energies

MDPI

Article

AIoT-Based Visual Anomaly Detection in Photovoltaic Sequence Data via Sequence Learning

Qian Wei [1,2,*], Hongjun Sun [1], Jingjing Fan [2], Guojun Li [3] and Zhiguang Zhou [3]

[1] Department of Intelligent Science and Technology, College of Artificial Intelligence, China University of Petroleum, Beijing 102249, China; sunhj68@cup.edu.cn
[2] Kunlun Digital Technology Co., Ltd., Beijing 102206, China; fanjingjing@cnpc.com.cn
[3] Big Data Visualization and Human Computer Collaborative Intelligent Laboratory, School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018, China; guojunli@hdu.edu.cn (G.L.); zhgzhou@hdu.edu.cn (Z.Z.)
* Correspondence: 2022345703@student.cup.edu.cn; Tel.: +86-18911730248

Abstract: Anomaly detection is a common analytical task aimed at identifying rare cases that differ from the majority of typical cases in a dataset. In the management of photovoltaic (PV) power generation systems, it is essential for electric power companies to effectively detect anomalies in PV sequence data, as this helps operators and experts understand and interpret anomalies within PV arrays when making response decisions. However, traditional methods that rely on manual labor and regular data collection are difficult to monitor in real time, resulting in delays in fault detection and localization. Traditional machine learning algorithms are slow and cumbersome in processing data, which affects the operational safety of PV plants. In this paper, we propose a visual analytic approach for detecting and exploring anomalous sequences in a PV sequence dataset via sequence learning. We first compare the sequences with their reconstructions through an unsupervised anomaly detection algorithm (Long Short-Term Memory) based on AutoEncoders to identify anomalies. To further enhance the accuracy of anomaly detection, we integrate the artificial intelligence of things (AIoT) technology with a strict time synchronization data collection and real-time processing algorithm. This integration ensures that data from multiple sensors are synchronized and processed in real time. Then, we analyze the characteristics of the anomalies based on the visual comparison of different PV sequences and explore the potential correlation factors to analyze the possible causes of the anomalies. Case studies based on authentic enterprise datasets demonstrate the effectiveness of our method in the anomaly detection and exploration of PV sequence data.

Keywords: photovoltaic sequence data; sequence learning; artificial intelligence of things; anomaly detection; visual exploration

check for updates

Citation: Wei, Q.; Sun, H.; Fan, J.; Li, G.; Zhou, Z. AIoT-Based Visual Anomaly Detection in Photovoltaic Sequence Data via Sequence Learning. *Energies* **2024**, *17*, 5369. https://doi.org/10.3390/en17215369

Academic Editor: Francesco Nocera

Received: 13 September 2024
Revised: 13 October 2024
Accepted: 18 October 2024
Published: 29 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Photovoltaic (PV) power generation has been studied and applied worldwide due to its green, clean, non-polluting, and sustainable characteristics. The PV inverter, an important component of a PV power system, is worth an in-depth study as its output power sequence data express the time-based PV conversion capacity well [1]. However, owing to the effects of sunlight intensity, temperature, and other objective factors, the output sequence data are subject to great uncertainty and fluctuation. Thus, large amounts of anomalies in the PV array sequence data emerge and seriously affect the management planning of power generation equipment and the stable operation of the grid system. To satisfy the requirements for intelligent management of PV power generation, it is necessary for electric power companies to develop a system and algorithm for detecting anomalies in PV sequence data. This integration ensures that data from multiple sensors are synchronized and processed in real time, which can help managers to take the appropriate measures to

avoid the occurrence of faults or accidents as well as reduce the economic impact on the power station [2,3].

To detect anomalies, different techniques have been proposed which can be divided into three categories. Manual simulation models [4] based on physical principles are generally created by domain experts. However, with the rapid growth of PV data quantity, these methods are time-consuming and labor-intensive to the extent that anomaly detection tasks cannot be completed in a timely and efficient manner. The detection performance of traditional machine learning models [5] is not ideal for time sequence data with complex data structures despite its ability to improve detection accuracy. Moreover, such models are sensitive to missing data. In contrast, deep learning models [6] are more powerful in handling complex data, as they are capable of learning clearer structural features in PV sequence data and enabling more accurate anomaly detection, and they are also more robust and fault tolerant to the data. Moreover, the addition of visual analytics can assist users in incorporating their domain knowledge into the analysis process so that they can quickly and accurately detect anomalous events and further explore such events in PV sequence data. Thus, it is necessary to integrate deep learning models and visualization techniques.

However, we are facing the following three main challenges in detecting the anomalies of PV sequence data: **CH1**. For unlabeled PV sequence data, the boundary between normal and anomalous data is not precisely defined, and anomalies are difficult to detect; **CH2**. Another challenge is to explore and explain the causes of anomalies in the PV array due to the blackbox effect of the deep learning method; **CH3**. An interactive visual analytics system needs to be developed to combine deep learning and visual analytics techniques for anomaly detection in PV sequence data.

To address these challenges, we systematically initiated foundational research preparations, which encompassed a comprehensive review of the related work, a detailed requirements analysis, and the systematic design of the system overview. Subsequently, we performed real-time data collection based on the AIoT system and further validated the LSTM-AE model, visualization design, and interactive functionalities in the context of anomaly detection in PV power generation. Following this, we developed an AIoT-based PV visual anomaly detection system and conducted case studies based on authentic enterprise datasets to verify the system's effectiveness in the visual exploration and analysis of anomalous events in PV arrays. Finally, we summarized the interim achievements and value of this research and delineated directions for future research. Our main innovations are highlighted as follows:

- An LSTM-AE model is leveraged to detect anomalies in PV sequence data with complex data structures and in the absence of anomaly labels, characterized by a significant difference between the original data and the predicted data.
- A set of visualization and interaction cues are designed to integrate with the model, thereby enabling interpretable detection of anomalies in the PV array.
- Case studies based on the AIoT system, employing strict time synchronization for data collection and real-time processing algorithms, are conducted to achieve real-time collection and monitoring of massive sequences of data from PV arrays.

2. Related Work

We have classified the existing methods into four categories, namely AIoT-Based PV Sequence Data Collection, Cloud-Based PV Sequence Data Analysis, Anomaly Detection in Sequence Data, and Visualization of Anomaly Detection.

2.1. AIoT-Based PV Sequence Data Collection

AIoT refers to the combination of Internet of Things (IoT) technology with artificial intelligence (AI). A PV power station based on an AIoT system realizes real-time collection and monitoring of massive sequences of data from PV arrays, establishing a reliable connection and intelligent management from the field equipment at the power station to the cloud, as is shown in Figure 1. First, using advanced sensors and software acquisition adaptation

technologies, it integrates multiple standard collection protocols to efficiently collect and dynamically monitor the operational data of PV arrays, including key parameters such as temperature, light intensity, current, and voltage. Second, IoT data preprocessing algorithms are utilized to remove noise, correct errors, and fill in missing values, providing accurate foundational data for PV sequence data analysis and anomaly detection. Then, these data are transmitted through wireless or wired networks to a cloud computing center for centralized governance and value extraction. Additionally, the AIoT system provides edge computing services and intelligent models, enabling real-time inference and local autonomy close to the data source while reducing transmission latency and bandwidth demand [7].

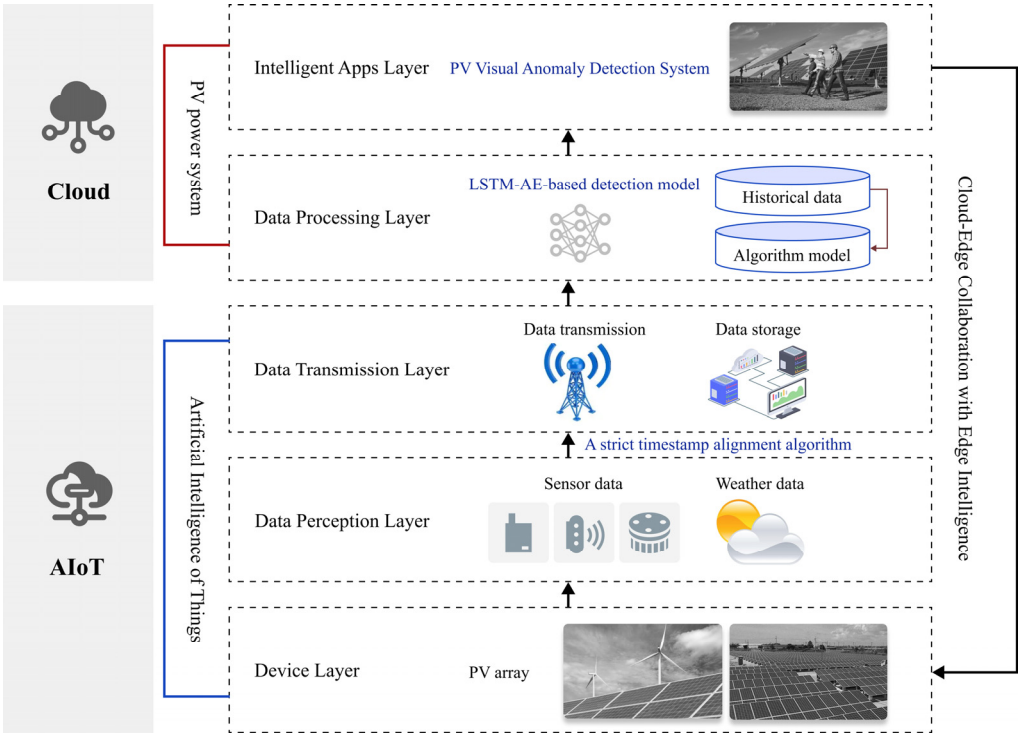


Figure 1. The Architecture of AIoT-Based PV Visual Anomaly Detection System.

