Representation Learning for Zero-Shot Anomaly Detection

Johannes Hölker

[1] University of Kassel
[2] Germany
johannes.hoelker@uni-kassel.de

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: Representation Learning · Zero Shot Learning · Anomaly Detection.

1 Introduction

In industrial settings several applications rely on multi variate time series data. This could be sensor measurements or machine state values. The machine reacts to these values and works according to them. Normally the values are changing constantly in repetetive patterns. This is when the measured data or the machine is running uninterrupted like it is supposed to. Sometimes the values change unpredicted because of differing surroundings or other influences. This can lead to faults which may break the machine.

Recognising and reacting to these changes can be important. Considering the example of a machine fault, the early detection can possibly prevent further damages. But interruptions occur in different forms. In some cases they never occured like this before. This raises the demand for a tool to detect anomalies in time series data without any further knowledge of the anomaly.

This paper conducts a detailed literature research to contribute in the development of a anomaly dection tool. It provides an overview of the latest trends in representation learning and extracts the possible solutions addressing the described probelm. The paper focuses on the following research question:

What are the different types of representation learning possible for Zero Shot Anomaly Detection in time series applications?

A literature research concerning the topic is conducted and the best choices are implemented on a test data set. We begin in 2 by defining necessary terms and how they are used. In 3 the found papers and their methodologies are presented and explained. The methodologies are compared and evaluated for usability in the context of Zero-Shot Anomaly Detection in 4. The implementation of suiting techniques is provided in 5. Finally the results are discussed and concluded in 6

2 Definitions and Conventions

2.1 Representation Learning

Representation Learning (RL) tries to detect meaningful interconnections in data relevant for further data analysis. There are several representation learning techniques to detect patterns and to store them in different ways. In Bishop (2006) representation learning occurs in several machine learning areas. In neural networks representations are learned in every hidden layer. In that case the representations are not symbolic representations that we as humans see. Cognitive repesentations can in that sense be seperated into neural, spatial and symbolic (Gärdenfors, 2000).

To extract symbolic or spatial features which are more comprehensible for us a knowledge discovery process with different methods of machine learning and data mining methods are used (Lavrač, Podpečan, & Robnik-Šikonja, 2021, p. 4). RL techniques are devided into Propositionalization as symbolic representations and Embeddings as numeric representations.

In the book of Goodfellow, Bengio, and Courville (2016) a general detailed description of representation learning is given. They summarize that representations should make the subsequent learning tasks easier. This implies that to find the best fitting representation and the underlying representation learning technique, we need to know the task it should perform afterwards.

One solution to learn representations is contrastive learning. Pairs of data points are labeled as similar and dissimilar. These data points are put into a feature space where the distance between the two represents their similarity. With a contrastive loss function and a label of similarity between two points, the model is trained by putting the similar data points together and separating dissimilar points. Using this method groups of similar data points are formed.

Autoencoders are another prominent method in representation learning. An autoencoder is a type of neural network used to learn efficient codings of input data in an unsupervised manner. It consists of an encoder that compresses the input into a latent-space representation and a decoder that reconstructs the input from this representation. The goal is to minimize the difference between the input and the reconstructed output.

Transformers, initially developed for natural language processing tasks, have become a powerful tool in representation learning. They use self-attention mechanisms to weigh the significance of each part of the input data differently, enabling the model to capture long-range dependencies (Vaswani et al., 2017).

In summary, representation learning can be achieved using different techniques, each suitable for different types of data and tasks. From neural networks and autoencoders to transformers, these methods provide the tools necessary to transform raw data into meaningful representations that facilitate further analysis and learning.

