Anomaly Classification of Tennessee-Eastman Process Data

Marius Johannes Peter[1], Naghmeh Ghanooni[1], Fabian Hartung[1,2], Billy Joe Franks[1], Dennis Wagner[1], Philipp Liznerski[1], Aparna Muraleedharan[8], Justus Arweiler[1], Daniel Reinhardt[1], Indra Jungjoha, Steffen Reithermann[1], Sophie Fellenz[1], Fabian Jirasek[1], Maja Rudolph[3], Daniel Neider[4], Frank Rhein[9], Heike Leitte[1], Chen Song[5], Stephan Mandt[6], Michael Bortz[7], Jakob Burger[8], Hans Hasse[1], and Marius Kloft[1]

[1] RPTU, Kaiserslautern, Germany
[2] BASF SE, Germany
[3] Bosch AI, USA
[4] Technische Universität Dortmund, Germany
[5] ABB Corporate Research Center Ladenburg, Germany
[6] University of California Irvine, USA
[7] Fraunhofer ITWM, Germany
[8] Technische Universität München, Germany
[9] Karlsruher Institut für Technologie, Germany

Abstract. This paper presents the first comprehensive evaluation and analysis of modern deep-learning-based anomaly classification (AC) methods applied to chemical process data. The focus is on the Tennessee Eastman process (TEP) dataset, which includes simulations of 20 distinct types of process faults. This dataset is widely recognized as a benchmark for evaluating anomaly detection methods on chemical process data. After detecting a fault in a plant, the crucial task is to identify the type of fault. From a machine learning perspective, process data are time-series well suited for analysis using deep-learning architectures, which have been driving breakthroughs such as ChatGPT. This paper evaluates the application of contemporary deep-learning-based time-series classification methods applied to chemical process data. The findings of this study may represent a significant step toward classifying anomalies in chemical plants.

Keywords: Anomaly classification · Chemical Process Data · Tennessee Eastman process · time-series

1 Introduction

Identifying data that deviates from the typical pattern—so-called *anomalies*—is a fundamental technique in machine learning and artificial intelligence. Anomaly detection (AD) is significant in various application areas, from identifying fake reviews in e-commerce and detecting bots on social networks to diagnosing tumors and monitoring industrial faults [22, 23, 30]. AD is especially relevant in safety-critical applications. For instance, failing to detect anomalies in a chemical plant or diagnose a malicious tumor in a patient puts lives at risk.

In chemical plants, data is predominantly collected during normal, fault-free operations. Anomalies are rare and can often be mistaken for normal behavior by process or control engineers. This is where computational methods, particularly machine learning, become essential. These methods can sift through massive datasets to accurately identify rare anomalies [9]. The literature on learning-based AD in chemical processes is extensive [13,25,33]. The Tennessee Eastman Process (TEP) has established itself as a standard benchmark for evaluating new anomaly detection techniques on chemical process data. Created by Downs and Vogel [8] using a model-based TEP simulator, Rieth later modified the TEP dataset [29], and now, most novel methods are evaluated on the TEP dataset.

However, AD is only the first step. After detecting an anomaly in a dataset, the crucial next step is *determining the specific type of anomaly.* This task of *anomaly classification* (AC)—also called *fault diagnosis* in engineering—is a well-known problem in the machine-learning community. From a mathematical perspective, AC aims to distinguish one anomaly from another. Achieving this opens up more in-depth knowledge about the data features, offers further analysis, and makes weighing different anomaly types possible. On the other hand, from a more practical point of view, since we consider the case of chemical engineering in this paper, for any plant operator, it is essential to diagnose a detected anomaly to determine the cause of the problem. This is particularly significant in complex industrial processes, such as those in chemical plants, where a precise classification is necessary to implement appropriate corrective actions. From a machine-learning perspective, this is a multiclass classification problem on multivariate time-series. Traditional anomaly detection methods are adept at flagging anomalies but fall short in categorizing them, which is essential for accurate diagnosis and resolution of the issue. Effective AC can provide deeper insights into the underlying causes of the detected anomalies, facilitating better decision-making and maintenance strategies [4, 7, 36].