Among these, in the data perception layer depicted in Figure 1, we employ multiple sensors to collect the same physical quantity and integrate data from different sensors using data fusion algorithms to enhance the accuracy and robustness of the data. Additionally, we implement adaptive algorithms that enable edge devices to dynamically adjust their processing strategies based on real-time load conditions, thereby augmenting the flexibility and responsiveness of the system. The selected sensors must possess self-diagnosis capabilities, allowing them to periodically assess their own health status and promptly report any potential failures. In the data transmission layer, we utilize encryption technology to ensure the security of data during transmission, safeguarding against interception or tampering. In the data processing layer, we leverage distributed storage solutions to enhance the reliability and scalability of data storage, as well as to improve the accuracy of sequence data analysis for PV arrays under various sensor modes, avoid data distortion, and enhance the speed of system fault localization and response. The AIoT system adopts a strict timestamp alignment algorithm during the data collection phase, which includes the following four steps:

Step 1. Hard synchronization of multi-sensor devices using time synchronization protocols (such as PTP or NTP) and thus regularly calibrating time offsets to ensure all the sensors' timestamps align at the same time point.

Step 2. Establishing data collection buffers and transmission delay mechanisms by employing interpolation techniques to ensure that they are sorted and soft-aligned at the same timestamp.

Step 3. Using real-time stream analysis algorithms and sliding windows in real-time anomaly detection analysis; when anomalies are detected, the system automatically adjusts its sampling frequency.

Step 4. By establishing a simulation debugging process, the system ensures that data from multiple sensors are precisely aligned and integrated on the same timeline, providing a solid foundation for subsequent comparative analysis and anomaly detection, thereby enhancing the overall monitoring and management efficiency of the PV visual anomaly detection system based on the AIoT system.

AIoT features a large capacity, high concurrency, and high reliability [8], as is shown in Figure 2. It supports the simultaneous connection and efficient data processing of a large number of devices, complemented by edge intelligence, thereby achieving intelligent interconnectivity among devices and enhancing their autonomous sensing, decision-making, and control capabilities [9].

Statistics

Expand all groups | Collapse all groups

Request	Executions					Response Time (ms)							
	Total	OK	KO	% KO	Req/s	Min	50th pct	75th pct	95th pct	99th pct	Max	Mean	Std Dev
Global Information	1,500,000	1,500,000	0	0%	34,090.909	1	54	75	275	638	942	79	114
Publish	1,500,000	1,500,000	0	0%	34,090.909	1	54	75	275	638	942	79	114

Figure 2. The results of the AIoT High Concurrency Performance Testing. This figure illustrates the success rate and response time analysis for each request, including the minimum, maximum, average response times, and standard deviation for the same request. When simulating 1.5 million concurrent real-time data requests, the success rate of each request exceeded 99.99%, and the average TPS reached 34,090.909 transactions per second. The minimum response time at this point is 1 ms, the maximum response time is 942 ms, and 99% of the response times are less than 638 ms, which means that 99% of the requests were successfully responded to within 638 ms. The request values counted here include successful requests and do not include failed requests.

2.2. Cloud-Based PV Sequence Data Analysis

The cloud computing center will further analyze the historical PV sequence data provided by the AIoT system. It focuses on the following three main areas: objective factors affecting generation, characteristics of the sequence data, and PV generation forecasting. Solar energy, as a special energy source, is largely influenced by objective factors and also affects the output of PV power generation. Mehmood et al. [10] processed a new centralized cloud-based solar conversion recovery system (SCRS). The proposed system utilizes the Internet of Things (IoT) and cloud-based centralized architecture, which allows users to remotely monitor the amount of soiling on PV panels, regardless of the scale. Chen et al. [11] constructed a PV output prediction model based on irradiance and the component's temperature. In addition, they analyzed the effect of dust particles in the air on PV power generation in an inverse relationship. Sawadogo et al. [12] analyzed the effect of temperature on PV power generation based on PV sequence data from West Africa and concluded that PV power generation is highest in summer and lowest in winter.

With the influence of the objective environment, the frequency of PV power generation changes constantly and shows periodicity, randomness, volatility, and instability thus bringing great challenges to the safe and stable operation and effective dispatch of the power grid. Experts have used PV sequence data to achieve predictions of PV power based on different methods such as physical modeling methods, statistical methods and artificial intelligence methods, as well as other innovative algorithms, respectively. Other innovative methods have also been applied. Emamian et al. [13] proposed an Intelligent Monitoring System (IMS) for photovoltaic (PV) systems using affordable and cost-efficient hardware and lightweight software that is capable of being easily implemented in different locations and having the capability to be installed in different types of PV power plants. Mahmud et al. [14] studied Alice Springs, one of the geographically solar energy-rich areas in Australia, and considered the various environmental parameters before providing a machine learning-based PV power generation forecasting for both the short- and long-term.

2.3. Anomaly Detection in Sequence Data

Sequential data anomalies refer to data samples with feature distributions that are significantly different from the norm. Anomaly detection, commonly used in areas such as network security, equipment maintenance, and medical diagnosis, is the process of screening out these anomalies.

Currently, existing anomaly detection methods can be broadly classified into the following three categories: traditional statistical methods, machine-learning methods, and deep learning methods. First, the traditional statistical approach assumes that normal data objects are generated by a statistical model, in which anomalies are identified as objects that do not fit the model perfectly. Abuzaid et al. [15] proposed a density-based anomaly detection method that identifies objects in low-density regions that are relatively far from their immediate neighbors as anomalies. However, traditional statistical methods are time-consuming and laborious. Hence, scholars have shifted their thinking and used machine-learning methods instead of manual methods for anomaly detection. Ergen et al. [16] investigated anomaly detection in an unsupervised framework and introduced Long Short-Term Memory (LSTM) neural network-based algorithms, obtaining anomaly detection algorithms that can process variable length data sequences while providing high performance, especially for time series data. Cook et al. [17] provided background on the challenges which may be encountered when applying anomaly detection techniques to IoT data, with examples of applications for the IoT anomaly detection taken from the literature, and summarized the other current challenges being faced in the anomaly detection domain with a view to identifying potential research opportunities for the future. Maleki et al. [18] developed an enhanced training algorithm for anomaly detection in unlabeled sequential data such as time-series, passing non-anomalous data through to train a deep Long Short-Term Memory (LSTM) autoencoder that distinguishes anomalies when the reconstruction error exceeds a threshold. Ullah et al. [19] extracted spatiotemporal features from a series of frames by passing each one through a pre-trained convolutional neural network (CNN) model. The features extracted from the sequence of frames were valuable in capturing anomalous events. They then passed the extracted deep features through a multi-layer bi-directional Long Short-Term Memory (BD-LSTM) model, which could accurately classify ongoing anomalous/normal events in the complex surveillance scenes of smart cities. Abdallah et al. [20] developed a hybrid intrusion detection system (IDS) by combining the convolutional neural network (CNN) and Long Short-Term Memory (LSTM) network; the proposed model is capable of capturing the spatial and temporal features of the network traffic. However, these models are often based on supervised learning implementations thus causing difficulty in achieving effective results for unsupervised learning that lacks training labels. In contrast, in the work on deep learning for anomaly detection, autoencoders (AEs) have received much attention due to their ability to identify anomalies in an unsupervised manner. Mavikumbure et al. [21] presented anomaly detection and types of anomaly detection, recent advancements in AEs typically used in anomaly

detection and AE-based Anomaly Detection (AE-AD) in selected critical infrastructures. Nguyen et al. [22] designed a novel hybrid approach employing clustering methods and autoencoders (AEs) for the detection of network anomalies in a semi-supervised manner.

In the current study, the PV sequence data we use do not have anomaly identification labels. Therefore, we introduce an unsupervised learning approach and use LSTM-AE to achieve anomaly detection of PV sequence data.

2.4. Visualization of Anomaly Detection

Visualization and analysis techniques are proposed to bring intuitive and convenient understanding to various types of research, such as machine learning [23,24], graphical perception [25] and sampling [26], contrastive dimensionality reduction [27], and also for the field of anomaly detection [28]. With the increasing volume of time series monitoring data of PV array, visualization techniques will assist practitioners in extracting anomalous information from the vast and complex volume of data and to support users in identifying and analyzing anomalies. These methods are of great significance to the processing of anomalies and even the development of PV power generation.

In recent years, a large number of visualization methods have been proposed for marking event sequence anomalies. Depending on their presentation, these methods can be classified as timeline-based, hierarchy-based, flow-based and matrix-based visualization methods. Timeline-based visualization methods typically plot the event sequence data in chronological order, with events represented graphically by color, size, or shape to distinguish among the different attributes. Yeshchenko et al. [29] developed the Event Sequence Visualization framework (ESeVis) that gives due credit to traditional information visualization; from process mining, the mapping study provides an integrated perspective on both fields and identifies potential for synergies for future research. Zhou et al. [30] proposed ConceptThread, a visual analytics approach to effectively show the concepts and the relations among them to facilitate effective online learning. Flow-based visualizations often emphasize the transition process between different event sequences. Jin et al. [31] used a stream-based visualization to depict a series of medical events in an electronic case record, with different color coding for different event types. As for the anomaly comparison aspect, Nguyen et al. [32] represented each event as a colored rectangle based on a timeline thus supporting users to verify the anomaly logs and explore the events that lead to sequence anomalies by comparing the similarities and differences between multiple sessions. Yang et al. [33] provided a comprehensive survey of the classical and deep learning-based approaches for visual anomaly detection in the literature.