Evaluation This chapter describes how to evaluate the performance of a RL approach. Bengio, Courville, and Vincent (2013) describe what makes a representation "good". They list the following factors:

- Smoothness
- Multiple Explanatory Factors
- A hierarchical organization of explanatory factors
- Semi-supervised learning
- Shared factors across tasks
- Manifolds
- Natural clustering
- Temporal and spatial coherence
- Sparsity
- Simplicity of factor dependencies

"We want to find properties of the data but at the same time we don't want to loose information about the input" (Goodfellow et al., 2016, S. 525)

2.2 Zero Shot Learning

Zero Shot Learning is an extreme form of transfer learning (Goodfellow et al., 2016, S. 536). While transfer learning is the concept of transferring the knowledge and weights gained at one task using them at solving another task, Zero-Shot Learning means there are no samples for the other task. The transformation of knowledge can help solving tasks where there are few or no samples available. The gained knowledge is normally stored as representations in the data. Representations which are abstract enough to not see a specific item but information about items which can be applicated to groups of items. This also means that Zero-Shot Learning is only possible because additional information has been discovered during training (Goodfellow et al., 2016, S. 536).

Palatucci, Pomerleau, Hinton, and Mitchell (2009) were the first to implement a successful Zero-Shot Detection followed by Socher et al. (2013) who used semantic word vector representations to classify words in groups and to sort new words with an accuracy of 90% with a fully unsupervised model.

2.3 Anomaly Detection

Several definitons of anomalies in data can be found in literature. In this paper the definition of Gruhl (2022, S. 54) is used. It seperates anomaly and novelty detection as different tasks. Anomalies can be understood as outliers from the regular class. But these anomalies can vary in their cause. If there is a specific cause and the anomalies occur in its own cluster, they form a novelty. If instead the outliers randomly occur with no specific root cause, they are called noise. The cause for noise then is of a different kind and cannot be classified.

Instead of dividing anomalies by their cause the shape of anomalies can vary in several ways. In real measurement data any shape is possible and it is totally unpredictable (Schwartz et al., 2024). For training purposes anomaly injection is crucial. Then the anomalies are simulated as point anomalies or subsequence anomalies. Point anomalies occur once and can be global or contextual. Subsequence anomalies on the other hand change the values in a given time window or on long term. They can be divided in seasonal, shapelet and trend anomalies. Seasonal and shapelet change the values in a limited time window, trend anomalies are changing all following values (Darban, Webb, Pan, Aggarwal, & Salehi, 2024, p. 9).

In this paper we want to focus on single time events, which are in any case anomalies. Potentially being caused by an unknown process, they cannot be classified (Gruhl, 2022). This defines our goal as an anomaly detection task.

3 Representation Learning for Time Series Data

In this chapter the found literature is put into context. Starting with classical literature about the fundamental findings followed by actual trends in the Area of Representation Learning. Finally the different Representation Learning Strategies are listed and compared.

3.1 Historical view

In this chapter the fundamental literature about Representation Learning is going to be discussed. Sensors and comparable applications produce values which vary over time. Sometimesthe values vary in an unforeseen way and for a short time window they may be completely random. We have to step back and observe longer time periods which could be days or weeks. Or, for very dense measuring it is shorter but there are way more data points to handle.
Sometimes it is possible for a human to see some patterns in the data when observing a long time window. Take for example the measuring of a solar plant. On a daily basis it is obvious to see the sun rising and setting, depending on the voltage of the panels. Starting at 0 at night the voltage is rising before noon and descending in the afternoon. This is one representation in the data. But there could be more represenations hidden, which are not likely to see. The shadow of a tree wandering over the panels happening every day or a one time event like the snow covering the plant.

These variations in data are not always visible for a human and even less possible to label them accordingly. Like (Bengio et al., 2013) mentioned it is important for artificial intelligence to detect these representations in data by machines. A machine should be able to extract information hidden in the low-level sensor measurings and continue working with the representations instead of the raw data. This is according to the paper the main requirement for a good representation, to be able using it as an input to a supervised predictor.

Since the paper came out in 2013, several representation learning techniques were developed and some of them are directly applicable for time series data. In (Sun & Ge, 2021) the importance of machine learning in sensor data is emphasized. They sum up several deep learning techniques on data-driven soft-sensors. Soft-sensors represent hard to measure variables by adapting available sensor data. Their observation of industry processes is a rapidly changing field which demands data processing for a huge amount of data.

3.2 Representation Learning Strategies

The different RL strategies are listed, explained and compared. The strategies are organized by their underlying concept. We begin with straight-forward methods which are based on one concept and increase the complexity throughout the chapter. In the end methods which use combinations of different concepts are presented.