Some research has been done on the AC of the TEP dataset, often relying on simple and shallow methods. Early studies investigated linear models and basic machine learning techniques, which struggle with the TEP data's highly non-linear and dynamic nature. Principle Component Analysis (PCA) and k-nearest Neighbors (k-NN) have been studied. However, they lack the robustness and accuracy required for complex AC tasks in industrial settings [4].

With the advent of deep learning, the classification of time-series data has been significantly improved. Deep learning models such as Convolutional Neural networks (CNNs) and Recurrent Neural Networks (RNNs) have shown superior performance in capturing the intricate temporal and spatial dependencies present in time-series data [5, 12, 21, 31]. Recently, self-attention mechanisms have been inserted into deep models, resulting in powerful transformer architectures and ultimately driving advances such as Chat-GPT. Transformers have also emerged as a powerful tool for time-series classification thanks to their ability to model long-range dependencies [16, 35]. The ability of transformer models to handle large and high-dimensional data makes them particularly suited for industrial applications where precise AC is crucial [35].

This benchmark evaluation presents the first and most comprehensive comparison of modern AC methods on chemical process data, including evaluations of many methods based on deep learning. This analysis also yields insights into which classification methods might most suitably apply to chemical process data. Establishing reliable methods for this classification task will open new ways to control these processes in the future to increase the safety and reliability of chemical plants. It will open new perspectives on the autonomous running of chemical processes.

The main contributions of this paper are as follows.

- This work identifies AC as the logical follow-up of the anomaly detection problem and evaluates numerous deep and shallow methods for this purpose.
- In a comprehensive analysis, 27 combinations of various classification methods and dimensionality reduction preprocessing are evaluated and ranked according to their average F1-Score.
- The analysis reveals that several deep-learning-based methodologies exhibit superior F1-Score in AC compared to their shallow counterparts. This underscores the necessity for sophisticated multivariate time-series classification techniques in the domain of chemical process data.

2 Related Work

Early methods in diagnosing faults in TEP data include Principal Component Analysis (PCA) based techniques and Fisher discriminant analysis (FDA) [4]. More recently, Bayesian networks have been studied. Verron et al. [36] first identified important variables by computing the mutual information between each process variable and the class variable. Then, a Bayesian classifier known as a tree-augmented network (TAN) was used to classify the faulty process. However, the new faults could not be classified since the classification was done using the reduced space of variables. Santos et al. [7] proposed a dynamic Markov blanket classifier to find relationships among the most relevant variables without any variable selection method.

Traditional approaches to anomaly detection on the TEP dataset often involved statistical methods and linear models, such as PCA and Partial Least Squares (PLS), which were used to monitor process variables and detect deviations indicative of faults [32]. Yin et al. [39] evaluated various methods, including PCA, PLS, independent component analysis (ICA), FDA, and subspace-aided approach (SAP) on TEP data. However, these methods typically struggle with capturing non-linear relationships in the data. Recent advancements have included adopting machine learning and deep learning techniques to improve the accuracy and robustness of anomaly detection in the TEP dataset. A few studies exist where the auto-associative neural networks (autoencoders - AE) have been used for fault detection [17,38,42]. However, they were not comparable because of the difference in hyperparameter settings, training objective functions, and the number of training samples. In addition, the principal components extracted from auto-associative neural networks were sometimes redundant due to co-adaptation in the early phase of training [20]. The objective of this work is to address these limitations. Heo et al. [11] evaluated these methods to address all their limitations. Sun et al. [34] proposed a probabilistic fault detection using Bayesian recurrent neural networks (BRNNs) with variational dropout, capable of modeling complex non-linear dependencies. AEs and variational autoencoders (VAEs) have also been used extensively for fault detection. Zhu et al. [43] applied VAEs to identify the deviation process from the normal ones. In contrast, Yu et al. [40] integrated convolutional neural networks (CNNs) and gated recurrent units (GRUs) within an autoencoder framework to capture both spatial and temporal features from the TEP data. Finally, Hartung et al. [9] extensively studied 27 anomaly detection methods on TEP data and concluded that reconstruction-based, generative, and forecasting-based methods were particularly effective.

