



# ANOMALY DETECTION

Revision 0

## SUMMARY

This report provides an overview of the activities on ML-driven demand side anomaly detection and identification performed in the context of the AISOP project.

Website  
Contact

[aisoproject.com](http://aisoproject.com)  
[hello@aisoproject.com](mailto:hello@aisoproject.com)



**Internal Reference**

<b>Deliverable No.</b>	D3.2
<b>Deliverable Name</b>	ML-based demand-side anomaly detection and identification
<b>Lead Participant</b>	ZEDO
<b>Work Package No.</b>	3
<b>Task No. &amp; Name</b>	3.4: ML-based demand-side anomaly detection and identification
<b>Document (File)</b>	500–Deliverables-WP3_D3_2
<b>Issue (Save) Date</b>	

**Document status**

	Date	Person(s)	Organisation
<b>Authors</b>	13.11.2024	Razieh Balouchi Ulf Häger Rajkumar Palaniappan	ZEDO
<b>Verification by</b>	03.12.2024	C. Yaman Evrenosoglu	ETHZ-FEN
<b>Approval by</b>	04.12.2024	C. Yaman Evrenosoglu	ETHZ-FEN

**Document sensitivity**

- X **Not Sensitive** Contains only factual or background information; contains no new or additional analysis, recommendations or policy-relevant statements
- Moderately Sensitive** Contains some analysis or interpretation of results; contains no recommendations or policy-relevant statements
- Sensitive** Contains analysis or interpretation of results with policy-relevance and/or recommendations or policy-relevant statements
- Highly Sensitive Confidential** Contains significant analysis or interpretation of results with major policy-relevance or implications, contains extensive recommendations or policy-relevant statements, and/or contain policy-prescriptive statements. This sensitivity requires SB decision.

**Disclaimer**

The content and views expressed in this material are those of the authors and do not necessarily reflect the views or opinion of the ERA-Net SES initiative. Any reference given does not necessarily imply the endorsement by ERA-Net SES.

**About ERA-Net Smart Energy Systems**

ERA-Net Smart Energy Systems (ERA-Net SES) is a transnational joint programming platform of 30 national and regional funding partners for initiating co-creation and promoting energy system innovation. The network of owners and managers of national and regional public funding programs along the innovation chain provides a sustainable and service oriented joint programming platform to finance projects in thematic areas like Smart Power Grids, Regional and Local Energy Systems, Heating and Cooling Networks, Digital Energy and Smart Services, etc.

Co-creating with partners that help to understand the needs of relevant stakeholders, we team up with intermediaries to provide an innovation eco-system supporting consortia for research, innovation, technical development, piloting and demonstration activities. These co-operations pave the way towards implementation in real-life environments and market introduction.

Beyond that, ERA-Net SES provides a Knowledge Community, involving key demo projects and experts from all over Europe, to facilitate learning between projects and programs from the local level up to the European level.

[www.eranet-smartenergysystems.eu](http://www.eranet-smartenergysystems.eu)

# Table of Contents

- 1 Abstract..... 5
- 2 Introduction: anomalies in power grids..... 6
- 3 Background on anomaly detection ..... 8
  - 3.1 Reason for using different methods for anomaly detection ..... 8
  - 3.2 Anomaly Detection in LV Grids Using ML Methods..... 9
  - 3.3 Challenges of data and ML methods in LV grids ..... 9
- 4 Methodology and illustrative results ..... 11
  - 4.1 Seasonal-trend decomposition using LOESS (STL) method ..... 11
  - 4.2 LSTM-Autoencoder method: ..... 14
- 5 Conclusions ..... 19
- 6 References ..... 20

# List of Figures

- Figure 1 Time series with different types of anomalies and outliers ..... 6
- Figure 2. Different types of anomalies in power grids ..... 7
- Figure 3. Methods for anomaly detection..... 8
- Figure 4. Anomaly detection process using STL decomposition and Isolation forest ..... 12
- Figure 5. Anomaly detection in net power data (a) anomalies in the trend data, (b) anomalies in the residual data, (c) detected anomalies in the year 2018, (d) temperature and GS data recorded in the year 2018. .... 13
- Figure 6. Structure of the LSTM method ..... 14
- Figure 7. Structure of the LSTM-AE method..... 15
- Figure 8. Net power analysis for 2020: (a) reconstruction error of training data, (b) comparison of original and predicted net power values..... 16
- Figure 9. Net power data analysis for test scenario in 2021: (a) real versus predicted data for 2021, incorporating future scenarios, (b) anomalies detected in 2021 with anomaly scores and consideration of future scenarios. .... 17
- Figure 10. Clustered anomaly into 3 clustered based on anomalies score ..... 18
- Figure 11. Analysis of anomaly characteristics for two clusters in 2021 ..... 18

# 1 Abstract

The widespread use of electric vehicles (EVs) and electric heat pumps (EHPs), along with the increasing expansion of distributed energy sources such as photovoltaics (PVs), have significantly increased the complexity of low voltage (LV) grids. This intensification of complexity has led to challenges such as reverse power flows from sources such as PVs, EVs, and energy storage systems, which complicate traditional voltage regulation methods [1].

These developments necessitate a highly adaptive grid infrastructure capable of managing increased data volumes and ensuring stability [2]. The surge in data from advanced metering infrastructure highlights the critical need for robust anomaly detection. Anomaly detection is useful for identifying deviations from expected energy patterns, which are essential for preemptive measures against potential disruptions like equipment failures, unauthorized energy usage, and integration issues with new energy sources.

Anomaly detection in the LV grid is increasingly crucial as it enables the identification of both critical and operational anomalies that could impact consumer-side stability. These anomalies may include unexpected changes in usage or generation, unregistered energy sources, and shifts due to external factors like weather changes or significant societal events.

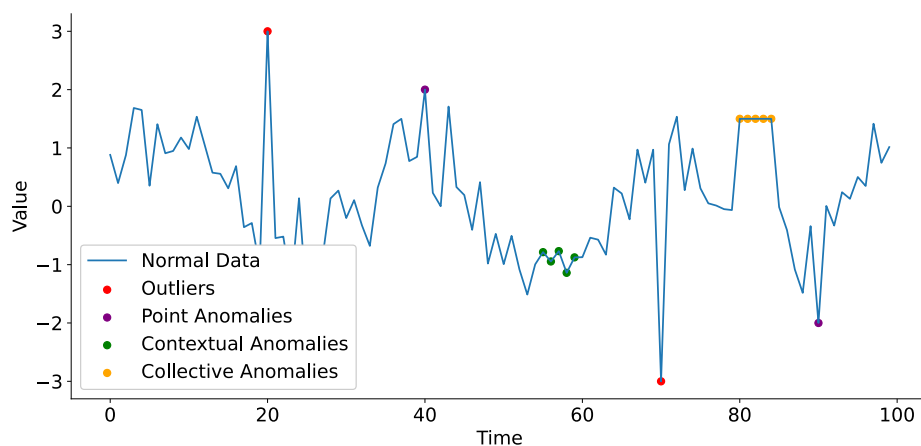
This report summarizes the application of machine learning (ML) techniques for demand-side anomaly detection within the AISOP project. Artificial intelligence (AI) methods are crucial for distribution system operators (DSOs), enabling the analysis of complex data and the identification of potential issues before they escalate. The effectiveness of these AI-driven methods depends heavily on the quality and specificity of data. Anomalies can range from significant spikes to subtle pattern shifts, with even minor variations often indicative of underlying grid sensitivities caused by factors such as weather changes. This project focuses on anomalies like newly installed PV systems and other notable data shifts, aiding DSOs by providing an alarm signal. This capability is vital for enhancing grid management and preparing for future operational demands.