MLP Using a simple Multi Layer Perceptron (MLP) is a straight-forward way to learn representations and to detect anomalies in time series data. The input variable for the MLP are time points and the output variable represents the value at these time points. The model is trained to learn this mapping. With the trained model, the values in a live scenario are predicted and the difference to the actual values is measured. If tis representation error exceeds a certain threshold, an anomaly is found (Jeong & Shin, 2022). The model is trained with data including anomalies so it is not suitable for Zero-Shot Learning. It is theoretically possible to adapt the model for Zero-Shot Anomaly Detection but no further publications based on MLP are found.

Contrastive Learning Learning representations in time series data is done in a several different ways. One solution according to Zhang, Cai, Zhou, and Liu (2024) is debiased contrastive learning. By comparing pairs of data points and rating the similarities as distances between the two, contrastive learning gets less dependant on labeled data. The data can be more general and the extracted representations are more robust. The pairs of data points are labeled as positive and negative pairs with a distance according to their similarities. With this distance they are put into a feature space where they form groups of data points. To minimize the bias between representations multigranularity augmented view generation and expert knowledge are used during training.

The proposed framework is applied on industrial fault detection. The two data sets consist of various vibration signals of industrial machines and stiction sensors. The effectiveness of the proposed framework is demonstrated through its application to these datasets, where it shows improved performance in fault detection compared to traditional methods (Zhang et al., 2024).

Contrastive Representation Learning is also used to tackle anomaly detection in time series data by Darban et al. (2024). They use CL combined with synthetic anomaly injection. CL enables them to capture patterns in time series data and the framework shows good results on common real world datasets. Similar to the previous paper, dissimilar pairs, the anomalies, build distant data points and similar data points are close to each other. In order to train the model artificial anomalies are injected which build distant pairs. In the next stage the classification is done by the proximity of the neighbours in the representation space. Additionally anchor points representing the nearest and furthest neighbour are given from each representation. Their methodology is called CARLA and is also not tested for Zero-Shot Learning.

The article by Ngu and Lee (2023) introduces CL-TAD, a novel method for time series anomaly detection that leverages contrastive learning and reconstruction-based techniques to address the challenges of temporal dynamics, label scarcity, and data diversity in real-world applications. The method comprises two main components: positive sample generation and contrastive-learning-based representation learning. Positive samples are generated by reconstructing masked parts of the time series data, helping the model learn the underlying normal patterns. These samples, along with the original data, are then fed into a contrastive learning framework, which contrasts pairs of similar (positive) and dissimilar (negative) samples to learn robust representations. This process helps the model map similar data points closer together in the feature space while distancing dissimilar points, making it easier to detect deviations indicative of anomalies. Extensive experiments on nine benchmark datasets show that CL-TAD outperforms ten other recent methods in detecting anomalies, highlighting its effectiveness in handling diverse and complex time series data (Ngu & Lee, 2023). While CL-TAD is not explicitly designed as a zero-shot learning method, its use of contrastive learning and reconstruction-based techniques suggests that it could have potential in zero-shot anomaly detection scenarios. However, this would depend on the model's ability to generalize from the learned normal patterns to detect unseen anomalies. Further empirical studies would be needed to validate its performance in zero-shot learning scenarios.

To succeed on Zero-Shot Anomaly Detection, the method of One-Class Classification (OCC) can solve the problem. By gathering all "normal" values into a single class the outliers are directly detected if they are outside of it. The paper by Lee, Byun, and Baek (2023) presents an approach for detecting anomalies using OCC in industrial time series data, which typically lacks labels for supervised learning. Thecombine OCC with contrastive learning to define a new objective function that can simultaneously learn from both models. This method enhances feature extraction while preserving temporal characteristics. The paper demonstrates the method's effectiveness through high anomaly detection performance on datasets with similar normal and anomalous data forms, highlighting its potential in industrial applications (Lee et al., 2023).