Intuitive visual elements, which can effectively assist experts in anomaly identification, combined with interactive interfaces can assist experts in integrating domain knowledge into the anomaly analysis process. Thus, a visual analysis system is needed to integrate visual elements and interactive interfaces to help us achieve anomaly detection and cause analysis of the sequence data.

3. Requirement Analysis and System Overview

In this section, we first describe the processing of data. A list of analytical requirements is then summarized after detailed discussions with domain experts. Thus, we propose a real-time data processing and anomaly detection framework based on AIoT to complete the exploration and analysis tasks of PV sequence data.

3.1. Data Description

The data employed in this study were divided into the following two distinct datasets: one pertaining to power generation and the other to meteorological conditions. The power generation dataset was collected at the inverter level, with each inverter comprising multiple solar panel lines. The meteorological data were obtained from a single array of sensors that were positioned in an optimal manner within the facility. The photovoltaic inverters and environmental sensors are integrated into the AIoT system for the purposes

of data collection, processing, and storage. Subsequently, the data underwent modeling and visualization, following the completion of data preprocessing. The two datasets were merged in a temporal fashion to create a multi-attribute sequence data set. To illustrate, Plant 1 comprised 22 inverters, with each inverter linked to multiple sequence data sets. Each of these corresponded to 15 min of measurements, including ambient temperature, module temperature, solar irradiance, and DC power output during the 15-min interval. Nevertheless, despite the exclusion of data points collected in the absence of lighting conditions, which would otherwise result in the cessation of electricity generation by the equipment, the dataset still exhibited some anomalous phenomena. These anomalous phenomena manifested as a lack of electricity generation by the equipment in daylight conditions. The presence of these anomalous phenomena may result in the inadvertent introduction of erroneous sequence features into the anomaly detection model. It was thus imperative to manually filter the data in order to identify and remove the sequence data points that should generate electricity under specified lighting conditions. In conclusion, a total of 359,040 data points were collected under normal lighting conditions, and the data was filtered to obtain 356,999 sequences. Given that the sample data size obtained did not reach one million, and that the ratio of training and testing sets in unsupervised learning is primarily determined by factors such as the characteristics of the dataset, computing resources, and research objectives, there was no fixed ratio. Accordingly, 70% of the randomly selected filtering sequences were designated as the training set, while the remaining data were allocated to the testing set.

3.2. Requirement Analysis

In this study, the research focus was derived from the discussions with two domain experts (E1 and E2). E1 is a professional researcher working in a research institute. E1 has worked for seven years on PV power generation and has extensive experience in the field of PV. E2 is a senior manager who has been working for an energy company for six years. He is involved in the maintenance and troubleshooting of PV power generation equipment, processes which have a strong need for the visual analysis of anomalous data from PV arrays. In the early stages of our collaboration, we held several meetings with the two experts to discuss the main issues related to PV power generation, e.g., what factors affect PV power generation. We interviewed the experts and reviewed the literature to understand that the amount of electricity that can be generated by PV arrays is largely influenced by objective environmental factors, such as climate, weather, and geography. These factors make PV power generation highly diurnal and prone to fluctuating. Moreover, the frequency of PV power generation changes rapidly, with changes occurring within minutes. Therefore, detecting anomalies from PV sequence data effectively is a difficult task. Meanwhile, the experts specified that analysts must be provided with an intuitive visual comparison that facilitates the interpretation of anomalies by visually demonstrating the inherent correlation and differences between the anomalous and normal events. In the form of structured interviews, a list of requirements was distilled and driven by the discussions with experts as follows:

R1. Anomaly presentation based on the temporal and spatial distribution. Anomalies may occur at any time or in any inverters that cannot be known before the occurrence. Thus, an intuitive interface is needed to present the real-time status, operation monitoring, and anomaly detection data of the PV power system, enabling users to select specific times or inverters and accurately locate anomalies. This is especially important when an anomaly examination is considered.

R2. Comparative analysis of anomalies based on sequences display. Intuitive comparisons of sequences are needed to detect anomalies, as the boundary between normal sequences and anomalous sequences is not clearly defined. By providing a set of visual cues, anomalous sequence data can be visually compared in detail and analyzed under unlabeled conditions.

R3. Cause analysis of anomalies with correlation factors. In addition to the PV array equipment failure, there are some critical factors, which affect the occurrence of anomalies in PV sequence data, such as environment, temperature, and sunlight hours. However, the blackbox effect of deep learning models during processing poses significant difficulties in interpreting the results, which need to be combined with a visual interface to allow users to explore the causes.

R4. Interactive Anomalous Sequence Data Analysis System: A visualization and interaction system needs to be developed, which integrates the anomalies detection model, causes analysis and a set of visual cues to help users to achieve the anomaly examination, comparison, detection, and cause exploration.

3.3. System Overview

Based on the identified requirements, we propose an interactive visualization anomaly detection system for PV power systems based on AIoT, enabling users to detect anomalies and conduct cause analysis on real PV sequence datasets. As shown in Figure 3, it primarily comprises three sections.

PV power generation equipment: As shown in Figure 3 (Section A), the PV power generation equipment involves installing smaller PV power generation devices at or near the point of consumption. The equipment includes a PV array, combiner box, DC distribution cabinet, inverter, transformer, power grid, and other components. Among these, the basic unit of PV array is the solar cell, which converts solar energy into electrical energy to drive the load. The combiner box consolidates multiple strings of PV arrays and has monitoring and protection functions. The DC distribution cabinet serves as a secondary combiner and provides protection for both upstream and downstream components. The inverter converts the DC electricity generated by the PV array into sinusoidal AC electricity, which can then be fed into the grid. The transformer regulates voltage levels; increasing the voltage helps reduce transmission losses. The power grid balances the supply and demand, regulates power supply, and ensures stable power delivery.

Artificial Intelligence Internet of Things (AIoT) System: As shown in Figure 3 (Section B), the AIoT system includes a sensor, AIoT data collection, AIoT edge intelligent network transmission modules, AIoT intelligent device management, and AIoT time-series database. Among these, the sensors include high-precision sensors such as temperature sensors, voltage sensors, and current sensors, each dedicated to monitoring specific performance parameters. AIoT data collection is used to monitor and collect the operational parameters of the PV arrays in real time, including temperature, illumination, voltage, and current, etc. AIoT edge intelligence preprocesses the collected real-time data, performs data cleaning, preprocessing, real-time computations, and uploads the data to the system. Network transmission modules offer multiple methods, including wired and wireless options such as Wi-Fi, 5G, and Zigbee, to achieve efficient data transmission. AIoT intelligent device management enables unified access and centralized control of all real-time collected power generation dataset and weather dataset from PV arrays, conducting comprehensive diagnostic analyses to promptly understand operational information and allowing for the optimized reverse control and energy consumption optimization of PV power generation equipment. The AIoT time-series database is used for storing operational data and historical records of the PV power generation system, providing data to the PV visual anomaly detection system.

PV Visual Anomaly Detection System: As shown in Figure 3 (Section C), the PV visual anomaly detection system utilizes sequence learning with PV sequence data, the distribution of spatial and temporal anomaly information, and comparisons between the sequences to provide a set of views that allow users to detect anomalies by selecting the items they want to explore. The calendar view Figure 3 (Section C(b)) is used to present the time-based distribution of anomaly levels, while the matrix view (Figure 3 (Section C(c))) presents the space-based (inverter-based) distribution of anomaly information. Figure 3 (Section C(d)) shows the projection of sequences (anomalous nor not)

that offers a prominent difference to identify the anomalies. We have leveraged a Long Short-Term Memory and AutoEncoder (LSTM-AE) model to achieve anomalous sequences detection with a threshold set by comparing the predicted data with the original data. A time-series-based line chart Figure 3 (Section C(e)) is designed to present the variation characteristics of PV sequence data in detail and to compare the differences between anomalous sequences and their predicted data (**CH1**). In addition, into connection with three correlation factors—irradiance, ambient temperature, and module temperature—a feature matrix is designed among the factors and power generation to explore and analyze the real causes of anomalies (**CH2**). The interactive visualization facilitates the examination, comparison, detection, and cause exploration of sequences in the PV data. Finally, an anomaly detection system is implemented to integrate the anomaly detection model, cause analysis, and a set of visual elements and interactions (**CH3**).

