A Domain Knowledge-Driven Visual Analytics System for Photovoltaic Power Time Series Forecasting

Xinjing Yi[1,2], Yurun Yang[1,2], Yin Song[1], Baozhu Zhou[3], Cheng Li[1], Shuhan Liu[4], Dazhen Deng[2], Di Weng[2*], Yingcai Wu[4]

[1]China Datang Corporation Technology Innovation Co., Ltd., Xiong'an New Area, 070001, Hebei, China.
[2]School of Software Technology, Zhejiang University, Ningbo, 315000, Zhejiang, China.
[3]China Datang Zhejiang Branch, Hangzhou, 310000, Hangzhou, China.
[4]State Key Lab of CAD & CG, Zhejiang University, Hangzhou, 310000, Hangzhou, China.

*Corresponding author(s). E-mail(s): dweng@zju.edu.cn;
Contributing authors: yixinjing@zju.edu.cn; yurunyang@zju.edu.cn;
song-y04@163.com; 24039221@qq.com; 601664999@qq.com;
shliu@zju.edu.cn; dengdazhen@zju.edu.cn; ycwu@zju.edu.cn;

Abstract

Accurate forecasting of photovoltaic (PV) power generation plays a critical role in optimizing production schedules for PV power stations and facilitating efficient maintenance and repair of PV equipment. Deep time series forecasting models have emerged as the leading approach in the field of time series prediction. However, the intricate architectures and numerous parameters in these models frequently lead to forecasts that conflict with established domain knowledge. Furthermore, domain experts face significant challenges in leveraging their specialized knowledge to enhance the predictive performance of these models. Additionally, when dealing with PV strings exhibiting diverse characteristics, it becomes particularly challenging to identify discrepancies in the model's predictive accuracy across different strings and pinpointing its performance flaws. To tackle the issues, we propose a domain knowledge-driven forecasting method that combines the deep time series forecasting model with decision tree regression. Empirical evaluations demonstrate that our proposed method yields substantial

improvements in forecasting accuracy compared to existing approaches. Furthermore, we have developed and implemented a comprehensive visual analytics system. This system aids experts in forecasting PV string power generation trends by employing multi-perspective model evaluation techniques to assess the rationality of forecasts.It also incorporates domain knowledge through an interactive decision tree construction process, thereby enhancing the model's predictive capabilities. The efficacy of our proposed system is substantiated through in-depth case studies and rigorous user evaluations.

Keywords: Visual analytics, Time series forecasting, Model interpretability

1 Introduction

The development of modern photovoltaic (PV) power generation technology and the proliferation of PV power stations have led to the accumulation of vast amounts of real-time monitoring data. This wealth of information significantly enhances our ability to forecast PV power generation, conduct fault analyses, make informed decisions, and optimize operations and maintenance procedures. PV power generation is inherently subject to considerable intermittency and volatility due to unpredictable factors, particularly meteorological conditions (Lai et al., 2019). These fluctuations not only give rise to issues such as frequency harmonic distortion (Ahmed et al., 2020), which compromises the reliability and stability, but also affects the dispatching and production processes of PV power stations. Consequently, accurate forecasting of PV power generation plays a crucial role in enhancing both resource utilization and overall power system security.

In addressing the challenge of accurate PV power generation forecasting, current research predominantly focuses on methods leveraging deep time series forecasting models (Zhang et al., 2018; Abdel-Nasser and Mahmoud, 2019; Qu et al., 2021; Phan et al., 2022). Despite their prevalence, these approaches continue to grapple with two significant limitations:

- **Limited interpretability**: The intrinsic "black box" nature of deep learning models engenders skepticism among experts regarding the reliability of forecast results (Liang et al., 2021). This opacity hinders the ability to effectively evaluate performance disparities across various components of PV power stations and impedes the development of comprehensive operation and maintenance strategies.
- **Constrained accuracy and flexibility**: Typical PV power stations comprise thousands of PV strings, and the status of the PV strings is affected by many factors such as geographical location, installation angle, and orientation. Existing models often fall short in incorporating domain-specific knowledge, rendering them ill-equipped to handle the highly dynamic forecasting scenarios encountered in real-world PV power stations.