Yue et al. (2022) introduce TS2Vec, a framework for learning robust and universal time series representations at multiple semantic levels through hierarchical contrastive learning. This approach utilizes timestamp masking and random cropping to create augmented context views, enhancing position-agnostic and comprehensive representations. By combining instance-wise and temporal contrastive losses, TS2Vec captures unique characteristics of different time series instances and dynamic temporal patterns within each series. Extensive experiments show that TS2Vec outperforms state-of-the-art methods in classification, forecasting, and anomaly detection tasks across 125 UCR and 29 UEA datasets, achieving average improvements of 2.4% and 3.0% in classification accuracy, respectively. The framework's efficiency in training time further underscores its practical utility. TS2Vec demonstrates its versatility by excelling in multiple time series analysis tasks, making it a significant contribution to the field. The framework's hierarchical contrastive learning at various scales encapsulates rich and meaningful patterns in time series data (Yue et al., 2022).

Autoencoder Nivarthi, Vogt, and Sick (2023) are the first to use a Unified Autoencoder (UAE) for time series data, namely the power forecast of wind and solar plants. They contribute to the challenge of predicting the possible outcome of renewable energy in a newly created plant, either wind or solar. To do so a UAE is combined with a Task Embedding Neural Network (TENN) They examine the usability divided in Single-Task, Multi-Task and Zero-Shot Learning. The method was first published in Nivarthi, Vogt, and Sick (2022). It is then extended by convolutional layers instead of the fully connected neural network layers (UCAE-TENN) and also Long Short-Term Memory layers (ULAE-TENN).

Realising few-shot anomaly detection of images is done by Schwartz et al. (2024). The method MAEDAY can detect objects newly added to the frames. To achieve this a masked autoencoder is used who recreates the former image but without the anomaly. The difference between the initial and reconstructed images is calculated and the object then visible. This method is useful for its ability to detect anomalies with very few examples, making it a powerful tool in scenarios where labeled data is rare. Schwartz et al. (2024)

demonstrate the effectiveness of MAEDAY in various applications, showcasing its potential for real-world anomaly detection tasks.

To detect anomalies in healthcare data a variational recurrent autoencoder is used by Pereira and Silveira (2019). The focus is on electrocardiogram (ECG) datasets. Their method tackles the challenge of finding anomalies in unlabelled time series data. They created an unsupervised framework where the model learns to represent the data and detect anomalies without needing labeled examples. The VRAE model works by learning to reconstruct the input sequences. During training, they add noise to the input data, and the model tries to reconstruct the original, uncorrupted data. This helps the model learn more robust representations of the data. To detect anomalies, they cluster these learned representations and use the Wasserstein distance to identify outliers. Their approach was tested on the ECG5000 dataset and showed that it could effectively detect unusual heartbeats, performing better than previous methods that required labeled data.

Another approach using VRAE involves creating synthetic anomalies to improve the detection process. In their method, they use a two-level hierarchical latent space representation. First, they distill feature descriptors of normal data points into more robust representations using autoencoders (AEs). These representations are then refined using a variational autoencoder (VAE) that creates a family of distributions. From these distributions, they select those that lie on the outskirts of the normal data as generators of synthetic anomalies. By generating these synthetic anomalies, they train binary classifiers to distinguish between normal and abnormal data. Their hierarchical structure for feature distillation and fusion helps create robust representations, enabling effective anomaly detection without needing actual anomalies during training. Their method performs well on several benchmarks for anomaly detection (Ramirez Rivera, Khan, Bekkouch, & Sheikh, 2022).

Pranavan, Sim, Ambikapathi, and Ramasamy (2022) present a novel approach for anomaly detection in multi-variate time series data using Contrastive Predictive Coding (CPC). Their method, named Time-series Representational Learning through Contrastive Predictive Coding (TRL-CPC), aims to effectively capture the temporal dependencies and correlations across multiple variables in time series data. The TRL-CPC framework consists of an encoder, an autoregressive model, and a non-linear transformation model. These components are jointly optimized to learn the representations of multi-variate time series data by predicting future segments from past segments. The core idea is to maximize the mutual information between the encoded representations of past and future segments, thereby learning robust representations. To detect anomalies, TRL-CPC calculates the prediction error between actual future segments and the predicted segments generated by the CPC model. Anomalies are identified where this prediction error exceeds a certain threshold, enabling unsupervised anomaly detection based on the structure of the data itself. Experimental results show that TRL-CPC outperforms traditional anomaly detection methods on several benchmark datasets, highlighting its effectiveness in capturing complex temporal dependencies and identifying anomalies in multi-variate time series data (Pranavan et al., 2022).