Multiclass classification on time-series data involves predicting one of several possible labels for each instance, where each label represents a distinct class. Traditional multiclass classification methods on time-series data rely on measuring the sequence similarity techniques, such as Dynamic Time Warping (DTW) and k-nearest neighbors (k-NN), often fail on high-dimensional and complex datasets due to their reliance on sequence similarity techniques [2]. Feature-based approaches apply a feature extraction method followed by classification algorithms such as Random Forests or Support Vector Machines [2], which may improve classification accuracy but still face challenges in capturing temporal dependencies. The emergence of deep learning marked a paradigm shift with Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Fawaz et al. and Wang et al. [15, 37] demonstrated that LSTM and CNNs have superior performance by leveraging their ability to model intricate patterns over time. More recently, Zervaes et al. [41] used transformer models to effectively model long-range dependencies using self-attention mechanisms. Cui et al. [6] proposed a novel Multi-Scale Convolutional Neural Networks (MCNNs), which integrated CNNs with attention mechanisms to enhance classification performance by capturing local and global patterns.

3 Benchmarking (modern) Time-series Classification Methods on the TEP

In this section, we begin by introducing the TEP dataset, followed by a review of the methods, including their implementation and evaluation. Finally, we present the results.

3.1 The Tennessee Eastman Process dataset

TEP dataset is a widely recognized benchmark in process control and fault diagnosis [28]. The TEP simulates a chemical production process, generating data that includes various normal and faulty operational states. The dataset contains a variety of variables that capture the dynamics of the process, including pressures, temperatures, and flow rates. Each record in the dataset is tagged with a fault number (0 for fault-free, 1-20 for faulty runs), representing the various issues introduced into the system. We will only focus on the 20 fault classes. This dataset is essential in developing and evaluating anomaly detection and classification algorithms due to its complexity and realistic reproduction of industrial scenarios.

3.2 Methods

This study compares several modern time-series classification methods to benchmark their performance on the TEP dataset. The methods include:

- **LSTM-FCN:** The Long Short-Term Memory Fully Convolutional Network [19] is a hybrid model that combines the strengths of LSTMs for capturing temporal dependencies and CNNs for feature extraction.
- **Deep CNN:** Deep Convolutional Neural Networks [1] are powerful for image-like data processing.
- **TCN:** Temporal Convolutional Networks [10] are effective for sequential data, offering long memory and stable gradients.
- **RNN:** Recurrent Neural Networks [24] are suitable for time-series data due to their sequential nature.
- **LSTM:** Long Short-Term Memory Networks [12] address the vanishing gradient problem in RNNs, making them practical for long-term dependencies.
- **XGBoost:** XGBoost [27] is a gradient-boosting framework known for its high performance with structured data.
- **Random Forest:** Random Forest [26] is an ensemble learning method that operates by constructing multiple decision trees.
- **SVM:** A Support Vector Machine [14] is a model used for classification and regression tasks, capable of finding optimal decision boundaries between different classes.
- **WaveNet:** WaveNet [18], initially designed for audio synthesis, is applied here for time-series classification.

3.3 Implementation Details

In the pre-processing phase, the data was scaled using StandardScaler, so all features have a mean zero and a unit variance. Afterward, all data was reshaped for sequence-based models into sequences of 500 timestamps for training and 960 timestamps for testing. The scaled data was aggregated for aggregation-based models by averaging features over fixed-length windows. This approach reduces the dimensionality and provides summary statistics for each window. Eventually, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were applied as specified for further dimensionality reduction (DR). PCA retained the most significant components, while LDA maximized class separability.

Dimensionality Reduction with PCA and LDA Dimensionality reduction plays a critical role in enhancing the performance of anomaly classification methods, particularly PCA and LDA. PCA and LDA contribute significantly to improving anomaly classification methods' efficacy by transforming high-dimensional TEP data into more informative and compact representations.