Visualization has been widely applied in deep learning, encompassing time series forecasting models (Hohman et al., 2019). Existing works on visual analytics for time

series forecasting ([Bogl et al., 2013](); [Sun et al., 2020](); [Shen et al., 2020](); [Xu et al., 2021](); [Jin et al., 2023](); [Hao et al., 2024]()) leverages model interpretation and visual analytics techniques to provide intuitive representations of a model's internal structural parameters, inputs, and output. While these approaches have undoubtedly enhanced our ability to comprehend model behaviors and facilitate model assessment and selection, they frequently fall short in empowering domain experts to incorporate their specialized insights into model optimization processes.

The integration of domain knowledge holds paramount importance in sectors such as PV energy production, intelligent manufacturing, and supply chain management. Consequently, the effective incorporation of this expertise into deep time series forecasting models presents a pressing issue ripe for resolution.

To address the aforementioned challenges, we propose an innovative PV power forecasting methodology that integrates the deep time series forecasting model with the decision tree algorithms. This hybrid approach is designed to incorporate domain knowledge through an interactive decision tree framework, thereby enhancing forecasting accuracy. Additionally, we have designed and implemented a sophisticated visual analytics platform that facilitates a comprehensive evaluation of model performance, enables in-depth exploration of power generation time series data, and aids in identifying model performance deficiencies. We also provide an interactive decision tree construction tool that empowers experts to refine models based on their domain-specific insights. To validate our approach, we present two case studies utilizing real-word historical data from PV strings and their corresponding models, illustrating how the visual analytics system assists experts in comprehending and assessing models, as well as applying interactive techniques to optimize model performance. These case studies serve to substantiate the system's practical utility and effectiveness. The key contributions of this study are summarized as follows:

- We propose a novel PV power forecasting approach that effectively incorporates domain knowledge and enhances the interpretability of the model.
- We have developed an interactive visual analytics system that integrates a suite of tailored visualizations, designed to assist domain experts in PV time series analysis, model evaluation and optimization.
- We have conducted two comprehensive case studies that demonstrate the efficacy of the proposed system, garnering positive feedback from domain experts.

2 Related Work

2.1 PV power generation forecasting

PV power generation forecasting models can be classified into three categories based on their underlying principles: physical, statistical, and artificial intelligence models ([Lai et al., 2019](); [Ahmed et al., 2020]()). Physical models, which rely on the fundamental principles of PV power generation, are limited by issues such as poor robustness to interference and restricted applicability ([Lai et al., 2019]()). Statistical approaches, primarily autoregressive models, are only suitable for short-term forecasting ([Ahmed

et al., 2020). With the rapid increase in data volume and the increasing complexity of power systems, deep learning models have shown superior performance in PV forecasting tasks (Sharda et al., 2021).

In recent years, numerous time series forecasting studies (Liu et al., 2024; Kitaev et al., 2020; Zhou et al., 2021; Wu et al., 2022) have employed the Transformer architecture, consistently achieving state-of-the-art performance across a spectrum of forecasting tasks. The iTransformer model (Liu et al., 2024), built on this architectural foundation, uses an input transposition technique to independently encode single-variable sequences into high-dimensional features. It employs the attention mechanism to model the interdependencies among variables, significantly boosting its predictive accuracy on a range of real-world datasets. This model effectively captures the complexities of multivariate correlations, exhibiting robust generalization capabilities.

Based on the model evaluation experimental results detailed in Sec. 3.3, we have selected the advanced iTransformer model for forecasting PV string power generation.