Zhang et al. (2024) point out, that AE-based methods have remaining challenges.

Stochastic Recurrent Neural Network Su et al. (2019) propose a method called OmniAnomaly for robust anomaly detection in multivariate time series data using a Stochastic Recurrent Neural Network (SRNN). This approach addresses the challenge of detecting anomalies in complex, high-dimensional time series data, which is common in applications such as network monitoring, industrial systems, and healthcare. Their method utilizes an SRNN to model the temporal dependencies and stochasticity in multivariate time series data. By incorporating stochastic units into the recurrent neural network, the model can capture the underlying uncertainty and variability in the data. This allows for more accurate detection of anomalies, as the model can differentiate between normal fluctuations and genuine anomalies. The key advantage of this method is its robustness to noisy and high-dimensional data. The SRNN learns to represent the normal patterns in the time series and identifies deviations from these patterns as anomalies. The model is evaluated on several benchmark datasets and demonstrates superior performance compared to state-of-the-art methods in terms of both precision and recall (Su et al., 2019).

Transformer To overcome the challenge of poorly available time series data sets (Ma et al., 2023), the model family MOMENT tries to learn general patterns on a pile of time series data (Goswami et al., 2024). The pile is a collection of different datasets which they assembled for their pretraining. According to the paper minimal finetuning is needed to perform well on time series tasks like anomaly detection. They published the model and made the usage easily accessible with its own python library. The constructed time serie pile consists of a widespread list of domains including Weather measurements, sensor values and power consumption datasets. They also included data not connected with the previous like the tongue and finger movement of humans. The different tasks which the model is evaluated on are forecasting (long and short horizon), classification, anomaly detection and imputation. Except for short-horizon forecasting all tasks are managed well.

TODO (Jiao, Yang, Song, & Tao, 2022) TimeAutoAD

TODO (Zhou, Pang, Tian, He, & Chen, 2024) AnomalyCLIP

TODO (Li et al., 2024)

Shapelet Learning Beggel, Kausler, Schiegg, Pfeiffer, and Bischl (2019) address the problem of detecting anomalies in time series data using a novel unsupervised method based on shapelet learning. This approach is particularly useful in scenarios where labeling data is difficult and expensive. Their method learns representative features that describe the shape of time series data from the normal class and simultaneously learns to accurately detect anomalies. The objective function encourages the learning of a feature representation in which normal time series lie within a compact hypersphere, while anomalous observations lie outside the decision boundary. This is achieved through a block-coordinate descent procedure. The advantage of their approach is that it can efficiently detect anomalies in unseen test data without retraining the model, by reusing the learned feature representation. Experimental results on multiple benchmark datasets demonstrate the robustness and reliability of the method in detecting anomalous time series, outperforming competing methods when the training data contains anomalies (Beggel et al., 2019).

In contrast, Alshaer, Garcia-Rodriguez, and Gouy-Pailler (2020) propose a method combining matrix profiles with shapelet learning to handle streaming time series data. The matrix profile efficiently identifies potential anomalies in real-time, and shapelet learning characterizes these anomalies accurately. This approach is particularly suited for environments requiring immediate anomaly detection, such as finance, healthcare, and industrial monitoring (Alshaer et al., 2020).

While both methods utilize shapelet learning, Beggel et al. focus on static datasets and robust feature representation, whereas Alshaer et al. emphasize real-time detection in dynamic, streaming environments.

combinations TODO (Aota, Tong, & Okatani, 2023)

TODO (Li et al., 2023)

4 Application on Time Series Data

Which of the proposed RL types are best suited for Zero Shot Anomaly Detection in time series data? In this chapter a selection of appropriate methods for Time Series Data Anomaly Detection out of 3 is extracted. Here the priors described in 2.1 are used in order to rate the RL types.