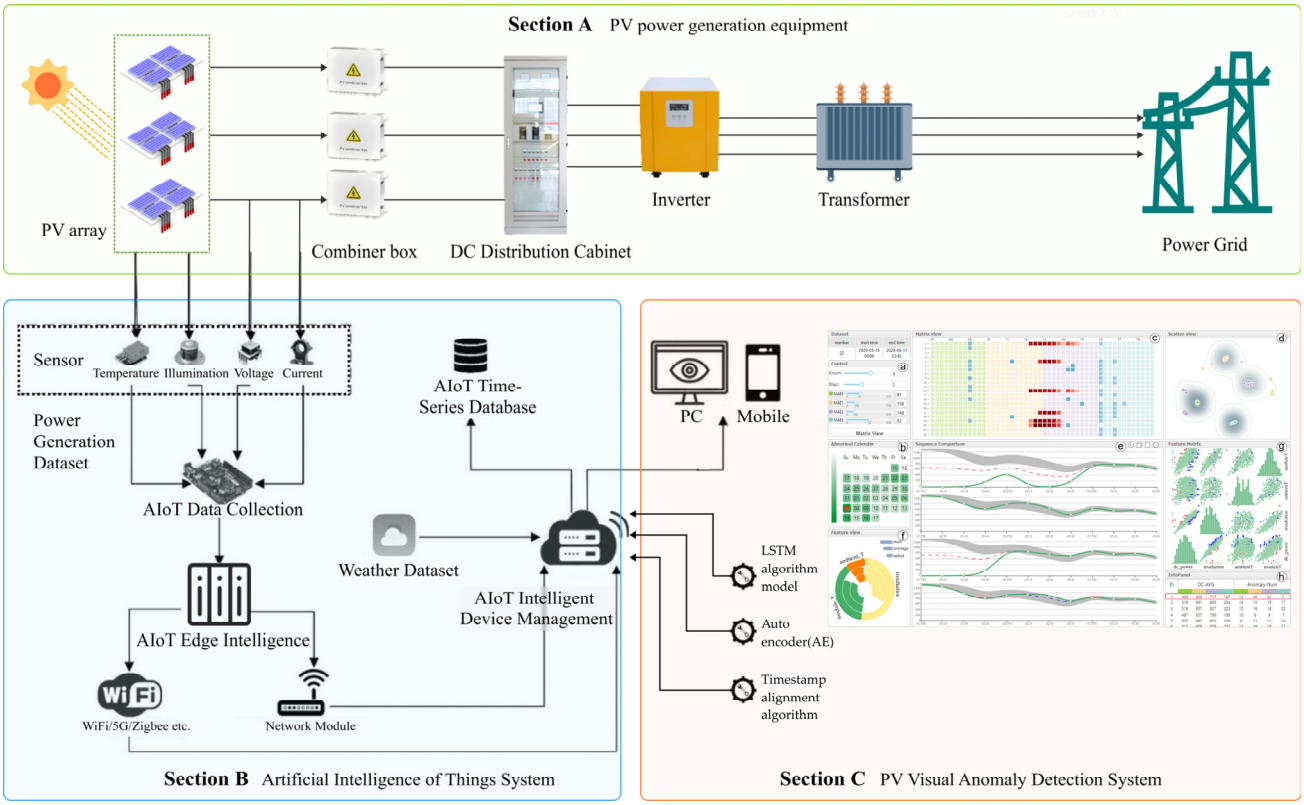


Figure 3. AIoT-Based Visual Anomaly Detection System in PV Sequence Data via Sequence Learning. **Section A** represents PV power generation equipment that includes the installation of smaller PV generation units at or near the point of consumption; **Section B** represents the Artificial Intelligence of Things (AIoT) system; and **Section C** represents the PV Visual Anomaly Detection system. In the **Section C**, (a) The control panel provides the interface for setting the parameters necessary for the identification of anomalies in sequence data. (b) The calendar view allows users to select the specific day for analysis. (c) The matrix view shows whether the power generated by each inverter is anomalous or not during the day. (d) The scatter view shows the distribution of the features of the sequence data. (e) The time analysis view shows specifically the differences in generation characteristics before and after the reconstruction of the sequence data. (f) The feature comparison view shows the differences in environmental factors between the selected data and its counterpart at the same time. (g) The feature matrix view compares the correlation between power generation and environmental variables for sequence data of the same category. (h) The information panel shows information about individual inverters.

3.4. Standard Workflow

The standard data visualization processing workflow of the AIoT-Based Visual Anomaly Detection System is shown in Figure 4 and includes the following steps:

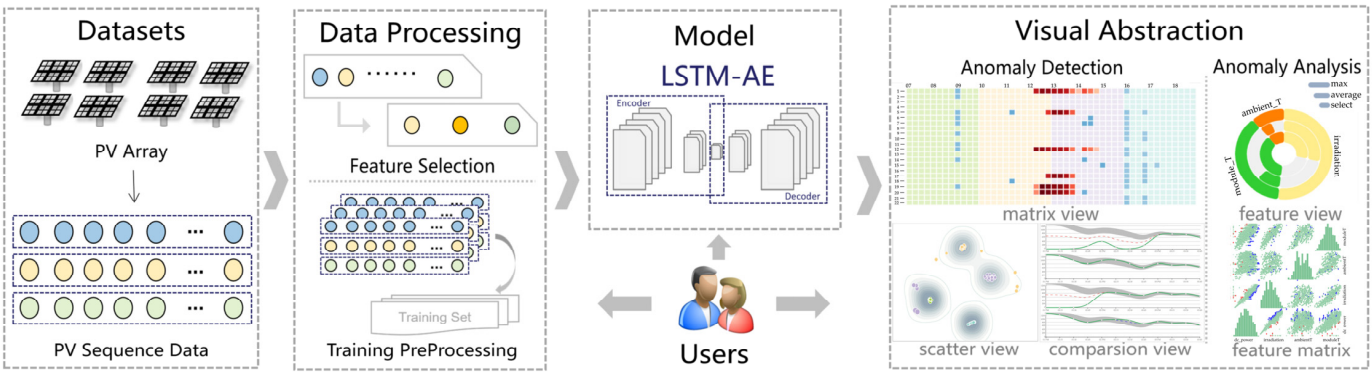


Figure 4. The workflow of our system for anomaly detection and analysis of PV sequence data.

First, the PV sequence dataset is loaded into the AIoT system for feature extraction to reduce the interference information in the dataset. Then, we propose an LSTM-AE-based detection model to select a portion of data as the training set, which is placed it into the model for training and learning. The power generation is predicted based on the training results with the remaining input data. However, when reconstructed by the autoencoder, the data are compressed such that normal values tend to have smaller reconstruction errors than anomalous values. Thus, data anomaly identification could be performed by comparing predicted values with actual values and identifying data with large reconstruction errors as anomalies (R1). The distribution characteristics of the sequence data are expressed through Multiple Dimensional Scaling (MDS) projection, and time-series-based line graphs are used to show in detail the changes in the specific sequence data and anomalies during each period (R2). Furthermore, three of the most critical influencing factors are introduced based on the results of feature extraction, namely PV exposure, module temperature, and ambient temperature, to help users with potential critical cause analysis (R3). A set of visualization and interaction interfaces are integrated into a system for users to detect and analyze the anomalous data and interactively evaluate the critical effects (R4).

4. Anomaly Detection Model

In this section, we detail the working of LSTM-AE and its application to the PV array analysis process.

4.1. LSTM-AE Model

To address the problem of obtaining anomalous event labels in PV sequence data with complex structural features (R1), we utilize the LSTM-AE model to achieve anomalous event detection in the PV sequence data. The first part of the model is an autoencoder (AE). An AE is a hierarchical model whose goal is to make the reconstructed data X' infinitely close to the original data X. The smaller the error is between the two, the better the training effect of AE will be. The AE consists of two parts, namely, the encoder and the decoder.

An M-dimensional vector $X = \{X_1, X_2, \ldots, X_m\}$ is taken as input by the encoder and abstracted into an N-dimensional potential feature vector $H = \{H_1, H_2, \ldots, H_n\}$, thus indicating that m is usually greater than n. Then, the decoder maps the N-dimensional potential feature vector to the expected output vector $X' = \{X'_1, X'_2, \ldots, X'_m\}$ close to the input vector.

$$\begin{aligned} Encoder &: X'_m \rightarrow H_n \\ Decoder &: H_n \rightarrow X'_m \end{aligned} \quad (1)$$

In the encoding section, the output of the encoder can be expressed as in Equation (2). μ_1 is the corresponding connection weight matrix from X to H, $\mu_1 \in R_{n*m}$, which is the bias vector, and α_1 is the activation function.

$$H = \alpha_1(\mu_1 X + \beta_1) \tag{2}$$

In the decoding section, the output of the decoder can be expressed as Equation (3). is the corresponding connection weight matrix from H to X', $\mu_2 \in R_{m*n}$, which is the bias vector, and α_2 is the activation function.

$$X' = \alpha_2(\mu_2 H + \beta_2) \tag{3}$$

Finally, we minimize the mean absolute error (MAE) to ensure that the input vector X is as similar as possible to the output vector. We take this value as the default threshold for detection. Therefore, the objective function of the AE can be described as follows:

$$argminE(X, X') = \frac{1}{m}\sum_{i=1}^{m}|X - X'| \tag{4}$$

The second part of the model is the LSTM network, which is a derivation of the RNN network aiming to avoid gradient disappearance and gradient explosion in the process of long sequence training.

The model used in this study consists of an encoder and a decoder, each of which is an LSTM. Information is selectively passed through a gating mechanism to enhance its time dependence. Figure 5 shows the architecture of the LSTM-AE model. In the encoder phase, a hidden state H_t at time t, where $i - 4 < t \leq i$, is affected by the input vector at the current time and the hidden state at the previous time. Then, the implicit state at time is passed to the decoder, where the process is just the opposite to that in the encoder. The decoder traverses gradually from time i to time $i - 4$. Hence, the hidden state H'_t at time t ($i - 4 \leq t < i$) will be affected by the implied state and the reconstructed vector at the subsequent time.