2.2 Visual analytics for time series forecasting models

Over the past few years, a multitude of studies have explored the application of visual analytics in time series forecasting. These efforts aim to assist professionals in model evaluation and selection (Bogl et al., 2013; Sun et al., 2020; Xu et al., 2021), as well as in understanding model behaviors and enhancing model effectiveness (Shen et al., 2020; Hao et al., 2024; Jin et al., 2023).

TiMoVA (Bogl et al., 2013) combines the Box-Jenkins methodology for model selection with interactive visualization techniques to compare ARIMA models across various parameter configurations. Sun et al. (2020) and Jin et al. (2023) have analyzed RNN models and attention mechanisms using model-driven interpretation (Liang et al., 2021), significantly improving model performance. In contrast to model-driven interpretation, data-driven methods (Liang et al., 2021) are independent of network architecture, focusing more on input data rather than model structure. DFSeer (Sun et al., 2020), designed for historical demand forecasting models, facilitates model comparisons and selections at various levels of detail while identifying predictive risks for specific products. Xu et al. (2021) propose interactive model-building approaches, incorporating SHAP to examine inputs contributions to outputs in multivariate time series forecasting models. TimeTuner (Hao et al., 2024) offers a suite of time series transformation techniques, exploring the relationships between the local characteristics, time series stationarity, and model behaviors.

Based on existing research, we aim to integrate visual analytics with PV power generation forecasting models to enhance the interpretability of complex data relationships and model behaviors. Furthermore, we propose a visual analytics system design that facilitates multi-level model assessment, deepening experts' understanding of time series forecasting models and aiding in the exploration and validation of their reliability.

2.3 Interactive decision tree construction

Decision trees are typically constructed using a recursive top-down divide-and-conquer approach (Van Den Elzen and Van Wijk, 2011). They offer the advantages of transparency and interpretability (Streeb et al., 2022), making them popular for both classification and regression tasks. Recent advancements in interactive decision tree construction have transformed the traditional iterative process, enabling experts to refine decision trees more effectively by incorporating domain-specific knowledge (Van Den Elzen and Van Wijk, 2011; Teoh and Ma, 2003a,b; Muhlbacher et al., 2018; Sarailidis et al., 2023).

BaobabView (Van Den Elzen and Van Wijk, 2011) outlines the essential requirements for the interactive decision tree construction, encompassing node creation, automatic subtree growth, selection of split points and attribute, pruning and analytical highlighting. This system provides various integrated views that clearly illustrate the decision tree's classification process. Sarailidis et al. (2023) have developed iDT, an open-source tool for interactive decision tree construction. Their case study demonstrates its effectiveness in empowering Earth science experts to incorporate domain knowledge into their models. Drawing inspiration from BaobabView (Van Den Elzen and Van Wijk, 2011) and iDT (Sarailidis et al., 2023), we aim to facilitate experts in applying domain knowledge to PV power forecasting through the interactive development of decision trees.

3 Data preparation and model evaluation

This section details the construction of a time series forecasting dataset for PV string power generation. It also presents benchmark experiments using various deep time series forecasting models based on the Transformer architecture, demonstrating the predictive performance of the iTransformer.

3.1 Data

We collected electrical, meteorological, and geographic data from a centralized PV power station in Zhejiang Province from November 27, 2022, to March 26, 2023. The electrical data, organized by PV inverter, records current and voltage readings for each parallel-connected string at one-minute intervals. Each PV string is uniquely identified using a standardized nomenclature: **"BT[Box Transformer Number]-I[Inverter Number]-DC[String Number]"**. The meteorological data were collected by the weather station within the PV power plant, encompassing site-specific parameters including ambient temperature, total solar irradiance, and wind speed, with a sampling frequency of one minute. The station management platform provides comprehensive geographic coordinate data for PV strings, detailing both string position number and corresponding latitude/longitude information.

