W. Cheung, "Enhancing effectiveness of outlier detections for low density patterns," in *Advances in Knowledge Discovery and Data Mining*, M.-S. Chen, P. S. Yu, and B. Liu, Eds. Heidelberg: Springer, 2002, pp. 535–548.

[23] W. Jin, A. K. H. Tung, J. Han, and W. Wang, "Ranking outliers using symmetric neighborhood relationship," in *Advances in Knowledge Discovery and Data Mining*, W.-K. Ng, M. Kitsuregawa, J. Li, and K. Chang, Eds. Heidelberg: Springer, 2006, pp. 577–593.

[24] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," *Pattern Recogn. Lett.*, vol. 24, no. 9–10, p. 1641–1650, jun 2003. [Online]. Available: https://doi.org/10.1016/S0167-8655(03)00003-5

[25] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," in *Proceedings of the ACM SIGKDD Workshop*. New York, NY, USA: Association for Computing Machinery, 2013, p. 8–15.

[26] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," in *Proceedings of International Conference on Data Mining*, 01 2003.

[27] P. Malhotra, L. Vig, G. M. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *ESANN*, 2015.

[28] T. Amarbayasgalan, B. Jargalsaikhan, and K. H. Ryu, "Unsupervised novelty detection using deep autoencoders with density based clustering," *Applied Sciences*, vol. 8, no. 9, 2018. [Online]. Available: https://www.mdpi.com/2076-3417/8/9/1468

[29] M. Schreyer, T. Sattarov, D. Borth, A. Dengel, and B. Reimer, "Detection of anomalies in large scale accounting data using deep autoencoder networks," 2017. [Online]. Available: https://arxiv.org/abs/1709.05254

[30] N. Laptev, S. Amizadeh, and I. Flint, "Generic and scalable framework for automated time-series anomaly detection," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1939–1947.

[31] Phyks, "Introducing practical and robust anomaly detection in a time series — twitter blogs," 2015.

[32] Etsy, Inc., "Skyline real time anomaly detection system," original-date: 2013-06-10. [Online]. Available: https://github.com/etsy/skyline

[33] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 1, mar 2012. [Online]. Available: https://doi.org/10.1145/2133360.2133363

[34] M. Smirnov, "Contextual anomaly detector," original-date: 2016-06-24T15:07:07Z. [Online]. Available: https://github.com/smirmik/CAD

[35] A. Alghanmi, A. Yunusa-Kaltungo, and R. Edwards, "A comparative study of faults detection techniques on hvac systems," in *2021 IEEE PES/IAS PowerAfrica*, 2021, pp. 1–5.

[36] M. Munir, S. Erkel, A. Dengel, and S. Ahmed, "Pattern-based contextual anomaly detection in hvac systems," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 1066–1073.

[37] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," 2021. [Online]. Available: https://arxiv.org/abs/2108.00941

[38] Geocene, Inc., "Trainset annotation tool," https://trainset.geocene.com/, last access: 2022-03-20.

[39] A. Liguori, R. Markovic, T. T. H. Dam, J. Frisch, C. van Treeck, and F. Causone, "Indoor environment data time-series reconstruction using autoencoder neural networks," *Building and Environment*, vol. 191, p. 107623, 2021.

[40] 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.

[41] A. V. Oppenheim, *Discrete-time signal processing*. Chennai, IN: Pearson Education, 1999.

[42] H. Wang, "Application of residual-based ewma control charts for detecting faults in variable-air-volume air handling unit system," *Journal of Control Science and Engineering*, vol. 2016, 2016.

[43] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo, "Cross-validation metrics for evaluating classification performance on imbalanced data," in *2019 Intl. Conference on Computer, Control, Informatics and its Applications IC3INA*, 2019, pp. 14–18.

[44] C. W. Tan, A. Dempster, C. Bergmeir, and G. I. Webb, "Multirocket: Multiple pooling operators and transformations for fast and effective time series classification," *arXiv preprint arXiv:2102.00457*, 2021.

[45] "sktime (framework for machine learning with time series)," https://www.sktime.org/en/stable/tutorials.html, accessed: 2022-04-04.

[46] G. A. Miller, "The cognitive revolution: a historical perspective," *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 141–144, 2003.

[47] P. Dendorfer, A. Osep, A. Milan, K. Schindler, D. Cremers, I. D. Reid, S. Roth, and L. Leal-Taixé, "Motchallenge: A benchmark for single-camera multiple target tracking," *International Journal of Computer Vision*, vol. 129, pp. 845–881, 2021.

[48] "Project description DataFEE," www.e3d.rwth-aachen.de/cms/E3D/Forschung/Projekte/~enekp/Datefee/?lidx=1, accessed: 2022-04-04.