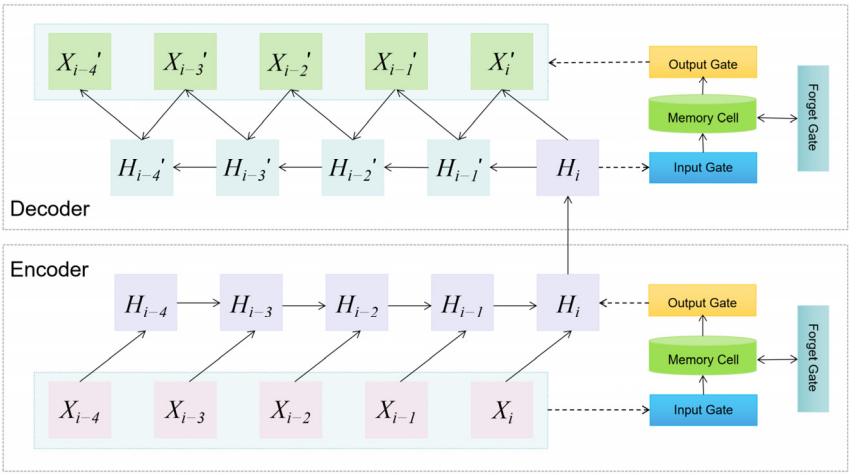


Figure 5. LSTM-AE with input sequence data of length = 5.

4.2. LSTM-AE-Based Anomaly Detection

In general, as is shown in Figure 6, the anomaly detection of PV sequence data based on LSTM-AE model is mainly divided into three steps as follows:

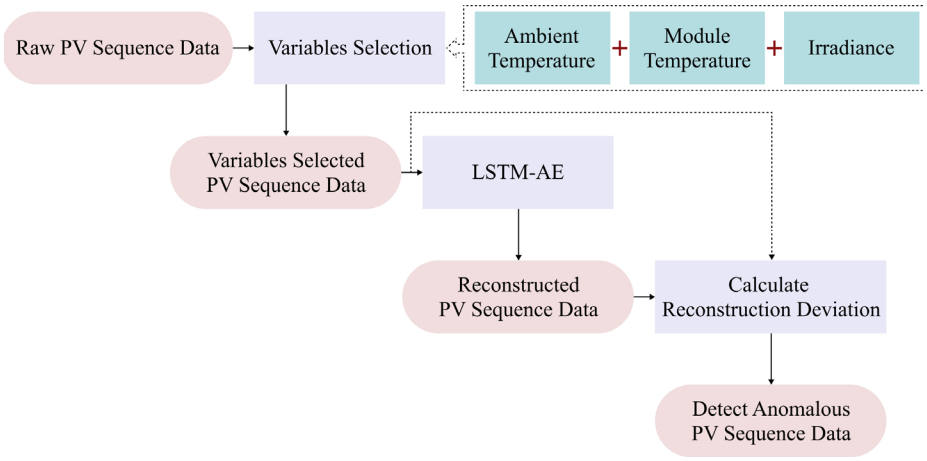


Figure 6. The processing of anomaly detection of PV sequence data.

Step 1. The top three variables (five variables in total) in feature importance, which are calculated by the permutation features' importance, are specified for modeling, namely, ambient temperature, module temperature, and irradiance.

Step 2. The selected variables are combined with power generation as the data for our model input. Then, an LSTM-based encoder is trained to extract a low-dimensional feature representation, the potential vector H, to characterize the progression of each input sequence. The potential vector H is fed into the decoder of the AE model for sequence reconstruction thus restoring the expected value of each factor in the input sequence.

Step 3. The deviation is calculated by comparing the power generation derived from the sequence reconstruction with the true power generation. If the deviation falls outside a fixed threshold, an anomalous event is issued.

However, the above process is not a good strategy for PV power generation data, with daily periodicity and large fluctuation. Thus, the overall attributes are clustered by the k-means method to distinguish periods where the attributes differ significantly. Then, appropriate thresholds are further selected based on the distribution of deviations in the generated categories. To be specific, the time sequences are clustered based on various time sequence features. To evaluate the results of anomaly detection, we utilize the mean absolute error (MAE), which is calculated by the predicted value and the true value. The appropriate MAE is selected as the threshold to display the anomalous detection results.

5. Visual Interfaces

To meet the design requirements discussed in Section 3, we further provide a set of visual designs and interactions to visualize (R2) and analyze (R3) anomaly detection and correlation factors.

5.1. Visual Design

Calendar view: As shown in Figure 3 (Section C(b)), the calendar heat view clearly reflects the distribution of anomalous values over a certain period thus facilitating users for comparative analysis. This view consists of time variables and anomaly levels, where each color block represents a day. In addition, the color encodes the degree of anomaly levels of the PV sequence data, with darker colors indicating a more severe anomaly in the sequence data for that day.

Matrix view: As shown in Figure 3 (Section C(c)), in the matrix view, the 22 inverters are set as the vertical axis and the time periods of the day for power generation as the horizontal axis, which is divided into units of every 15 min. At the same time, we divide the time series of electricity generation within a day into categories and quantities according to user analysis needs and then map them with different colors to meet the convenience needs of subsequent analysis. In this paper, we have classified them into four categories

and mapped them using light green, light orange, light purple, and light blue, respectively. Each of the matrix box indicates whether the power generated by that inverter at the corresponding time is normal or anomalous. When the power generated is normal, the color of the original time category is presented. Red indicates that the predicted value of sequence data is higher than the actual value, blue indicates that the predicted value of sequence data is lower than the actual value, and the degree of the anomaly is indicated by the shade of the color.

Scatter view: The scatter view presents the average features of each inverter at the currently set time step by multidimensional scaling (MDS), where each point represents sequence data at a fixed time step, thus illustrating the distribution of different sequence data in the potential space (Figure 3 (Section C(d))). The vector distribution is calculated by its average features, with the distance between two points reflecting their similarity. Intuitively, multiple high-density regions may imply the aggregation of different patterns, and sequence data located in low-density regions (the outliers) are more prone to anomalies. In addition, we rendered contours based on Kernel Density Estimation (KDE) to show different density regions, with high to low density regions plotted in white to blue, respectively. The density at the location of a sequence is defined by the following equation:

$$D_h(x_s, x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_s - x_i}{h}\right) \tag{5}$$

$D_h(x_s, x_i)$ denotes the kernel density function, K stands for the weight function, x_i represents the location of the rest of the sequences except s, and h denotes the bandwidth. Intuitively, multiple high-density regions may imply the aggregation of different patterns, and sequences located in low-density regions (high outliers) with large radii are more likely to have complex anomalies.

Sequence Comparison view: Once a particular sequence of data is selected, the selected data and other sequence data occurring at the same time step are displayed in the comparison view Figure 3 (Section C(e))). A stream-based format has been devised for the presentation of details pertaining to sequential data, together with a summary of analogous sequential processes, within the aforementioned view. The horizontal axis represents the timestamp of the selected sequence, which corresponds to the time category in the matrix view, with each time interval of 15 min. The vertical axis represents the corresponding power generation, with the power generation unit being kW. It is divided into 2000 kW each and gradually increases to display its power output. Furthermore, the gray background shadow is used to indicate the typical interval range generated in comparable sequences. The normal power generation sequence is mapped in green. To direct users' attention to anomalies, an intuitive glyph was devised to identify anomalies (R1, R2). In instances where the actual generated value is less than the predicted value, this is indicated by an upward-pointing glyph and displayed in red. In the absence of such a case, the glyph is depicted in a downward protrusion and is represented by the color blue. Furthermore, in order to more effectively convey the contrast between anomalous and typical sequence data, upon selecting an anomalous sequence data segment for analysis, the sequence display window will highlight the sequence data exhibiting greater similarity to the selected segment's changing features. The box button, located in the upper right corner of the sequence data display window, is used for selection purposes.

Feature view: As shown in Figure 3 (Section C(f)), we proposed three key factors (irradiance, ambient temperature, and module temperature) affecting output to guide the user in correlating the causes of anomalies. Furthermore, we combined the pie chart with the circular bar chart to correlate the factors for the selected data. The sector size is intended to indicate the extent to which these three factors influence PV generation. In the process of training the model, each feature is predicted and a loss is obtained; each loss is the importance of the feature. If the loss is larger, the feature is more important. The innermost, middle, and outermost bars in the circular bar chart are mapped in two ways, with one representing the correlation value of the selected data point, the average

value of the correlation coefficient of the data points within different dates at the same time, and the maximum value of the correlation coefficient of the total data, and the other representing the correlation coefficient value of the selected data point, the average value of the correlation coefficient of the corresponding data point in the comparison data within the box, and the maximum value of the correlation coefficient of the total data.

Feature matrix view: As shown in Figure 3 (Section C(g)), the feature matrix is drawn to compare correlations between features and power generation. The dispersion of correlations between different features at non-diagonal positions is represented by the horizontal and vertical axes, which correspond to specific values of the corresponding features. Each data point in the figure represents the position of each sequence in the feature correlation of the selected same feature sequence directory. Green represents the normal point of the data, red represents the outlier point where the predicted value of the sequence data is higher than the actual value, and blue represents the outlier point where the predicted value of the sequence data is lower than the actual value. The diagonal position represents a histogram of the relationship between each feature factor and the number of photovoltaic inverters. The horizontal axis represents the specific value of each feature, while the vertical axis represents the number of photovoltaic inverters within the corresponding feature value.