## 2 Introduction: anomalies in power grids

Anomaly detection is a critical process that identifies unusual patterns in the observed data deviating from the expected behavior. These atypical patterns are known as anomalies, outliers, or exceptions. There is no absolute definition of anomaly, but there are some classic definitions of anomaly as follows [3]:

- “Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior [4].”
- “An anomaly is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism [5].”

There are different types of anomalies in data analysis: (a) **Point anomalies**, which are single data points that significantly deviate from the rest, (b) **Contextual anomalies**, which depend on the context of occurrence—for example, the energy data might be normal during the day and could be abnormal during the night, and (c) **Collective anomalies**, which involve a group of data points that are classified as anomalies in the entire dataset. Outliers, though similar to anomalies, are defined slightly differently; they are data points that significantly deviate from other observations in a dataset. In a time series, outliers may not impact the overall pattern or trend of the data. Figure 1 shows all types of anomalies and outliers.



**Figure 1 Time series with different types of anomalies and outliers**

Defining what constitutes normal behavior in data can be challenging, particularly in low-voltage (LV) grids. Different anomalies could be considered in power grids, which can be categorized into three groups, as shown in Figure 2.

- **Equipment anomalies:** Equipment anomalies occur when the physical equipment or infrastructure in the grid experiences a fault or deviates from its normal operating conditions. For example: transformer overheating due to excessive load or malfunction, circuit breaker faults, and corrosion or wear on lines or equipment.
- **Operational anomalies:** Operational anomalies refer to unexpected deviations in the functioning of the grid, often related to power flow, grid stability, or control systems. For example, voltage fluctuations due to load imbalances or equipment failures, frequency deviations, phase imbalance, failure in communication, cybersecurity attacks, congestion, and islanding.

- Consumer-side anomalies:** Consumer-side anomalies arise from irregular power consumption and/or generation or problems with end-user equipment that can affect the grid and consumer devices. One growing source of these anomalies is the increasing presence of distributed renewable energy sources, such as roof-top PV systems installed in households. For example, anomalies can be caused by factors such as unregistered PV systems, changes in consumption patterns, or meter failures. Several factors contribute to these anomalies, including shifts in weather, which impact both consumption and generation patterns, fluctuations in electricity prices, and major events like the COVID-19 pandemic or large public gatherings such as football matches. Additionally, the increased use of new technologies, such as EVs and EHPs, can significantly alter the demand on the grid.

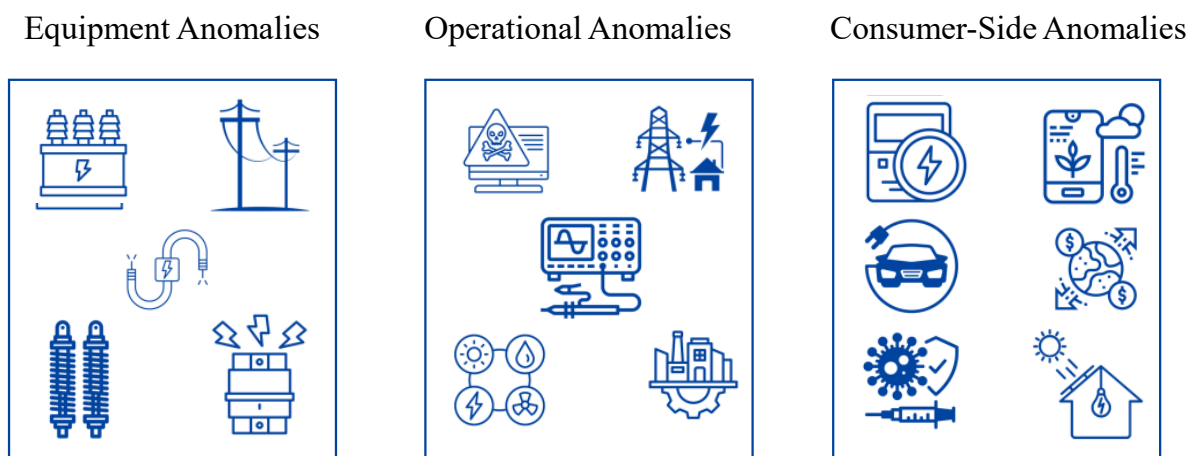


Figure 2. Different types of anomalies in power grids

The rapid adoption of EVs and EHPs, driven by efforts to reduce CO2 emissions, adds complexity to the LV grids. The increasing integration of distributed energy resources like PVs further enhances this complexity. Such developments demand a more flexible grid capable of managing significant power consumption and generation variability. Anomaly detection methods use extensive data from the meter to increase grid visibility, support real-time detection, and reduce undesired events and recognition of possible future grid problems.

The focus of the activity reported in this deliverable is on the consumer-side anomalies in LV grids. These anomalies are usually caused by irregular power consumption or generation or problems with consumers that can disrupt both households and the overall network. In the context of LV grids, the challenges are compounded by the lack of detailed metering for each household or component, which makes pinpointing the exact source of an anomaly more difficult. The lack of labeled data is another obstacle to training anomaly detection models, as most data sets in the power grid do not clearly distinguish between normal and abnormal data points.

### 3 Background on anomaly detection

There are several methods to detect anomalies in time series data in electrical grids. As shown in Figure 3, these methods can be categorized into traditional statistical methods and AI-based methods (This category includes ML and deep learning models, which are increasingly popular due to their ability to learn from data). Each of these methods offers unique advantages and can be chosen based on specific needs and the nature of the data being analyzed. For example, statistical methods involve simpler approaches, like checking if data deviates significantly from expected norms, and are quick and easy to apply but may struggle with rare or unusual data patterns. ML methods, on the other hand, learn from examples and are more effective with complex data. It has to be noticed that ML methods can make errors if trained incorrectly. Deep Learning methods are more advanced and are adept at handling data with complex or changing patterns, often finding hidden structures that other methods miss. However, deep learning models require substantial computational resources and can sometimes be difficult to interpret.