To reduce noise and facilitate the models' ability to capture long-term trends and cyclic patterns, we downsampled the dataset to a one-hour frequency. The operational dataset of the PV power station is composed of sequentially recorded data points,

with each PV string sequence represented in:

$$P_m = (p_1, p_2, ..., p_n) \tag{1}$$
$$p_i = (I_i, V_i, t_i) \tag{2}$$

Here, P_m represents the sequence data of the m-th string, with a sequence length of n. Each data point p_i in the time series consists of current (I_i), voltage (V_i), and timestamp (t_i).

After extracting the PV string data, we conducted a data cleansing process to eliminate anomalies due to communication or monitoring issues. The anomalies and corresponding processing methods are as follows:

- **Missing electrical data.** Missing data may occur during the collection of electrical data from PV strings. Electrical data includes voltage and current. However, for some strings, only voltage or current is recorded, preventing power output calculation. We retain only the matched current and voltage data. Additionally, a large amount of electrical data was missing from December 1st to December 8th, 2022. Due to the extensive missing data, we discarded the data within this period.
- **Two strings connected to one port.** The PV power station's inverter has 18 DC input ports. Theoretically, each port should be connected to one PV string. However, in actual installation, there is a situation in which two strings are connected to one port. As shown in Fig. 1(b), there is an obvious twofold relationship between the historical currents of the abnormal string and the normal string. This likely results from incorrect port configuration. We removed the abnormal data of the cases where two strings are connected to one port.
- **Meteorological data noise.** Significant noise interferes with the total solar irradiance in the meteorological data. We performed time-series correction and smoothing to reduce noise and enhance data usability.

After cleansing the data, we calculated the power output for each string. The final dataset comprises 1,246 strings.

3.2 Time series forecasting dataset

Time series forecasting datasets typically have a three-dimensional structure ($N \times T \times D$), where:

- N is the number of sequences
- T is the sequence length
- D is the feature dimension

A dataset composed of N sequences is denoted as $\mathbb{X} = \{X_1, X_2, ..., X_N\}$. Each sequence X_i consists of T multidimensional feature vectors, represented as $X_i = \{x_{i,1}, x_{i,2}, ..., x_{i,T}\}$. Each feature vector $x_{i,t}$ at time step t contains D individual features, denoted as $x_{i,t}^k$, where $k \in [1, D]$. In this context, $x_{i,t}^k$ indicates the power generation of string k at time step t in the sequence.

For training our time series forecasting models, we use processed PV string power generation data. The dataset includes 1,246 independent historical power generation
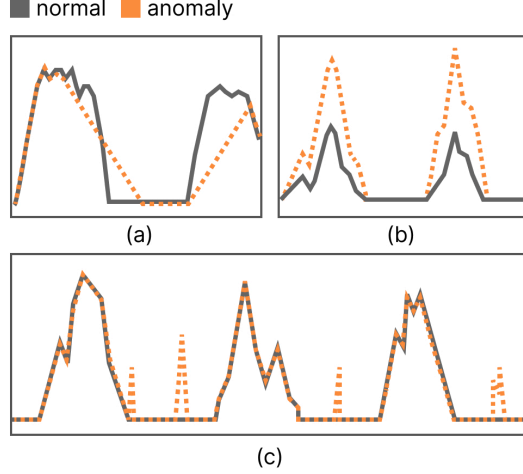
Fig. 1: The anomalies review: (a) missing electrical data, (b) two strings connected to one port, (c) meteorological data noise.

time series from PV strings, created using a sliding window approach. PV string's power generation data is segmented into 120-hour time series, which serves as input features for the model. The sliding window advancs at 1-hour intervals. The model is designed to forecast power generation for the 24 hours following each input sequence. We chronologically divided the dataset into training (70%), validation (10%), and test (20%) sets. To address the issue of varying feature scales in model training, we standardized the entire dataset based on the statistics of the training set.