Information view: As shown in Figure 3 (Section C(h)), we counted the average power generation as well as the number of abnormal events for the 22 inverters in the different time categories based on the time steps delineated.

5.2. Interaction

The system also supports the following user interaction to help users analyze anomalies in the PV sequence data.

First, as shown in Figure 3 (Section C(a)), in the control panel, users can determine the step size of the sequence data segments (Step) and the number of clusters to be performed on the sequence data segments (Knum). Alternatively, they can adjust the threshold (MAE) for detecting anomalies in a specific category. In the calendar view, users can select the date to analyze by clicking on the drop-down box or on a day in the calendar. Then, the matrix view and scatter view will show the details of the sequence data for that day. In addition, when the user clicks on a matrix block in the matrix view or a sequence data point in the scatter view that they wish to analyze, the comparison view will display the selected sequence data as well as other sequence data from the same period. When the user wishes to further analyze the causes of anomalies that arise, they can click on a point on the series, such that the feature view will show a comparison of the selected data with data from different dates at the same time in terms of environmental factors. Optionally, the upper right-hand corner of the comparison view can be selected to frame a specific sequence segment, which will then highlight sequence data with high similarity to the generation characteristics of the selected segment. When clicking on a specific point in the framed data, the feature map will show a comparison of the features of the selected data with the average and maximum features of the matched data. By selecting the sequence data, the feature matrix view will show the correlation of all features in the category in which the selected sequence is located. The information view will display basic information about the selected inverter.

6. Evaluation

In this section, the authentic enterprise datasets detailed in Section 3 are utilized to demonstrate the effectiveness of our system in the anomaly detection and analysis of PV sequence data. The validity of our system is then discussed further.

6.1. Case Study

The number of feature clusters for the sequence data were set to four. Excluding the time periods without power generation, we divided the main time periods of the day from

7:00 a.m. to 7:00 p.m. into three-hour intervals. Then, we clustered them according to their generation characteristics into 7:00 a.m. to 10:00 a.m. (green markings), 10:00 a.m. to 1:00 p.m. (yellow markings), 1:00 p.m. to 4:00 p.m. (purple markings), and 4:00 p.m. to 7:00 p.m. (blue markings). The results of the LSTM-AE model indicated that the optimal thresholds for anomaly detection in the four types of sequence data were 91 kW, 158 kW, 148 kW, and 52 kW, respectively.

Case 1. Anomalies identification. For the anomaly detection of unlabeled sequence data, the accuracy of the results would directly affect the analysis and exploration of the anomalous events in subsequent work. Thus, we observed and compared the anomaly identification results to demonstrate that conclusion. A representative day (7 June) was selected to illustrate the case. A large number of serious anomalies during the day were mainly concentrated between 12:00 p.m. and 3:00 p.m. but accounted for two sequence categories (Figure 3 (Section C(c))). Thus, we selected a random sequence data in the corresponding category separately for analysis. In Figure 7, we selected a sequence (A) containing anomalies in the period from 1:00 p.m. to 4:00 p.m. During this period, the actual photovoltaic power generation (green solid line) was relatively low, while the reconstructed sequence data values (red dashed line) were significantly higher than the actual values. The difference between the two exceeded the threshold of 148 kW. In the scatter view, the selected sequence corresponded to scatter A, which was located at a position significantly different from the normal sequence in the category thus indicating that the feature structure of the selected sequence was significantly different compared with the normal sequence. Meanwhile, another selected anomalous sequence (B) within the time interval from 10:00 a.m. to 1:00 p.m. also had a true value of generation significantly lower than its reconstructed predicted value between 12:00 p.m. and 1:00 p.m. It also differed significantly from normal generation, which was reflected as outlier B in the scatter view. The difference between the two is higher than the set threshold of 158 kW. The experimental results proved the effectiveness of our methods in identifying anomalies of PV sequence data.

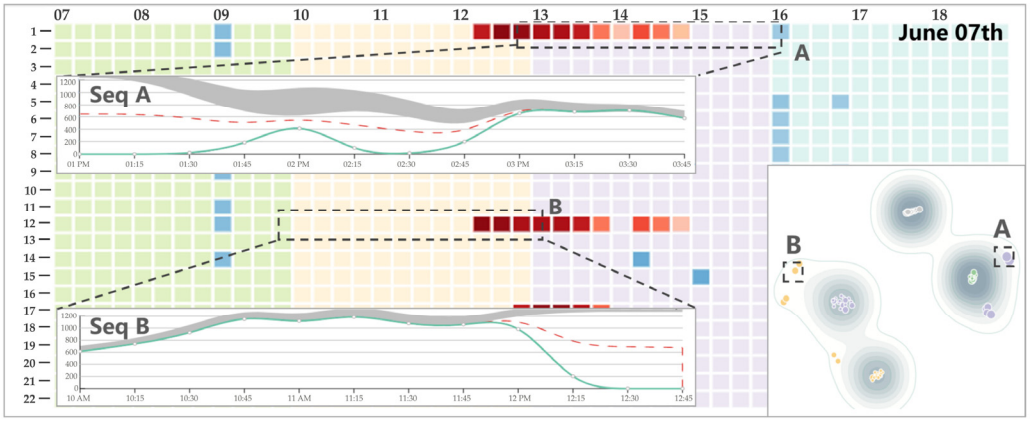


Figure 7. Two different types of sequence data (sequence A and B) and their attribute positions in the scatter plot (scatter points A and B in the bottom right corner of the graph) were used to verify the accuracy of anomaly recognition.

Case 2. Comparison of anomalies. As shown in Figure 8, the output sequence data from inverters No. 1 (sequence A containing a large number of serious anomalies), No. 4 (sequence C is normal), and No. 7 (sequence B containing a few minor anomalies) from the same period in the comparison view were selected for analysis. Under adequate light conditions, the PV output should remain within a high range (gray background range) at around 1:00 p.m. However, the actual output of inverter No. 1 was nearly non-existent and did not return to the normal range until after 3:00 p.m. This level of anomaly is not normally caused by external factors but is most likely due to the failure of its own equipment during

this period. While sequence C showed a slightly higher deviation from the predicted value than the true value at around 2:30 p.m., the occurrence of such anomalous events is usually the result of external factors. The specific analysis will be explained in Case 3.

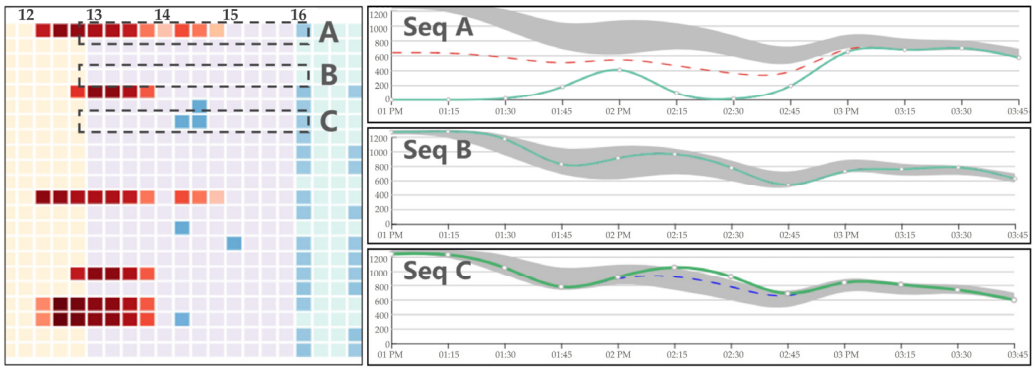


Figure 8. Three different types of sequence data (Sequence A, B and C) were used for comparative analysis of anomaly detection. Among them, Sequence A is an abnormal sequence marked in red, Sequence B is a normal sequence, and Sequence C is an abnormal sequence marked in blue.

In addition, we can also match and compare the characteristics of the anomaly data by framing it. As shown in Figure 9, the sequence data mapped to the light gray rectangular background is the selected anomaly sequence and its matching sequence data. The matching sequence data have similar characteristics to the blue anomaly sequence, which is concentrated around 11:30 a.m., 1:00 p.m., and 2:30 p.m. Notice that the corresponding blue anomalies are also present in the matched sequence segment around 2:30 p.m. This indicates that the features of the currently selected sequence segments cannot match the normal sequence data at the same time, further validating the accuracy of the system in detecting anomalies.

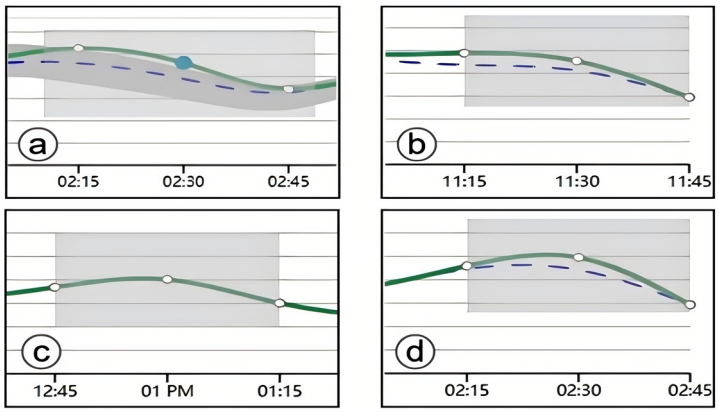


Figure 9. Illustration of the sequence segments comparison. Sequence segment to be framed for analysis (**a**) and those to be matched for comparison (**b**–**d**). Among them, (**a**,**b**,**d**) are both abnormal data marked in blue.