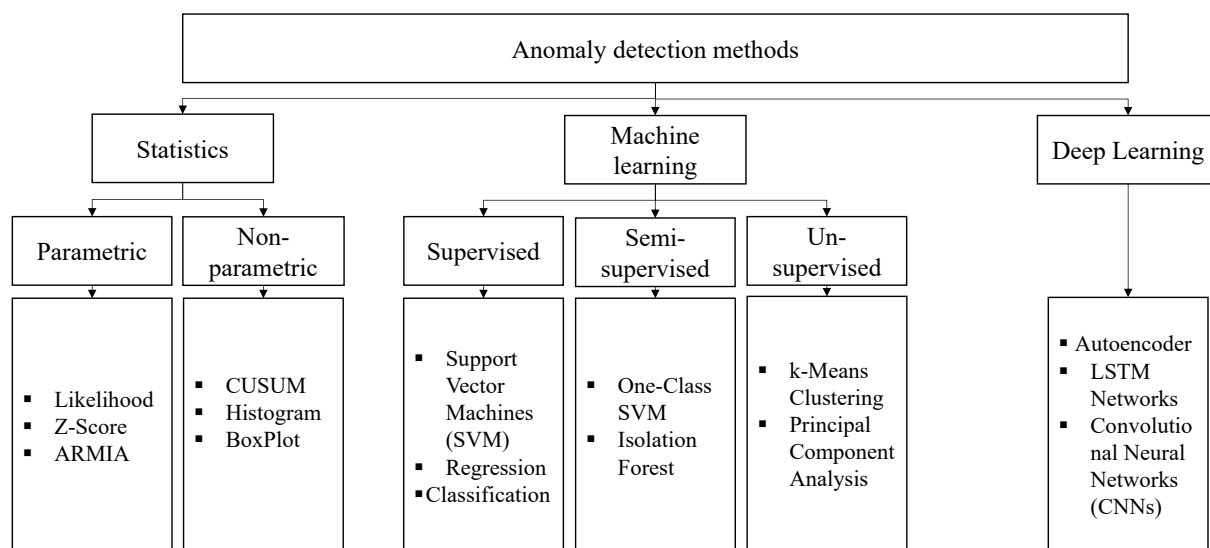


Figure 3. Methods for anomaly detection

#### 3.1 Reason for using different methods for anomaly detection

The types of anomalies in an LV grid can differ greatly in terms of their characteristics and potential impacts. For instance, some anomalies may appear suddenly because of an outage and failure or one special event, such as a planned event that changes the pattern suddenly. Others may emerge gradually, such as changes in behavior related to new loads (e.g., EVs, HPs) to the grid or shifts in the overall usage patterns.

Different ML models can process time series data in varied ways to enhance detection accuracy. For example, **Decision Trees** and **Random Forests** can handle varied data types and are good at classification tasks. **Deep Learning Models** can assimilate large volumes of data to model non-linear relationships and interactions among features, Long Short-Term Memory (LSTM) is adept at spotting anomalies in sequential data, **Clustering Algorithms** like K-Means could be deployed to identify groups of similar behavior patterns where deviations from these clusters signal potential issues.



Each dataset comes with its own set of challenges, including but not limited to its distribution, dimensionality, and the level of noise it contains. A detection method that performs exceptionally well on one dataset might not yield the same results on another due to these variations.

Different anomaly detection methods are superior in various fields, for example:

- Some might offer high precision, which means finding a high rate of true positives. Others might excel in recall, which means they are good at catching as many anomalies as possible.
- Some methods are particularly robust against noise and missing data, which is advantageous in messy real data, such as the isolation forest method, which is highly effective in datasets with a lot of noise and can handle missing data. Others offer adaptability, learning, and adjusting to new patterns as data evolves, such as the LSTM method, which can adjust to new patterns over time due to its ability to "remember" long-term dependencies, making it ideal for dynamic environments.

## 3.2 Anomaly Detection in LV Grids Using ML Methods

Detecting anomalies with ML methods depends heavily on the data quality and the specific goals of the analysis. Due to the unpredictable nature of power consumption and PV generation, training data might include small, hidden anomalies. Real data from LV grids often contain uncertainties, including minor anomalies that are not immediately obvious. Furthermore, the measurement data may not be available for every sampling interval, and in some cases, the data used may be generated as pseudo-values rather than actual measurements. Over time, the patterns in this data can change, making older training data less effective.

Additionally, incorporating extra information, such as weather data, can further complicate the modeling process. Weather data, especially critical variables like global radiation, may not be available for the locations closest to the PV panels. The use of such incomplete or low-quality weather data as an additional input to ML models can, in some cases, increase errors instead of improving model accuracy. Different ML methods vary in how they handle inputs, outputs, and parameter settings. Some methods are more flexible and adapt better to changes in grid data, which is constantly evolving due to varying consumption and generation patterns.

## 3.3 Challenges of data and ML methods in LV grids

The real LV grid data poses several challenges for ML methods. While real data is valuable for developing accurate ML models, it also presents several challenges:

1. **High Fluctuations and Inconsistent Patterns:** The data shows high levels of fluctuation, and the patterns are not consistent throughout the years. This variability makes it difficult for ML models to capture clear trends or learn stable patterns, leading to potential inaccuracies in predictions.
2. **Presence of Anomalies and Noise:** The dataset contains various types of anomalies. Some of these appear as clear outliers and can be removed during the preprocessing stage. However, smaller, more subtle anomalies are harder to detect and may remain in the training data. As a result, the models are sometimes trained with noisy data that includes uncertainties, anomalies, and errors, which can reduce the model's overall accuracy.
3. **Black-Box Nature of ML Models:** Many ML models, especially deep learning methods, function as "black boxes," meaning that it is difficult to explain why certain anomalies are detected and other are not. When the training data contains noise and anomalies, this issue becomes more pronounced. In these cases, the

model may detect unclear or unexplained anomalies, making it hard to provide a clear justification for why these points were flagged.

4. **Inaccurate or Incomplete External Data:** In some cases, additional information, such as weather data (e.g., temperature, wind speed, and global radiation), is used to improve anomaly detection. However, this data is not always completely accurate. For instance, the weather data might come from a station distant from the LV grid measurements, which is problematic for applications like PV generation, where local conditions such as shading and cloud cover are critical. This lack of precise external data can make it difficult to explain some anomalies.
5. **Sensitivity of ML Methods:** Different ML models have different levels of sensitivity to anomalies. This means that some models may detect certain points in the data as anomalies, while others may not. This variation can make it harder to interpret the results. In some cases, anomalies can be explained by considering external factors like weather conditions, historical patterns, or daily/weekly trends. In contrast, in other cases, the detected anomalies remain unclear due to the noise and errors in the training data.