In order to achieve a successful implementation in the next chapter, the focus is on the opensource availability of the described models. Only models which are available and well documented are chosen for further examination. A list of the availability is given in table X

Method Name	Author	Concept	Implementation	Tested on ZSL
INRAD	Jeong et al.	MLP	good	no
	Zhang et al.	CL		no
CARLA	Darban et al.	CL		no
CL-TAD	Ngu et al.	CL		yes
	Lee et al.	CL		
TS2Vec	Yue et al.	CL		
UCAE-TENN	Nivarthi et al.	AE		
MAEDAY	Schwartz et al.	AE		
	Pereira et al.	AE		
	Ramirez et al.	AE		
TRL-CPC	Pranavan et al.	AE		
OmniAnomaly	Su et al.	Stochastic RNN		
MOMENT	Goswami et al.	Transformer	python library	yes
TimeAutoAD	Jiao et al.			
AnomalyCLIP	Zhou et al.			
	Li et al.			
	Beggel et al.	Shapelet Learning		
	Alshaer et al.	Shapelet Learning		
	Aota et al.			
	Li et al.			

(Fung, Qiu, Li, & Rudolph, 2024)

5 Implementation

The best fitting strategies are implemented on a small test data set in order to demonstrate how it works.
TODO Include Link to code repo

5.1 Data Set including Anomalies

Which data set to choose for a valid proof of concept. The structure of the chosen data set is described in this chapter.
While NLP and image processing tasks are common and a variety of data sets exists, time series data sets are not available that much (Ma et al., 2023).
The transferability between time series datasets is difficult due to the fact that the data between domains is huge (Ma et al., 2023)
In the test data the learning data is seperate from the data including anomalies. The important thing about Zero Shot Learning is that a specific anomaly never occured like this before. In the test data, all chosen representation learning techniques are applied using the same data for learning and afterwards testing the anomaly detection with the same anomalies. According to chapter (Evaluation) the characteristics are evaluated for each RL technique chosen in the previous chapter.

To test the model with anomalies in a consistent data set the Server Machine Dataset (SMD) provided by Su et al. (2019) is used. The SMD (Server Machine Dataset) is a 5-week-long dataset made up by data from 28 different machines. The anomalies are pre labeled.

how to inject artificial (Darban et al., 2024) /real anomalies. Which real world scenarios do we have? Are there anomalies in SMA data?

5.2 Results

Maybe like in Darban et al. (2024, p. 19)

6 Summary

6.1 Discussion

6.2 Future Work

References

Alshaer, M., Garcia-Rodriguez, S., & Gouy-Pailler, C. (2020, November). Detecting Anomalies from Streaming Time Series using Matrix Profile and Shapelets Learning. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 376–383). Baltimore, MD, USA: IEEE. Retrieved 2024-07-30, from `https://ieeexplore.ieee.org/document/9288261/` doi: `https://doi.org/10.1109/ICTAI50040.2020.00066`

Aota, T., Tong, L. T. T., & Okatani, T. (2023, January). Zero-shot versus Many-shot: Unsupervised Texture Anomaly Detection. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 5553–5561). Waikoloa, HI, USA: IEEE. Retrieved 2024-08-01, from `https://ieeexplore.ieee.org/document/10030870/` doi: `https://doi.org/10.1109/WACV56688.2023.00552`

Beggel, L., Kausler, B. X., Schiegg, M., Pfeiffer, M., & Bischl, B. (2019, September). Time series anomaly detection based on shapelet learning. *Computational Statistics*, *34*(3), 945–976. Retrieved 2024-07-30, from `http://link.springer.com/10.1007/s00180-018-0824-9` doi: `https://doi.org/10.1007/s00180-018-0824-9`

Bengio, Y., Courville, A., & Vincent, P. (2013, August). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. Retrieved 2024-07-03, from `http://ieeexplore.ieee.org/document/6472238/` doi: `https://doi.org/10.1109/TPAMI.2013.50`

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York: Springer.

Darban, Z. Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024, April). *CARLA: Self-supervised Contrastive Representation Learning for Time Series Anomaly Detection.* arXiv. Retrieved 2024-07-18, from `http://arxiv.org/abs/2308.09296` (arXiv:2308.09296 [cs])

Fung, C., Qiu, C., Li, A., & Rudolph, M. (2024, February). *Model Selection of Zero-shot Anomaly Detectors in the Absence of Labeled Validation Data.* arXiv. Retrieved 2024-08-01, from `http://arxiv.org/abs/2310.10461` (arXiv:2310.10461 [cs])

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* Cambridge, Massachusetts: The MIT Press.

Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., & Dubrawski, A. (2024, May). *MOMENT: A Family of Open Time-series Foundation Models.* arXiv. Retrieved 2024-07-18, from `http://arxiv.org/abs/2402.03885` (arXiv:2402.03885 [cs])

Gruhl, C. M. (2022). Novelty Detection for Multivariate Data Streams with Probalistic Models. Retrieved 2024-07-15, from `https://kobra.uni`

-kassel.de/handle/123456789/13902 (Publisher: Universität Kassel) doi: https://doi.org/10.17170/KOBRA-202205106160

Gärdenfors, P. (2000). *Conceptual spaces: the geometry of thought.* Cambridge, Mass: MIT Press.

Jeong, K.-J., & Shin, Y.-M. (2022, January). *Time-Series Anomaly Detection with Implicit Neural Representation.* arXiv. Retrieved 2024-07-30, from http://arxiv.org/abs/2201.11950 (arXiv:2201.11950 [cs])

Jiao, Y., Yang, K., Song, D., & Tao, D. (2022, May). TimeAutoAD: Autonomous Anomaly Detection With Self-Supervised Contrastive Loss for Multivariate Time Series. *IEEE Transactions on Network Science and Engineering*, *9*(3), 1604–1619. Retrieved 2024-08-01, from https://ieeexplore.ieee.org/document/9705079/ doi: https://doi.org/10.1109/TNSE.2022.3148276

Lavrač, N., Podpečan, V., & Robnik-Šikonja, M. (2021). *Representation Learning: Propositionalization and Embeddings.* Cham: Springer International Publishing. Retrieved 2024-07-09, from https://link.springer.com/10.1007/978-3-030-68817-2 doi: https://doi.org/10.1007/978-3-030-68817-2

Lee, Y., Byun, Y., & Baek, J.-G. (2023, February). Time Series Anomaly Detection Using Contrastive Learning based One-Class Classification. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (pp. 330–335). Bali, Indonesia: IEEE. Retrieved 2024-08-01, from https://ieeexplore.ieee.org/document/10067089/ doi: https://doi.org/10.1109/ICAIIC57133.2023.10067089

Li, A., Qiu, C., Kloft, M., Smyth, P., Rudolph, M., & Mandt, S. (2023, November). *Zero-Shot Anomaly Detection via Batch Normalization.* arXiv. Retrieved 2024-08-01, from http://arxiv.org/abs/2302.07849 (arXiv:2302.07849 [cs, stat])

Li, A., Zhao, Y., Qiu, C., Kloft, M., Smyth, P., Rudolph, M., & Mandt, S. (2024, June). *Anomaly Detection of Tabular Data Using LLMs.* arXiv. Retrieved 2024-08-01, from http://arxiv.org/abs/2406.16308 (arXiv:2406.16308 [cs])

Ma, Q., Liu, Z., Zheng, Z., Huang, Z., Zhu, S., Yu, Z., & Kwok, J. T. (2023, May). *A Survey on Time-Series Pre-Trained Models.* arXiv. Retrieved 2024-07-18, from http://arxiv.org/abs/2305.10716 (arXiv:2305.10716 [cs])

Ngu, H. C. V., & Lee, K. M. (2023, October). CL-TAD: A Contrastive-Learning-Based Method for Time Series Anomaly Detection. *Applied Sciences*, *13*(21), 11938. Retrieved 2024-08-01, from https://www.mdpi.com/2076-3417/13/21/11938 doi: https://doi.org/10.3390/app132111938

Nivarthi, C. P., Vogt, S., & Sick, B. (2022, December). Unified Autoencoder with Task Embeddings for Multi-Task Learning in Renewable Power Forecasting. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1530–1536). Nassau, Bahamas: IEEE. Retrieved 2024-06-27, from https://ieeexplore.ieee.org/document/

10068974/ doi: https://doi.org/10.1109/ICMLA55696.2022.00240

Nivarthi, C. P., Vogt, S., & Sick, B. (2023, September). Multi-Task Representation Learning for Renewable-Power Forecasting: A Comparative Analysis of Unified Autoencoder Variants and Task-Embedding Dimensions. *Machine Learning and Knowledge Extraction*, *5*(3), 1214–1233. Retrieved 2024-06-27, from https://www.mdpi.com/2504-4990/5/3/62 doi: https://doi.org/10.3390/make5030062

Palatucci, M., Pomerleau, D., Hinton, G. E., & Mitchell, T. M. (2009). Zero-shot Learning with Semantic Output Codes.