Case 3. Causes analysis of anomalies. The generation of anomalous events depends largely on the degree of failure of the PV array equipment. However, it can also be influenced by environmental factors. Case 2 shows that Inverter No. 1 is highly likely to have an equipment failure at the 1:00 p.m. value of 3:00 p.m. time. Meanwhile, Inverter No. 7 is, to a greater extent, due to extraneous factors. Therefore, based on the training results of the model, the three most influential environmental characteristic variables (irradiance, module temperature, and ambient temperature) are selected for further explanation of the causes of the anomalies. As shown in Figure 10, after comparative analysis of the

blue anomaly, we could further click on the point when the anomaly occurred to see how the analysis of the three factors corresponding to that anomaly has varying degrees of influence on the anomalous event. Among the degrees, irradiance is the highest, and ambient temperature is the lowest. In addition, by clicking on the data point analyzed before and after being framed, a comparison of the environmental factors for the selected data with its data for different dates at the same time and for different matching data for the same date could be viewed separately. In the comparison for different dates at the same time, the module temperature and ambient temperature for the selected data are above average, while the irradiance is below average. For the comparison of matched data on the same date, however, the selected data have the same module temperature and ambient temperature as the average, while the irradiance is below average. Therefore, combining the two comparative analyses shows that the actual power generation at this time is likely to be higher than predicted due to the current higher module temperature of the inverter itself and the ambient temperature where it is located.

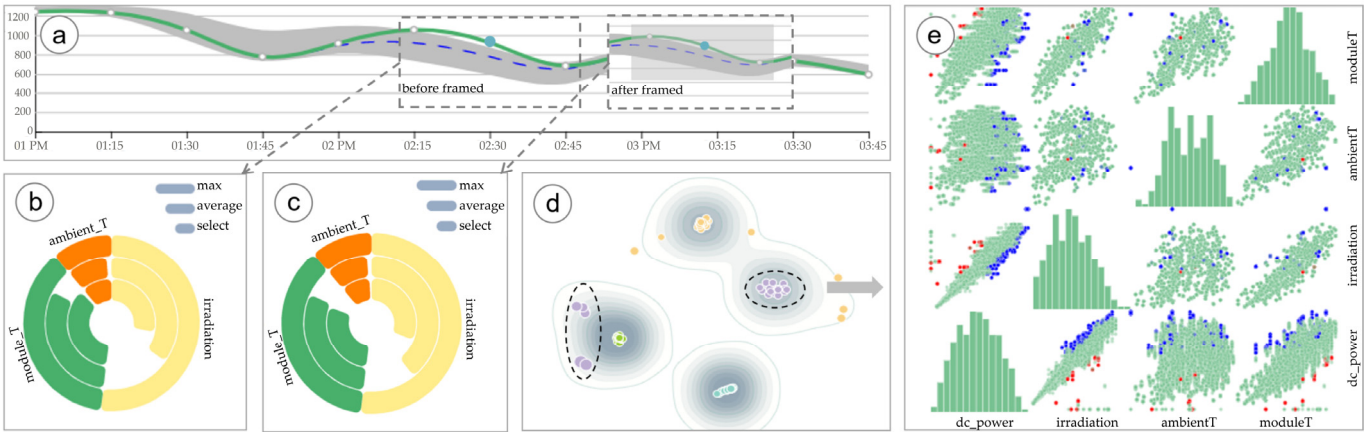


Figure 10. Illustration of the causes analysis. (**a**) presents the anomalous sequence to be analyzed (Sequence C in case 2). (**b**,**c**) present the comparison of environmental factors before and after the anomalous sequence is framed respectively. (**d**) presents the category of the sequence analyzed and (**e**) is the corresponding feature matrix.

Concurrently, the selected sequence category can be characterized as the sequence data exhibiting the same power generation characteristics as the selected sequence (illustrated in the circle in Figure 10d). The resulting data are presented in Figure 10e, and each data point in the figure represents the position of each sequence in the feature correlation of the selected same feature sequence directory. Green represents the normal point of the data, red represents the outlier point where the predicted value of the sequence data is higher than the actual value, and blue represents the outlier point where the predicted value of the sequence data is lower than the actual value. The amount of electricity generated is positively correlated with the amount of PV illumination and the module temperature, i.e., When the amount of PV illumination and module temperature are high, the electricity generated is high as well. A positive correlation also exists between the amount of PV light and the module temperature itself. Notably, in the scatter of the correlation among power generation, PV illumination, and module temperature, some outliers may exist. In these outliers, the PV illumination or the module temperature is high, but the power generation is close to zero. This trend entails that the near-zero power generation is not correlated with the PV illumination and module temperature. Thus, the equipment likely has issues. This scenario further confirms that outliers A and B in Figure 7 are due to a fault in the PV equipment.

6.2. Discussion

In this section, we discussed with the two experts on three aspects—technological advantages, practical impacts, and research limitations. The results of the discussion are as follows.

Technological advantages. The objective of this study is to propose an interactive system analysis tool that will assist users in identifying and exploring abnormal phenomena and their causes in PV sequence data, without the need for abnormal labels. Both experts concur that our work has the following technical advantages. The use of AIoT technology enables the precise collection and time alignment of PV sequence data. A deep learning LSTM-AE model is employed for the effective processing of PV sequence data devoid of anomalous labels, thereby facilitating the more efficient identification and presentation of anomalies. Furthermore, the combination of information and visualization techniques based on global sequence data enables users to compare and analyze anomalous sequence data, thereby facilitating a more comprehensive understanding of anomalous features and their underlying causes. Consequently, our approach to identifying anomalies in photovoltaic power generation sequence data not only effectively processes sequence data without anomaly identification labels but also enables the comparison, analysis, and interpretation of the identified anomalous features.

Practical impacts and contributions. The system tools provided by this research institute facilitate the timely detection and localization of anomalies in photovoltaic systems. Furthermore, they enable the investigation of potential causes of anomalies through the analysis of their characteristics and associated factors. This is of great significance for the preventive maintenance of photovoltaic power generation systems, as it can reduce the incidence of system failures, thereby improving power generation efficiency and stability. This indicates that the research method can assist photovoltaic power plant operators in developing a more comprehensive understanding of their systems, thereby reducing the potential economic impact on power plants. By reducing maintenance costs and improving power generation efficiency, this research contributes to enhancing the economic viability and market competitiveness of photovoltaic power generation projects.

Research limitations and future work. Nevertheless, there remain some outstanding issues in this study that will be addressed in future research. Firstly, at the algorithmic level, improvements can be made. While the LSTM-AE algorithm demonstrates efficacy in anomaly detection, the primary objective of ongoing research is to enhance its precision and reduce the occurrence of false positives. Meanwhile, while existing anomaly detection algorithms may perform well on specific datasets, their performance may be less effective when applied to different photovoltaic systems and environments. This presents a challenge in improving the algorithm's generalization ability [34]. In future work, we will further optimize the deep learning models with a view to reduce errors in anomaly recognition models and improve the accuracy of predicted values thus meeting the requirements of anomaly analysis. The next stage is the analytical level. The causes of abnormal events are influenced by a number of factors, both directly and indirectly. This study introduced three environmental characteristic variables as influencing factors for explanatory causal analysis, but this is believed to be an insufficient approach [35]. In future work, we will introduce additional influencing factors and investigate the correlation between each characteristic factor to ascertain their impact on abnormal events, thereby facilitating more effective exploration and analysis of the causes of various abnormal events. Finally, at the level of practical application. As technology advances, photovoltaic systems may incorporate novel components and equipment, necessitating the ability of anomaly detection systems to adapt, maintain scalability, and facilitate maintenance. Furthermore, the integration of AIoT technology can enhance data synchronization and real-time processing capabilities; however, ensuring data accuracy and real-time performance remains a significant challenge [36]. In future work, the focus will be on further developing anomaly detection applications for photovoltaic sequence data, with a view to enhance the scalability and maintainability of the system.

7. Conclusions

This study proposes a data collection and preprocessing method utilizing a rigorous timestamp alignment algorithm for intricate photovoltaic power generation data structures. Furthermore, it employs an LSTM-AE model to identify anomalies in photovoltaic sequence data. Moreover, a visual analysis system for anomaly detection has been developed to provide technical support for photovoltaic array monitoring. The system provides a set of visual comparisons to describe photovoltaic sequence data of varying granularities in multiple views, thereby facilitating comprehension of anomalous events. Concurrently, the combination of feature views enables correlation analysis between anomalies and associated variables in photovoltaic power generation thus facilitating the further elucidation and investigation of anomalies through the utilization of attribute features by analysts. The practicality of our system for anomaly detection has been demonstrated by case studies and quantitative comparisons based on authentic enterprise datasets from different perspectives, as well as explorations of PV sequence data. The method proposed in this study is therefore designed to assist photovoltaic power plant operators in developing a deeper understanding of their systems, thereby reducing the potential economic impact on power plants.