## 4 Methodology and illustrative results

The data used in this study comes from a real LV grid by a DSO in Germany. It measures net power from the secondary side of the transformer, including both consumption and generation data. Additional factors like weather, weekdays, and holidays are also included. Data was taken from two LV stations, selected from over 100 stations because they showed higher fluctuations. The data is recorded every 15 minutes.

The study explains two methods for detecting anomalies. The seasonal trend method, published in a conference paper [6], shows results for data from an urban station, while the LSTM Autoencoder (AE) method, also detailed in a conference paper [7], presents results from a rural station.

### 4.1 Seasonal-trend decomposition using LOESS (STL) method

In this project, a combination of statistical decomposition and ML methods is used to analyze the high fluctuations in real LV grid data. The first step involves decomposing the time series data using the **STL** method. This technique decomposes the data into three key components: **trend**, **seasonality**, and **residual**.

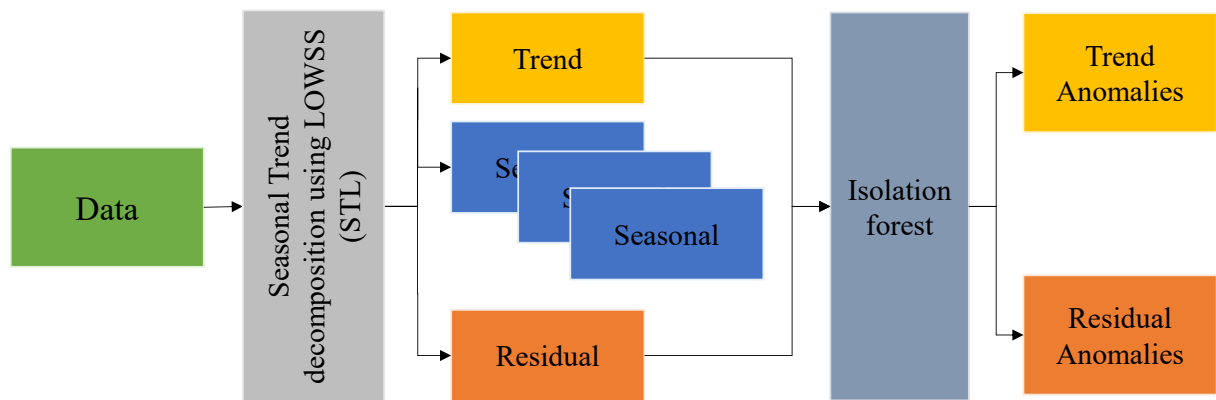
- **Trend:** This component captures the long-term movement in the data, smoothing out short-term fluctuations. It provides a clear view of the overall direction of the grid behavior, which is useful for understanding long-term changes, identifying significant shifts in data patterns, and providing actionable insights.
- **Seasonality:** This captures recurring patterns, such as daily or weekly cycles, which can help identify predictable changes in the grid, like regular demand peaks. It adjusts for periodic changes, revealing anomalies that might be hidden due to regular fluctuations.
- **Residual:** The residual component shows the short-term, irregular fluctuations that aren't explained by the trend or seasonality. It highlights sudden or unexpected changes in the grid's behavior, which may indicate anomalies.

Seasonal-trend decomposition is particularly effective for time series data like that from LV grids because it [8]:

1. manages seasonal fluctuations and sudden changes in trends,
2. stays robust even when anomalies are present in the data, and
3. handles long seasonal periods effectively.

For detecting anomalies, the focus is on the **trend component** for long-term anomalies and the **residual** component for short-term anomalies. Anomalies in the trend refer to gradual changes in the grid's behavior over time, while anomalies in the residual signal sudden and unexpected shifts.

To detect anomalies, an ML model like **Isolation Forest** is applied, as it is shown in Figure 4. The model takes the trend and residual components as input to learn the normal patterns in the data and identify outliers that signal anomalies. This combination of STL and ML (Isolation Forest) provides a method to detect both long-term and short-term anomalies in the LV grid data.



**Figure 4. Anomaly detection process using STL decomposition and Isolation Forest**

Figure 5 presents the results of anomaly detection applied to both the trend and residual data. Figure 5(a) highlights significant anomalies in the trend component, such as meter failures in October 2019, shifts due to the COVID-19 pandemic in April 2020, and changes observed in an August 2022 scenario. Figure 5(b) focuses on anomalies in the residuals, which capture minor variations in the data. For instance, Figure 5(c) illustrates anomalies detected in both the trend and residual components from 2018, with identified causes. Residual anomalies on January 1st are linked to distinct New Year's Day energy usage patterns, while trend anomalies in early March correspond to a notable increase in power consumption. Figure 5(d) further explores these variations by comparing them with global solar radiation and temperature data. A significant temperature drop to around -10°C was recorded during that week. While global radiation levels were higher, possible cloudy conditions may have reduced PV generation, contributing to the observed anomalies in power consumption.