3.3 Model evaluation

All experiments were conducted on a single NVIDIA A100 80GB GPU using PyTorch. We optimized the model using the Adam optimizer and L2 loss function, with an initial learning rate of 0.0005. The sequence representation dimension was consistently set to 512 across all models. We used a batch size of 32 and trained for 10 epochs.

To evaluate the model's predictive performance, we employed Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics, where lower values indicate more accurate predictions. The model training and evaluation were performed on the cleaned dataset described in Sec. 3.2. As shown in Table 1, the iTransformer exhibited superior performance compared to other models.

To more clearly demonstrate model performance differences, we analyzed the power generation data of string BT001-I001-DC001 over two sunny days and one cloudy or rainy day. As shown in Fig. 2, the blue line represents actual power generation, while the orange lines indicate predicted power generation. The prediction results show that Reformer, Informer, and iTransformer accurately captured the daily variation pattern of power generation—an increase followed by a decrease during the daytime. In contrast, Transformer and Flowformer failed to grasp this trend. Among these models,

Table 1: Performance evaluation of different time series forecasting models

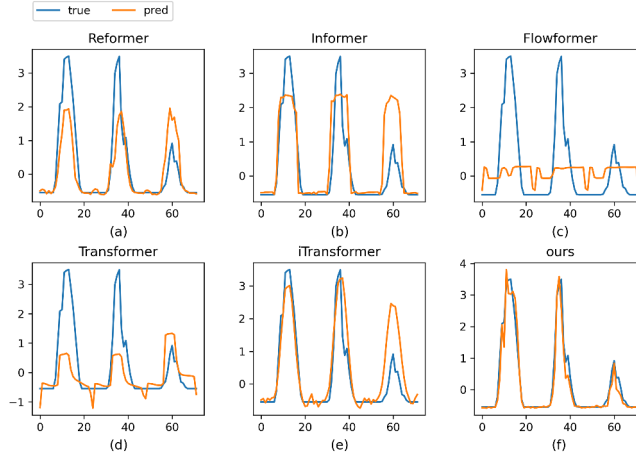| Model Type	MSE	MAE
iTransformer (Liu et al., 2024)	**0.268029**	**0.305569**
Reformer (Kitaev et al., 2020)	1.957360	1.125238
Informer (Zhou et al., 2021)	2.829479	1.156674
Flowformer (Wu et al., 2022)	1.222513	0.673625
Transformer (Vaswani et al., 2017)	1.894715	1.051079

Fig. 2: Comparison of power generation prediction performance of different models on string BT001-I001-DC001.

iTransformer's predictions on sunny days closely matched the actual power generation. While Reformer and Informer captured the overall trend, they failed to distinguish power generation characteristics under different weather conditions, leading to similar predicted peak values regardless of the weather. Notably, iTransformer's forecasting accuracy on cloudy and rainy days was suboptimal, indicating room for improvement in its adaptability to different weather conditions.

After a comprehensive evaluation, we select the iTransformer model due to its superior performance and apply it to forecast the power generation.

4 Design requirement

Our goal is to assist PV experts in understanding the dynamics of PV time series forecasting models and to streamline the process of model evaluation and analysis. To comprehensively assess the analytical requirements for these models, we collaborated with three domain experts, EA, EB, and EC. Both EA and EB are seasoned researchers with over a decade of expertise in energy studies. EC is an expert with five years of specific experience in PV research. We progressively identified the design

requirements and built design prototypes through a careful review of relevant literature and weekly online meetings with experts spanning across three months. Through ongoing engagement, we gained insights into the methodologies and challenges these experts face when analyzing time series forecasting models. These insights were crucial in formulating our design requirements.

R1: Enable comprehensive presentation and analysis of multivariate time series datasets. Experts require a clear understanding of both the operational status and geographical distribution of various PV strings.While time series data visualization is crucial, providing an overview of PV string information enhances experts' quick comprehension of the data. The system should facilitate rapid identification and precise location of diverse PV strings, enabling efficient comparison and analysis across these strings. This capability is essential for effectively identifying variations in model performance and uncovering potential issues.