Each method underwent the same pre-processing procedure, followed by a grid search to fine-tune their hyperparameters. The optimal hyperparameters were chosen based on the training accuracy, and subsequently, the test data was used to evaluate performance.

3.4 Evaluation Metric

To compare and evaluate the performance of the classification models, the classification report library was used, which includes precision, recall, and F1-Score for each class, offering a comprehensive view of the model's performance across all fault types.
Using a standard evaluation metric across all methods is crucial for fair comparison. The F1-Score is used as the primary metric in this study, as it balances precision and recall, providing a comprehensive measure of model performance, especially on imbalanced datasets.

3.5 Results

Table 1 summarizes the experiments' results.

The LSTM-FCN model consistently achieved competitive performance across all feature extraction and dimensionality reduction methods and was ranked first in the average F1-score. Deep CNN and TCN also showed strong performance, highlighting their ability to handle complex data patterns. Traditional machine learning models such as XGBoost and Random Forest showed competitive results but were generally outperformed by the deep learning approaches concerning F1-score and accuracy. These results underscore the importance of leveraging advanced time-series models for complex datasets like TEP. Future work could explore hybrid models and optimize the architectures to improve performance. Also, we see that PCA is not a suitable dimensionality reduction method here, as the F1-Score decreased for every model when applied before training. The same does not hold for LDA.

In addition to the above findings, it is noteworthy that WaveNet exhibited overfitting tendencies on the TEP dataset. Despite achieving an initial high accuracy in training (over 0.90), its F1-Score on the test set was notably lower, at an average of 0.57. This discrepancy suggests that WaveNet may have struggled to generalize effectively to unseen data even if Regularizer, BatchNormalization, and dropout layers have been applied. This highlights the importance of addressing overfitting in model training and evaluation strategies for future studies.

Table 1. Performance Comparison of AC Methods on TEP Dataset showing deep learning methods outperforming traditional machine learning methods. The evaluation was done without further dimensionality reduction (DR) and with DR using PCA and LDA. The average result for all three DR options was computed for every method to receive a comparable ranking.

Model	Preprocess	DR	F1-Score	Rank	AVG F1	AVG Rank
LSTM-FCN	reshaped	No DR	0.90	6	0.92	1
		PCA	0.89	9		
		LDA	0.98	1		
Deep CNN	reshaped	No DR	0.96	2	0.91	2
		PCA	0.86	12		
		LDA	0.91	4		
TCN	reshaped	No DR	0.90	6	0.91	3
		PCA	0.89	9		
		LDA	0.93	3		
RNN	reshaped	No DR	0.90	6	0.86	4
		PCA	0.78	14		
		LDA	0.91	4		
LSTM	reshaped	No DR	0.84	13	0.81	5
		PCA	0.71	19		
		LDA	0.88	11		
XGBoost	aggregated	No DR	0.72	18	0.71	6
		PCA	0.66	21		
		LDA	0.74	16		
Random Forest	aggregated	No DR	0.70	20	0.70	7
		PCA	0.62	23		
		LDA	0.77	15		
SVM	aggregated	No DR	0.62	23	0.65	8
		PCA	0.60	26		
		LDA	0.73	17		
WaveNet	reshaped	No DR	0.62	23	0.57	9
		PCA	0.44	27		
		LDA	0.66	21		

4 Conclusion and Future Work

In this study, we benchmarked various modern time-series classification methods on the Tennessee Eastman Process (TEP) dataset, including LSTM-FCN, Deep CNN, TCN, RNN, LSTM, XGBoost, Random Forest, SVM, and WaveNet. Our results, summarized in Table 1, indicate that deep learning models, particularly those with sequence-based preprocessing, generally outperform traditional machine learning approaches.