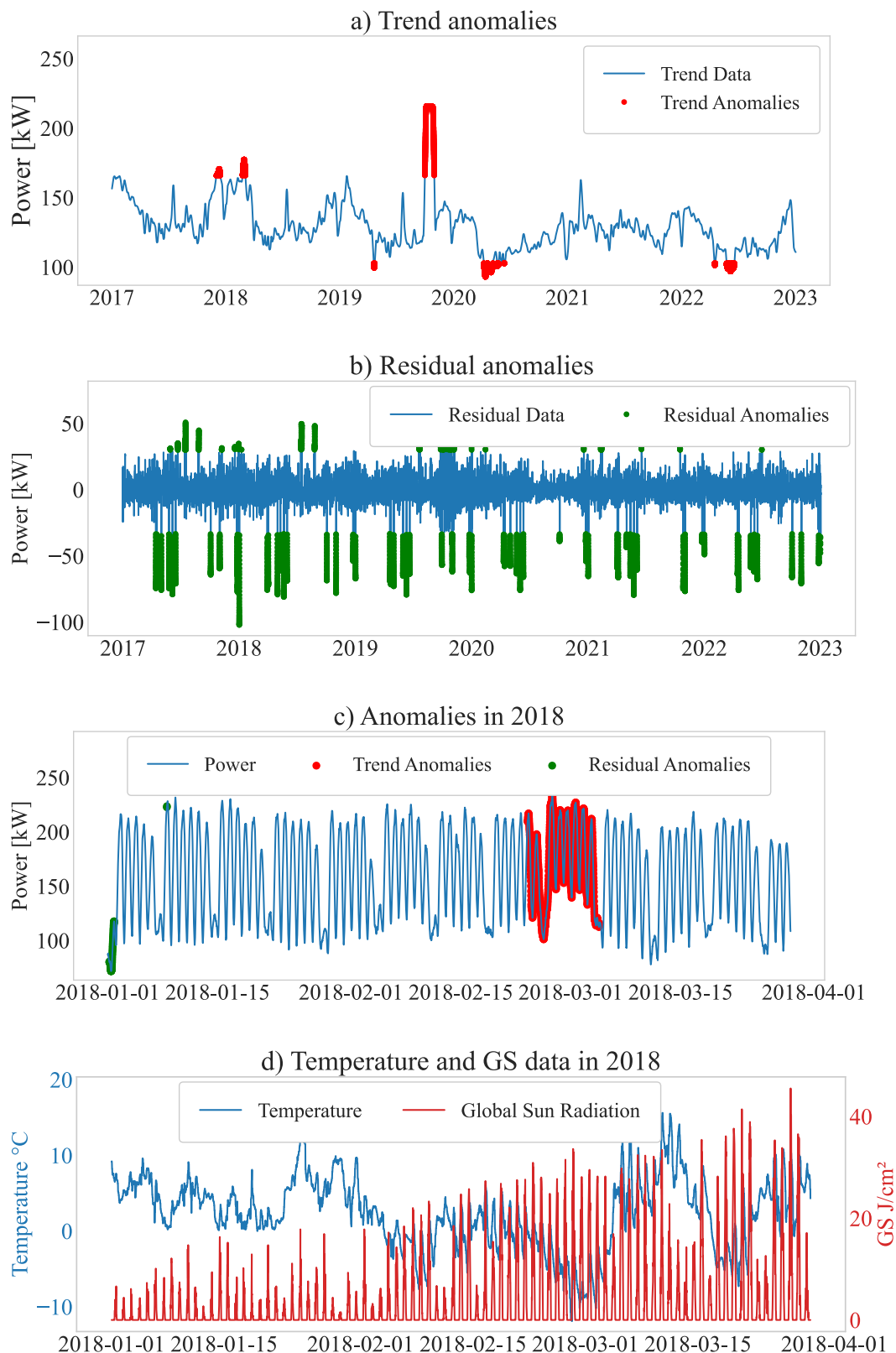


Figure 5. Anomaly detection in net power data (a) anomalies in the trend data, (b) anomalies in the residual data, (c) detected anomalies in the year 2018, (d) temperature and GS data recorded in the year 2018.

## 4.2 LSTM-Autoencoder method:

The LSTM-AE is a hybrid approach that combines LSTM networks with AE architecture to model sequential data for anomaly detection. The proposed LSTM-AE acts as a stochastic process model, capturing the normal behavior of the data and allowing the prediction of the probability distribution of the data. Any deviation between the observed data and the predicted probability density distribution is identified as a potential error, fault, or anomaly. LSTM networks are an extension of Recurrent Neural Networks (RNN) designed to overcome the limitations of RNN in retaining important information in long sequences. LSTMs solve this problem by introducing memory cells with feedback connections, enabling the network to keep or discard information over longer periods selectively [9].

As shown in Figure 6 each LSTM cell is composed of three key gates: the input gate, which determines what new information should be added to the memory; the forget gate, which decides what information to discard; and the output gate, which controls what part of the memory is used to produce the output. Through these gates, the network can maintain and update the cell’s state over time, allowing it to process entire sequences of data effectively.

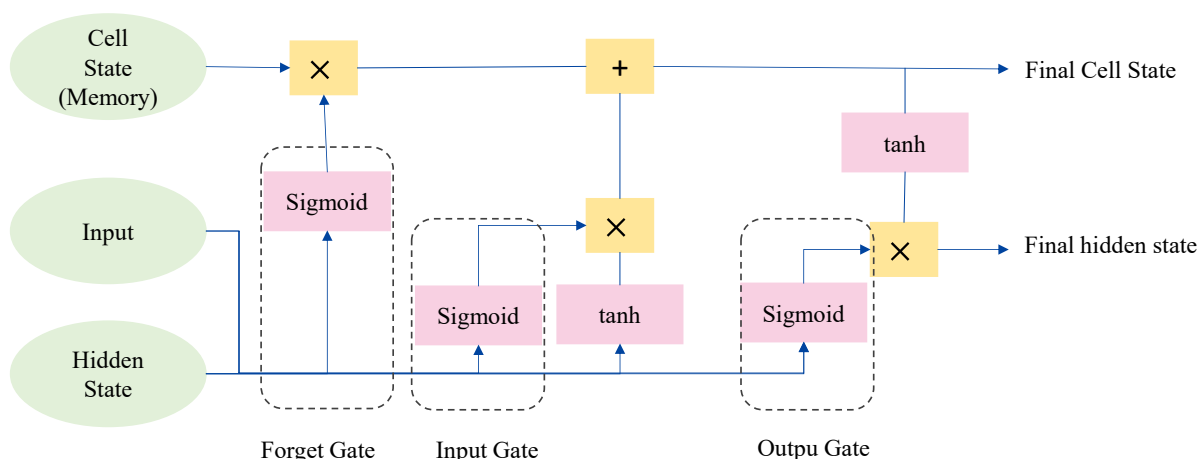


Figure 6. Structure of the LSTM method

AEs, on the other hand, are neural networks designed to learn efficient representations of data by compressing high-dimensional input into a lower-dimensional space (encoding) and then reconstructing the input (decoding) from this compressed form. The encoder module captures the essential features of the data in a compact representation while the decoder reconstructs the original data, aiming to minimize reconstruction error. This process also helps to reduce the noise in the data. AEs are especially popular in anomaly detection tasks because the model is trained to represent normal data patterns.

The LSTM-AE first uses its LSTM architecture to learn temporal dependencies within the data and capture short-term and long-term relationships. The encoder then compresses these relations into a dimensionally reduced representation while the decoder reconstructs the original sequence from this compressed form. The reconstruction error is used as an indicator of normal or abnormal behavior, where large errors indicate the presence of errors or anomalies in the data.

Figure 7 shows the structure of the LSTM-AE method used in this project. The model consists of 4 layers:

1. **Input Layer:** Takes in sequential data, with each sequence representing seven days of energy usage. This sequence length was chosen to capture weekly patterns and potential anomalies while maintaining manageable complexity. A sliding window with a stride of one day ensures that each day’s data is included in multiple sequences, providing comprehensive coverage and overlap for effective model training.
2. **Structure Layer:** The structure of the LSTM-AE consists of two LSTM layers followed by a hidden layer in the encoder. The first LSTM layer processes the input sequence, transforming each time step into a 128-dimensional vector to capture temporal dependencies and key features. This output is then passed to the second LSTM layer, which further compresses the 128-dimensional vectors into 64-dimensional representations, distilling the essential features from the data. The final hidden layer further compresses the sequence, potentially utilizing techniques like averaging or attention mechanisms. The decoder part of the AE then reconstructs the original input sequence from this compressed representation.
3. **Output Layer:** The output layer of the LSTM-AE ensures that the decoder’s output matches the dimensions of the original input sequence, effectively reconstructing the data from its compressed, encoded state. This ability to reduce the data’s dimensionality and then accurately reconstruct it makes the LSTM-AE especially useful for tasks like dimensionality reduction, feature extraction, and anomaly detection in complex datasets.
4. **Analysis Layer:** After training, the LSTM-AE is applied to predict test data. Anomalies are detected by calculating the reconstruction loss, which is the difference between the actual data and the reconstructed data from the model. A threshold is set based on the distribution of these losses, and data points with a reconstruction loss above this threshold are flagged as anomalies. To further analyze these anomalies and understand potential causes, the anomalies are clustered, and the features of each cluster are examined. These features might include factors like temperature, global radiation, or other relevant variables that could explain the reason for the anomalies.