R2: Implement multi-perspective model assessment capabilities. Experts aim to evaluate the predictive accuracy of time series forecasting models across different PV strings while assessing overall model performance. To facilitate comprehensive model analysis, the system should incorporate an anomaly detection feature that automatically identifies and displays anomalies in forecast results. This functionality will assist users in exploring instances of model inefficiency and provide direction for subsequent troubleshooting and improvement efforts.

R3: Investigate the factors influencing model forecast results. Model interpretability tools can assist the experts in comprehending the model's behavior and decision-making processes. These tools should facilitate the examination of relationships between data from various PV strings and their corresponding forecast outcomes. Furthermore, the system should provide a user-friendly configuration interface, allowing users to modify both the model and datasets. This feature will enable the assessment of how various factors impact model performance, enhancing overall model understanding and optimization.

R4: Facilitate the incorporation of expert knowledge and models optimization. The system should provide interactive tools that allow experts to incorporate their domain knowledge into the model, including capabilities for structural modifications. Additionally, a snapshot feature is essential to enable efficient explorationand comparison of various model configurations. This will support experts in identifying and retaining the most effective model setups, streamlining the optimization process.

5 Model

Based on the aforementioned requirements, we present an analysis framework for time series forecasting models in PV power generation and design the interactive visual analytics system. Fig. 3 illustrates both the system architecture and the model analysis pipeline. In the following sections, we detail the four core modules of the framework, including time series forecasting, PV string clustering analysis, model feature importance computation and anomaly detection in forecasting results.
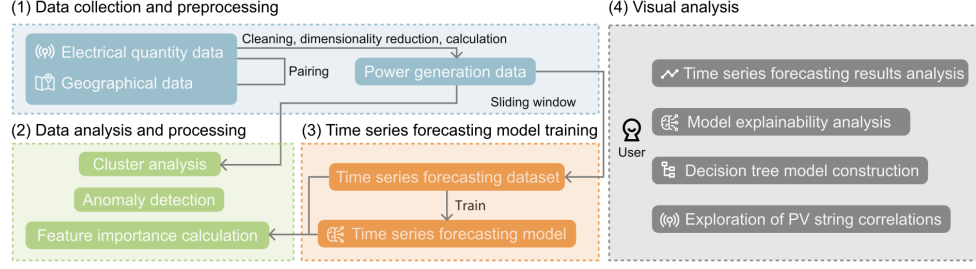
Fig. 3: The system overview that includes four part: (1) data collection and preprocessing, (2) data analysis and processing, (3) time series forecasting model training, and (4) visual analysis.

5.1 Time series forecasting

Deep time series forecasting models are inherently opaque, exhibiting black-box characteristics that make it challenging for experts to grasp the decision-making processes underlying their predictions or to apply their domain knowledge for model refinement. Decision tree models possess clear structures that are easily understandable and interpretable, enabling the extraction of decision rules from data characteristics. Decision tree regression is frequently employed for predictions involving complex data and facilitates expert application of domain knowledge. Inspired by hybrid models, we propose an innovative forecasting approach that combines a deep time series forecasting model with a decision tree regression model.

The methodology comprises the following steps:

1. Construct a deep time series forecasting model utilizing historical power generation data. Train and validate the deep time series forecasting model on the training dataset, with detailed training configurations specified in Sec. 3.3.
2. Build and train a decision tree regression model using sampled forecast outputs from the time series model, along with relevant covariates. Specifically, we extract 100 data points from the forecast power generation of each string, complemented by associated temporal and environmental covariates, to form the decision tree dataset. The feature details are provided in Table 2.
3. Combine the deep time series forecasting model and the decision tree regression model obtained from steps 1 and 2 to achieve the integrated prediction, where the output of the deep time series forecasting model serves as input to the decision tree, ultimately generating the final power generation forecasts.