The LSTM-FCN model with LDA preprocessing achieved the highest F1-Score of 0.98, demonstrating its effectiveness in capturing temporal dependencies and distinguishing between the different classes in the TEP dataset. Deep CNN and TCN models also performed well, particularly when combined with LDA, highlighting the importance of feature extraction techniques in enhancing model performance. However, it is noteworthy that PCA did not increase performance in any model, indicating that not all dimensionality reduction techniques are equally beneficial for this task.

Traditional machine learning models such as XGBoost, Random Forest, and SVM, while generally less effective than deep learning models, still provide valuable insights. The aggregation preprocessing technique for these models helped improve their performance, but they were unable to match the accuracy of the deep learning models.

Future work could explore the integration of hybrid models that combine the strengths of both deep learning and traditional machine learning approaches. Further investigation into advanced feature extraction techniques and their impact on model performance could yield even better results. Exploring other deep learning architectures, such as attention-based models or transformers, could also provide new insights and improve classification accuracy on the TEP dataset. Applying transfer learning, where models pre-trained on similar industrial datasets are fine-tuned on the TEP dataset, could be another promising direction for future research. Moreover, utilizing representation learning on the TEP dataset can facilitate more accurate identification and classification of patterns within the data, leveraging the intrinsic relationships between samples for better anomaly classification outcomes [3, 23].

References

1. Aloysius, N., Geetha, M.: A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing (ICCSP). pp. 0588–0592 (2017). https://doi.org/10.1109/ICCSP.2017.8286426
2. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data mining and knowledge discovery **31**, 606–660 (2017)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)

4. Chiang, L.H., Russell, E.L., Braatz, R.D.: Fault detection and diagnosis in industrial systems. Springer Science & Business Media (2000)
5. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
6. Cui, Z., Chen, W., Chen, Y.: Multi-scale convolutional neural networks for time series classification. arXiv preprint arXiv:1603.06995 (2016)
7. Dos Santos, E.B., Ebecken, N.F., Hruschka Jr, E.R., Elkamel, A., Madhuranthakam, C.M.: Bayesian classifiers applied to the tennessee eastman process. Risk Analysis **34**(3), 485–497 (2014)
8. Downs, J.J., Vogel, E.F.: A plant-wide industrial process control problem. Computers & chemical engineering **17**(3), 245–255 (1993)
9. Hartung, F., Franks, B.J., Michels, T., Wagner, D., Liznerski, P., Reithermann, S., Fellenz, S., Jirasek, F., Rudolph, M., Neider, D., et al.: Deep anomaly detection on tennessee eastman process data. Chemie Ingenieur Technik **95**(7), 1077–1082 (2023)
10. He, Y., Zhao, J.: Temporal convolutional networks for anomaly detection in time series. Journal of Physics: Conference Series **1213**(4), 042050 (jun 2019). https://doi.org/10.1088/1742-6596/1213/4/042050, https://dx.doi.org/10.1088/1742-6596/1213/4/042050
11. Heo, S., Lee, J.H.: Statistical process monitoring of the tennessee eastman process using parallel autoassociative neural networks and a large dataset. Processes **7**(7), 411 (2019)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
13. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artificial intelligence review **22**, 85–126 (2004)
14. Islam, M.M., Sultana, N.: Comparative study on machine learning algorithms for sentiment classification. International Journal of Computer Applications **182**(21), 1–7 (2018)
15. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data mining and knowledge discovery **33**(4), 917–963 (2019)
16. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. Advances in neural information processing systems **28** (2015)
17. Jiang, L., Ge, Z., Song, Z.: Semi-supervised fault classification based on dynamic sparse stacked auto-encoders model. Chemometrics and Intelligent Laboratory Systems **168**, 72–83 (2017)
18. Jonathan, B., Philippe, G., roch, L.: A literature review of wavenet: Theory, application, and optimization. Journal of the Audio Engineering Society (10171) (march 2019)
19. Karim, F., Majumdar, S., Darabi, H.: Insights into lstm fully convolutional networks for time series classification. IEEE Access **7**, 67718–67725 (2019). https://doi.org/10.1109/ACCESS.2019.2916828
20. Kramer, M.A.: Nonlinear principal component analysis using autoassociative neural networks. AIChE journal **37**(2), 233–243 (1991)
21. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
22. Li, A., Qiu, C., Kloft, M., Smyth, P., Mandt, S., Rudolph, M.: Deep anomaly detection under labeling budget constraints. In: International Conference on Machine Learning. pp. 19882–19910. PMLR (2023)