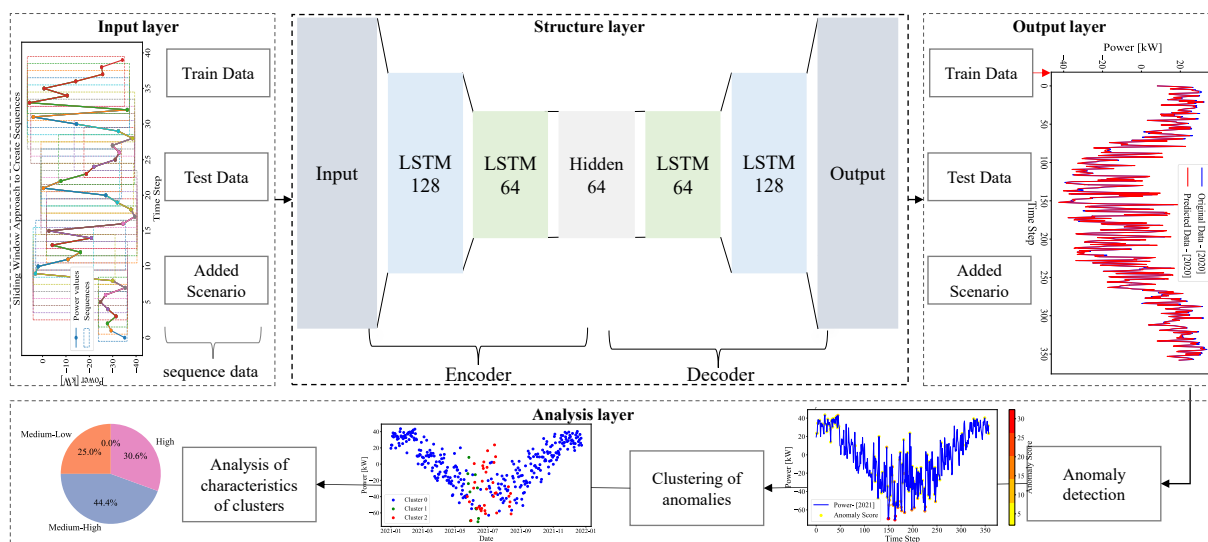
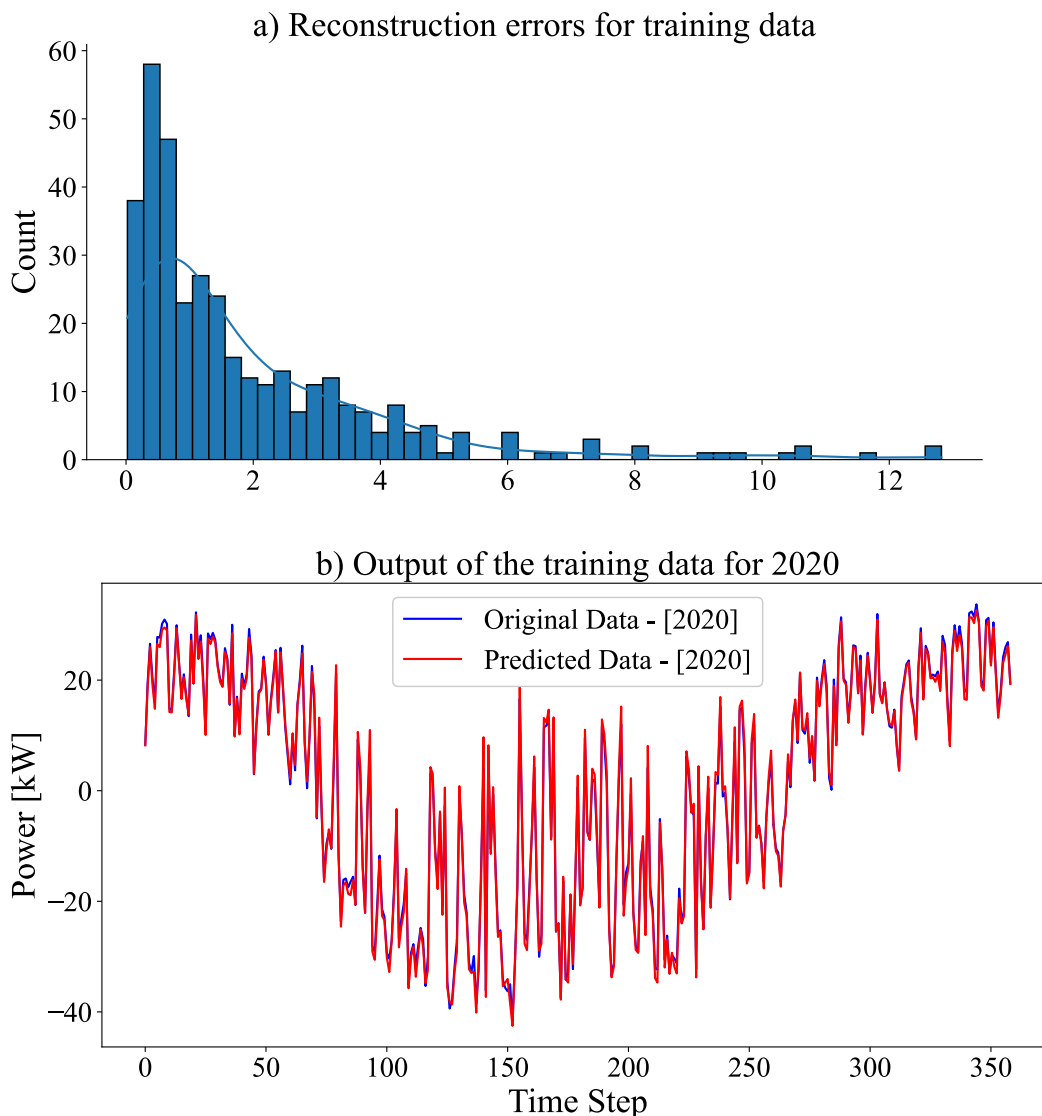


Figure 7. Structure of the LSTM-AE method

Data from 2017 to 2019 are treated as historical data to train the model, while data from 2020 are used as test data to evaluate the effectiveness of the algorithm. To simulate future challenges, scenarios with additional PV systems and additional EHPs were introduced into the 2021 data. This scenario-based data serves as an evaluation set to test how well the algorithm copes with potential future grid conditions.