Author Contributions: Conceptualization, Q.W. and Z.Z.; methodology, H.S. and G.L.; software, J.F. and G.L.; validation, J.F. and Z.Z.; formal analysis, Z.Z. and G.L.; investigation, Q.W. and G.L.; resources, Q.W.; data curation, J.F. and G.L.; writing—original draft preparation, J.F. and G.L.; writing—review and editing, Q.W., H.S. and Z.Z.; visualization, J.F., Z.Z. and G.L.; supervision, Q.W. and Z.Z.; project administration, Q.W.; funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52377020; the Public Welfare Project of Zhejiang Provincial Science and Technology Department No. LGF22F020034; and the APC was funded by National Natural Science Foundation of China.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical restrictions.

Conflicts of Interest: Authors Qian Wei and Jingjing Fan were employed by the Kunlun Digital Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Rosato, A.; Altilio, R.; Araneo, R.; Panella, M. Prediction in Photovoltaic Power by Neural Networks. *Energies* **2017**, *10*, 1003. [CrossRef]
2. Wang, K.; Qi, X.; Liu, H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* **2019**, *251*, 113315. [CrossRef]
3. Guo, S.; Jin, Z.; Chen, Q.; Gotz, D.; Zha, H.; Cao, N. Interpretable Anomaly Detection in Event Sequences via Sequence Matching and Visual Comparison. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 4531–4545. [CrossRef]
4. Humada, A.M.; Darweesh, S.Y.; Mohammed, K.G.; Kamil, M.; Mohammed, S.F.; Kasim, N.K.; Tahseen, T.A.; Awad, O.I.; Mekhilef, S. Modeling of PV system and parameter extraction based on experimental data: Review and investigation. *Sol. Energy* **2020**, *199*, 742–760. [CrossRef]
5. Zhao, Y.; Liu, Q.; Li, D.; Kang, D.; Lv, Q.; Shang, L. Hierarchical Anomaly Detection and Multimodal Classification in Large-Scale Photovoltaic Systems. *IEEE Trans. Sustain. Energy* **2018**, *10*, 1351–1361. [CrossRef]
6. Li, Y.B.; Du, X.Q.; Wan, F.Y.; Wang, X.Z.; Yu, H.C. Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. *Chin. J. Aeronaut.* **2020**, *33*, 427–438. [CrossRef]
7. Matin, A.; Islam, M.R.; Wang, X.; Huo, H.; Xu, G. AIoT for sustainable manufacturing: Overview, challenges, and opportunities. *Internet Things* **2023**, *2023*, 100901. [CrossRef]
8. Zhang, J.; Tao, D. Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet Things J.* **2020**, *8*, 7789–7817. [CrossRef]
9. Muhammed, D.; Ahvar, E.; Ahvar, S.; Trocan, M.; Montpetit, M.J.; Ehsani, R. Artificial Intelligence of Things (AIoT) for smart agriculture: A review of architectures, technologies and solutions. *J. Netw. Comput. Appl.* **2024**, *228*, 103905. [CrossRef]
10. Mehmood, M.; Ulasyar, A.; Ali, W.; Zeb, K.; Zad, H.S.; Uddin, W.; Kim, H.J. A new cloud-based IoT solution for soiling ratio measurement of PV systems using artificial neural network. *Energies* **2023**, *16*, 996. [CrossRef]

11. Chen, J.; Pan, G.; Ouyang, J.; Ma, J.; Fu, L.; Zhang, L. Study on impacts of dust accumulation and rainfall on PV power reduction in East China. *Energy* **2020**, *194*, 116915. [CrossRef]
12. Sawadogo, W.; Abiodun, B.J.; Okogbue, E.C. Impacts of global warming on photovoltaic power generation over West Africa. *Renew. Energy* **2020**, *151*, 263–277. [CrossRef]
13. Emamian, M.; Eskandari, A.; Aghaei, M.; Nedaei, A.; Sizkouhi, A.M.; Milimonfared, J. Cloud computing and IoT based intelligent monitoring system for photovoltaic plants using machine learning techniques. *Energies* **2020**, *15*, 3014. [CrossRef]
14. Mahmud, K.; Azam, S.; Karim, A.; Zobaed, S.; Shanmugam, B.; Mathur, D. Machine learning based PV power generation forecasting in alice springs. *IEEE Access* **2021**, *9*, 46117–46128. [CrossRef]
15. Abuzaid, A.H. Identifying density-based local outliers in medical multivariate circular data. *Stat. Med.* **2020**, *39*, 2793–2798. [CrossRef]
16. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3127–3141. [CrossRef]
17. Cook, A.A.; Mısırlı, G.; Fan, Z. Anomaly detection for IoT time-series data: A survey. *IEEE Internet Things J.* **2019**, *7*, 6481–6494. [CrossRef]
18. Maleki, S.; Maleki, S.; Jennings, N.R. Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering. *Appl. Soft Comput.* **2021**, *108*, 107443. [CrossRef]
19. Ullah, W.; Ullah, A.; Haq, I.U.; Muhammad, K.; Sajjad, M.; Baik, S.W. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimed. Tools Appl.* **2021**, *80*, 16979–16995. [CrossRef]
20. Abdallah, M.; An Le Khac, N.; Jahromi, H.; Delia Jurcut, A. A hybrid CNN-LSTM based approach for anomaly detection systems in SDNs. In Proceedings of the 16th International Conference on Availability, New York, NY, USA, 17 August 2021.
21. Mavikumbure, H.S.; Wickramasinghe, C.S.; Marino, D.L.; Cobilean, V.; Manic, M. Anomaly detection in critical-infrastructures using autoencoders: A survey. In Proceedings of the IECON 2022—48th Annual Conference of the IEEE Industrial Electronics Society, Brussels, Belgium, 17–20 October 2022.
22. Nguyen, V.Q.; Nguyen, V.H.; Le-Khac, N.A.; Cao, V.L. Clustering-based deep autoencoders for network anomaly detection. In Proceedings of the 7th International Conference on Future Data and Security Engineering (FDSE 2020), Quy Nhon, Vietnam, 25–27 November 2020.
23. Yuan, J.; Chen, C.; Yang, W.; Liu, M.; Xia, J.; Liu, S. A survey of visual analytics techniques for machine learning. *Comput. Vis. Media Vol.* **2021**, *7*, 3–36. [CrossRef]
24. Shah, N.; Bhagat, N.; Shah, M. Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention. *Vis. Comput. Ind. Biomed. Art* **2021**, *4*, 9. [CrossRef] [PubMed]
25. Zhao, Y.; Shi, J.; Liu, J.; Zhao, J.; Zhou, F.; Zhang, W.; Chen, K.; Zhao, X.; Zhu, C.; Chen, W. Evaluating Effects of Background Stories on Graph Perception. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 4839–4854. [CrossRef] [PubMed]
26. Zhao, Y.; Jiang, H.; Chen, Q.; Qin, Y.; Xie, H.; Wu, Y.; Liu, S.; Zhou, Z.; Xia, J.; Zhou, F. Preserving Minority Structures in Graph Sampling. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 1698–1708. [CrossRef] [PubMed]
27. Xia, J.; Huang, L.; Lin, W.; Zhao, X.; Wu, J.; Chen, Y.; Zhao, Y.; Chen, W. Interactive Visual Cluster Analysis by Contrastive Dimensionality Reduction. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 734–744. [CrossRef] [PubMed]
28. Wang, X.; Chen, W.; Xia, J.; Wen, Z.; Zhu, R.; Schreck, T. HetVis: A Visual Analysis Approach for Identifying Data Heterogeneity in Horizontal Federated Learning. *IEEE Trans. Vis. Comput. Graph.* **2023**, *29*, 310–319. [CrossRef]
29. Yeshchenko, A.; Mendling, J. A survey of approaches for event sequence analysis and visualization. *Inf. Syst.* **2024**, *120*, 102283. [CrossRef]
30. Zhou, Z.; Ye, L.; Cai, L.; Wang, L.; Wang, Y.; Wang, Y.; Chen, W.; Wang, Y. Conceptthread: Visualizing threaded concepts in MOOC videos. *arXiv* **2024**, arXiv:2401.11132. [CrossRef]
31. Jin, Z.; Cui, S.; Guo, S.; Gotz, D.; Sun, J.; Cao, N. CarePre: An Intelligent Clinical Decision Assistance System. *ACM Trans. Comput. Healthc.* **2020**, *1*, 1–20. [CrossRef]
32. Nguyen, P.H.; Turkay, C.; Andrienko, G.; Natalia, A.; Olivier, T.; Jihane, Z. Understanding User Behaviour through Action Sequences: From the Usual to the Unusual. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 2838–2852. [CrossRef]
33. Yang, J.; Xu, R.; Qi, Z.; Shi, Y. Visual anomaly detection for images: A systematic survey. *Procedia Comput. Sci.* **2022**, *199*, 471–478. [CrossRef]
34. Akbarian, H.; Mahgoub, I.; Williams, A. Autoencoder-LSTM Algorithm for Anomaly Detection. In Proceedings of the 2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), Boca Raton, FL, USA, 4–6 December 2023.
35. Ibrahim, M.; Alsheikh, A.; Awaysheh, F.M.; Alshehri, M.D. Machine learning schemes for anomaly detection in solar power plants. *Energies* **2022**, *15*, 1082. [CrossRef]
36. Ullah, I.; Adhikari, D.; Su, X.; Palmieri, F.; Wu, C.; Choi, C. Integration of data science with the intelligent IoT (IIoT): Current challenges and future perspectives. *Digit. Commun. Netw.* **2024**, *5*, 12–26. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.