23. Li, A., Qiu, C., Kloft, M., Smyth, P., Rudolph, M., Mandt, S.: Zero-shot anomaly detection via batch normalization. Advances in Neural Information Processing Systems **36** (2024)
24. Medsker, L.R., Jain, L., et al.: Recurrent neural networks. Design and Applications **5**(64-67), 2 (2001)
25. Monroy, I., Escudero, G., Graells, M.: Anomaly detection in batch chemical processes. In: Computer Aided Chemical Engineering, vol. 26, pp. 255–260. Elsevier (2009)
26. Primartha, R., Tama, B.A.: Anomaly detection using random forest: A performance revisited. In: 2017 International Conference on Data and Software Engineering (ICoDSE). pp. 1–6 (2017). https://doi.org/10.1109/ICODSE.2017.8285847
27. Ramraj, S., Uzir, N., Sunil, R., Banerjee, S.: Experimenting xgboost algorithm for prediction and classification of different datasets. International Journal of Control Theory and Applications **9**(40), 651–662 (2016)
28. Ricker, N.L.: Tennessee eastman process simulation dataset (1995), https://depts.washington.edu/control/LARRY/TE/download.html, accessed: 2024-06-21
29. Rieth, C.A., Amsel, B.D., Tran, R., Cook, M.B.: Additional tennessee eastman process simulation data for anomaly detection evaluation (2017). https://doi.org/10.7910/DVN/6C3JR1
30. Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.R.: A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE **109**(5), 756–795 (2021)
31. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. nature **323**(6088), 533–536 (1986)
32. Russell, E.L., Chiang, L.H., Braatz, R.D.: Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. Chemometrics and intelligent laboratory systems **51**(1), 81–93 (2000)
33. Song, B., Suh, Y.: Narrative texts-based anomaly detection using accident report documents: The case of chemical process safety. Journal of loss prevention in the process industries **57**, 47–54 (2019)
34. Sun, W., Paiva, A.R., Xu, P., Sundaram, A., Braatz, R.D.: Fault detection and identification using bayesian recurrent neural networks. Computers & Chemical Engineering **141**, 106991 (2020)
35. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
36. Verron, S., Tiplica, T., Kobi, A.: Fault diagnosis with bayesian networks: Application to the tennessee eastman process. In: 2006 IEEE international conference on industrial technology. pp. 98–103. IEEE (2006)
37. Wang, Z., Yan, W., Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 International joint conference on neural networks (IJCNN). pp. 1578–1585. IEEE (2017)
38. Yan, W., Guo, P., Li, Z., et al.: Nonlinear and robust statistical process monitoring based on variant autoencoders. Chemometrics and Intelligent Laboratory Systems **158**, 31–40 (2016)
39. Yin, S., Ding, S.X., Haghani, A., Hao, H., Zhang, P.: A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. Journal of process control **22**(9), 1567–1581 (2012)

40. Yu, J., Liu, X.: A fault detection method based on convolutional gated recurrent unit auto-encoder for tennessee eastman process. In: 2020 Chinese Automation Congress (CAC). pp. 1234–1238. IEEE (2020)
41. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C.: A transformer-based framework for multivariate time series representation learning. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. pp. 2114–2124 (2021)
42. Zhang, Z., Jiang, T., Li, S., Yang, Y.: Automated feature learning for nonlinear process monitoring–an approach using stacked denoising autoencoder and k-nearest neighbor rule. Journal of Process Control **64**, 49–61 (2018)
43. Zhu, J., Jiang, M., Liu, Z.: Fault detection and diagnosis in industrial processes with variational autoencoder: A comprehensive study. Sensors **22**(1), 227 (2021)