Figure 8 shows the reconstruction errors and the method's output applied to the 2020 training data. The relatively low reconstruction errors suggest that the method has successfully captured the underlying patterns in the data, leading to accurate predictions and strong training performance.



**Figure 8. Net power analysis for 2020: (a) reconstruction error of training data, (b) comparison of original and predicted net power values.**

Figure 9(a) displays the net power for the year 2021 data, including the added future scenario. The trend for 2021 differs from previous years, reflecting the presence of additional PV installations. While the model's performance in 2021 is slightly weaker than in prior years, it still effectively tracks the trend during most non-summer months. Figure 9(b) highlights the anomalies detected in the year 2021 data, with high-scoring anomalies shown in red and lower-scoring ones in yellow. The result illustrates the method's capability to identify the impact of added PVs and EHPs in the grid data.



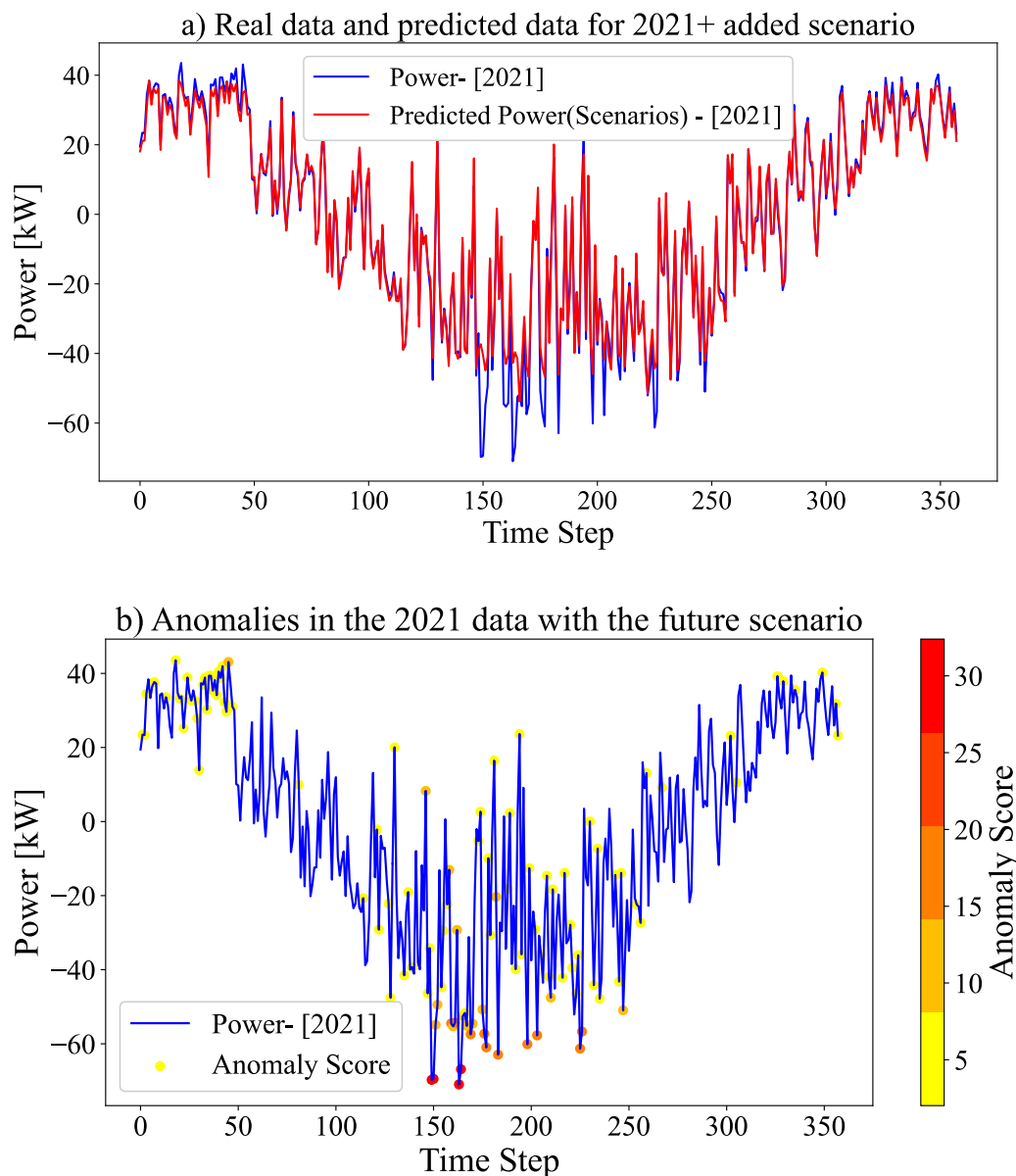


Figure 9. Net power data analysis for test scenario in 2021: (a) real versus predicted data for 2021, incorporating future scenarios, (b) anomalies detected in 2021 with anomaly scores and consideration of future scenarios.

Figure 10 displays anomalies grouped into three clusters, categorized by an anomaly score. The goal is to uncover the root causes of these anomalies. While clustering could be based on different features, in this example, the clustering is done using the anomaly score.

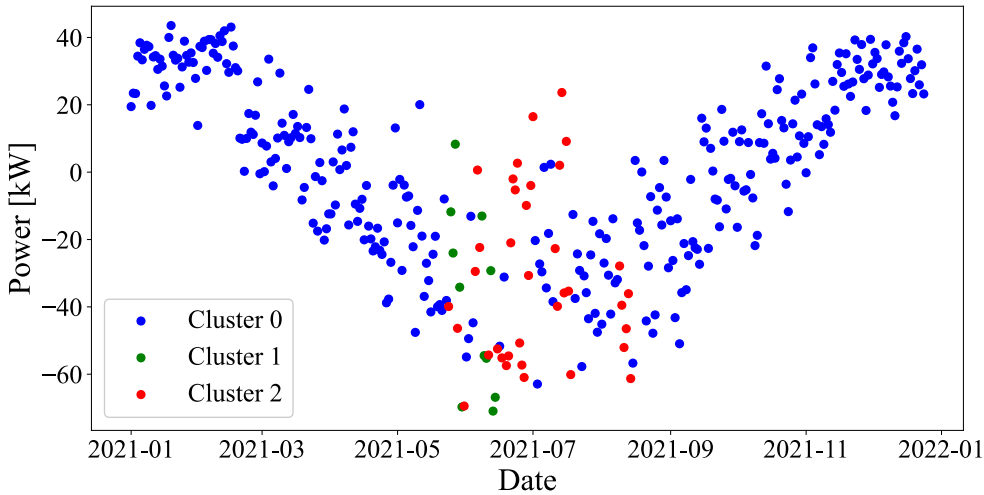


Figure 10. Clustered anomaly into 3 clustered based on anomalies score

Figure 11 highlights some key features within each cluster. For instance, Cluster 0 reveals that the majority of anomalies occur under conditions of medium-low global radiation (GS) (53%) and moderate-to-low temperatures (47.9%) and are concentrated during colder months (October to February, 56.3%). This pattern suggests that these anomalies are likely driven by an increase in energy consumption from EHPs. In contrast, Cluster 2 exhibits a high frequency of anomalies in situations involving medium-high to high GS, often during higher temperatures. These anomalies tend to appear in the warmer months (May through August), suggesting that they may be linked to extra PV generation.