Pereira, J., & Silveira, M. (2019, February). Learning Representations from Healthcare Time Series Data for Unsupervised Anomaly Detection. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 1–7). Kyoto, Japan: IEEE. Retrieved 2024-07-30, from https://ieeexplore.ieee.org/document/8679157/ doi: https://doi.org/10.1109/BIGCOMP.2019.8679157

Pranavan, T., Sim, T., Ambikapathi, A., & Ramasamy, S. (2022, February). *Contrastive predictive coding for Anomaly Detection in Multi-variate Time Series Data.* arXiv. Retrieved 2024-07-30, from http://arxiv.org/abs/2202.03639 (arXiv:2202.03639 [cs])

Ramirez Rivera, A., Khan, A., Bekkouch, I. E. I., & Sheikh, T. S. (2022, January). Anomaly Detection Based on Zero-Shot Outlier Synthesis and Hierarchical Feature Distillation. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(1), 281–291. Retrieved 2024-07-30, from https://ieeexplore.ieee.org/document/9228891/ doi: https://doi.org/10.1109/TNNLS.2020.3027667

Schwartz, E., Arbelle, A., Karlinsky, L., Harary, S., Scheidegger, F., Doveh, S., & Giryes, R. (2024, February). *MAEDAY: MAE for few and zero shot AnomalY-Detection.* arXiv. Retrieved 2024-06-27, from http://arxiv.org/abs/2211.14307 (arXiv:2211.14307 [cs])

Socher, R., Ganjoo, M., Sridhar, H., Bastani, O., Manning, C. D., & Ng, A. Y. (2013, March). *Zero-Shot Learning Through Cross-Modal Transfer.* arXiv. Retrieved 2024-07-14, from http://arxiv.org/abs/1301.3666 (arXiv:1301.3666 [cs])

Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019, July). Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2828–2837). Anchorage AK USA: ACM. Retrieved 2024-08-01, from https://dl.acm.org/doi/10.1145/3292500.3330672 doi: https://doi.org/10.1145/3292500.3330672

Sun, Q., & Ge, Z. (2021, September). A Survey on Deep Learning for Data-Driven Soft Sensors. *IEEE Transactions on Industrial Informatics*, *17*(9), 5853–5866. Retrieved 2024-07-14, from https://ieeexplore.ieee.org/document/9329169/ doi: https://doi.org/10.1109/TII.2021.3053128

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. (2022, February). *TS2Vec: Towards Universal Representation of Time Series.* arXiv. Retrieved 2024-08-03, from `http://arxiv.org/abs/2106.10466` (arXiv:2106.10466 [cs])

Zhang, K., Cai, R., Zhou, C., & Liu, Y. (2024, May). Debiased Contrastive Learning for Time-Series Representation Learning and Fault Detection. *IEEE Transactions on Industrial Informatics*, *20*(5), 7641–7653. Retrieved 2024-07-09, from `https://ieeexplore.ieee.org/document/10443248/` doi: `https://doi.org/10.1109/TII.2024.3359409`

Zhou, Q., Pang, G., Tian, Y., He, S., & Chen, J. (2024, March). *AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection.* arXiv. Retrieved 2024-08-01, from `http://arxiv.org/abs/2310.18961` (arXiv:2310.18961 [cs])

7 Source Statement

I hereby declare that the content of this paper is written on my own and sources from literature are declared as such.
The use of artificial intelligence is limited to the help in understanding and summarizing the subjects and specific papers for the author. For reassuring the correctness of this paper a GPT helped in finding potential issues. None of the generated output is copied to this paper.

Kassel, 15.09.2024