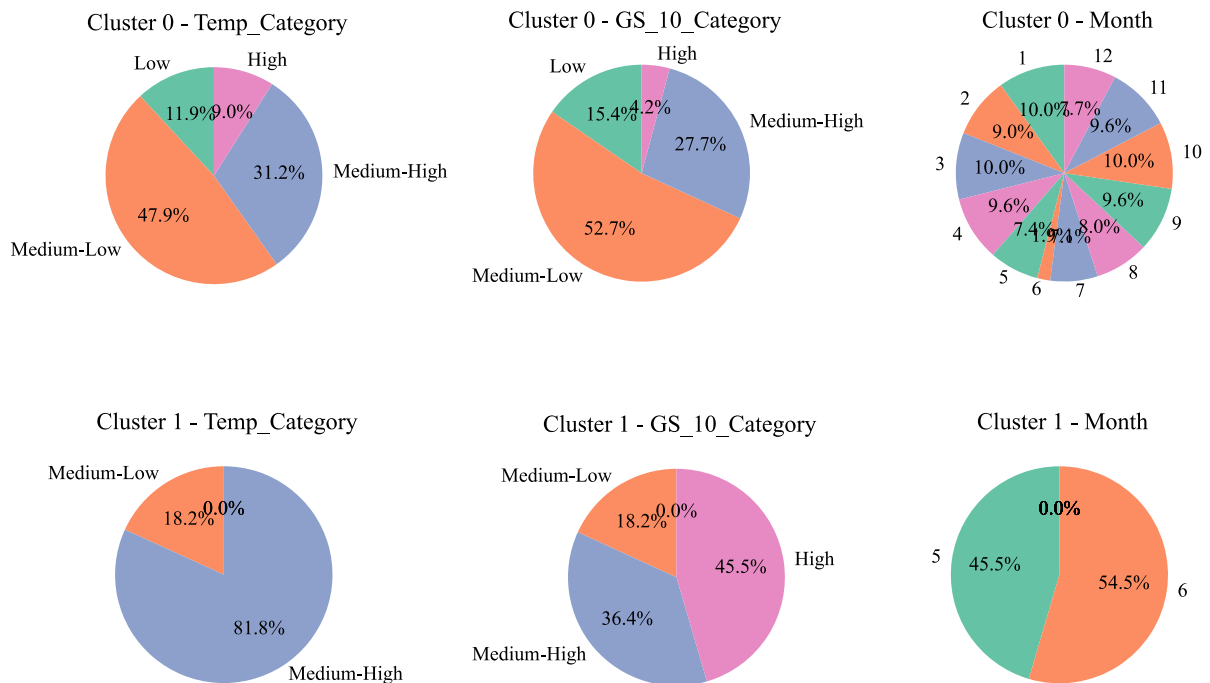


Figure 11. Analysis of anomaly characteristics for two clusters in 2021

## 5 Conclusions

In conclusion, as the share of PVs and new loads in the grid increases, the grid faces new challenges, including shifts in consumption and generation patterns due to changes in consumer behavior or sudden events like temperature fluctuations. DSOs must stay aware of these unexpected patterns and new behaviors to make better decisions for the present situation and to prepare for the future of the grid.

AI-driven anomaly detection methods play a key role in assisting DSOs by analyzing data and signaling potential issues in undesirable situations. Although there are multiple anomaly detection techniques available, it is essential to evaluate the potential of each method, understand the specific objectives of anomaly detection, and address the challenges related to available data. The performance of AI methods heavily depends on the quality and nature of the data, and it is important to note that anomalies from the consumer side are not limited to extreme spikes or drops (and not just outliers). Even small changes in data patterns can indicate anomalies, signaling potential sensitivity in the grid due to minor factors like weather changes.

In this project, detecting newly installed PVs and other significant shifts in data patterns is particularly valuable for DSOs. Whether dealing with large or small anomalies, the key takeaway is that anomaly detection helps operators identify and respond to grid changes, allowing for more effective grid management and improved preparedness for future challenges. [4]

## 6 References

- [1] H. Hatta, M. Asari, and H. Kobayashi, "Study of energy management for decreasing reverse power flow from photovoltaic power systems," in *2009 IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE)*, pp. 1–5.
- [2] A. Anwar and A. N. Mahmood, "Anomaly detection in electric network database of smart grid: Graph matching approach," *Electric Power Systems Research*, vol. 133, pp. 51–62, 2016.
- [3] D. Samariya and A. Thakkar, "A comprehensive survey of anomaly detection algorithms," *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, 2023.
- [4] V. Chandola, A. Banerjee, and V. Kumar, "Outlier detection: A survey," *ACM Computing Surveys*, vol. 14, p. 15, 2007.
- [5] D. M. Hawkins, *Identification of Outliers*, London, U.K.: Chapman & Hall, 1980.
- [6] R. Balouchi Anaraki, R. Palaniappan, U. Häger, C. Rehtanz, "Data-driven Approaches for Anomaly Detection in Low-Voltage Grid Net Power," in *Proc. of IEEE/PES ISGT Europe, Dubrovnik, 2024*.
- [7] R. Balouchi Anaraki, R. Palaniappan, U. Häger, C. Rehtanz, "Anomaly Detection in Low-Voltage Grids with LSTM Autoencoders: A Study on Future Scenario Impacts," in *Proc. of IEEE/PES ISGT Europe, Dubrovnik, 2024*.
- [8] *RobustSTL: A robust seasonal-trend decomposition algorithm for long time series*, 2019.
- [9] P. Mobtahej *et al.*, "An LSTM-Autoencoder Architecture for Anomaly Detection Applied on Compressors Audio Data," *Computational and Mathematical Methods*, vol. 2022, no. 1, p. 3622426, 2022.

## FUNDING



This project has received funding in the framework of the joint programming initiative ERA-Net Smart Energy Systems' focus initiative Digital Transformation for the Energy Transition, with support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883973.