Using all temporal and environmental features to construct the decision tree training set, without further optimization, results in a MSE of 0.148532 on the test set. This represents an improvement of approximately 44.58% over the iTransformer-only method, thus validating the effectiveness of our approach.

As shown in Fig. 2(f), our approach not only accurately forecasted the power generation during the night but also effectively captured the power generation characteristics on cloudy and rainy days, significantly enhancing the model's predictive capability under complex weather conditions.

Furthermore, we developed an interactive decision tree construction function leveraging scikit-learn that enables domain experts to optimize PV string power forecasting models through feature selection. The function automates model construction and deployment, encompassing four core functionalities:

Manual construction. While traditional decision regression trees are constructed by recursively identifying optimal partition points with user-defined methods and parameters, this process often remains opaque to users. Our algorithm enables users to manually create split nodes, select leaf nodes, and specify attributes and boundaries of partition points to constrain the tree structure. This manual approach allows users to leverage their domain knowledge fully, enhancing the effectiveness of decision tree construction.

Automatic expansion. Recognizing the limitations of domain knowledge, our algorithm offers an automatic expansion feature. When users are unable to clearly define attributes or boundaries for partition points, this functionality builds upon the user-initialized decision tree to generate a comprehensive tree structure.

Modify split points. After the construction of the decision tree, users can apply domain knowledge to change the split attributes and boundary values of nodes, further enhancing the model's performance. Following any modifications to the decision tree, the system will automatically expand and optimize the affected leaf nodes.

Pruning. As decision trees are constructed solely based on the training dataset, excessively tall trees may lead to overfitting. Moreover, the interpretability of a decision tree decreases as its complexity increases. Our system empowers users to prune subtrees, either by removing low-accuracy branches or by reducing overall tree complexity, thus mitigating the risk of overfitting and enhancing model interpretability.

5.2 Clustering analysis

In the operation of PV power stations, electrical time series data can form clusters with physical significance (Liu et al., 2019), revealing potential patterns within operational data. To investigate the correlations in power generation among PV strings, we employ a k-means algorithm based on Dynamic Time Warping (DTW) for clustering. DTW is a technique for measuring distances between time series that is widely used to evaluate their similarity.

DTW aligns time series and accurately determines distances between sequences, effectively overcoming limitations of Euclidean distance such as non-linear distortions (Ali et al., 2019). We use DTW to quantify similarity and apply k-means clustering to categorize PV strings. To improve computational efficiency, the clustering process utilizes resampled power generation data, reducing each sequence to 100 data points. PV strings can operate in various states, including normal operation, shading, open circuit conditions, dust accumulation, hot spots, and diode failures. Based on expert insights and the elbow method, we determined that setting the number of clusters to 6 optimally distinguishes between different power generation patterns of

11

PV strings. In the visual analytics system, we use a unified color coding to distinguish the categories of PV strings identified through clustering.

5.3 Feature importance

To gain a deeper understanding of the model's decision-making mechanism and the contribution of different input sequences, this study employs sensitivity analysis to compute feature importance. This method quantifies the extent to which input features impact prediction outcomes, assigning importance scores to each feature and thereby revealing the causal relationships between inputs and outputs. Specifically, in time series analysis, the model's decision process often exhibits dynamic characteristics. By tracking the temporal relationships between hidden states and input data (Strobelt et al., 2018, 2019), we can capture the time-varying characteristics of feature importance, thereby revealing the model's dynamic decision patterns. This analytical approach not only identifies key features but also elucidates their impact across different time steps.

In deep learning, gradients are commonly used to indicate the sensitivity of model outputs to input data (Liang et al., 2021), facilitating feature-level analysis to evaluate the contribution of different features. Gradient-based sensitivity analysis effectively measures the model's output sensitivity to input variations, providing reliable metrics for feature importance assessment. Based on this principle, this paper defines two key metrics derived from local gradients: string importance (WS) and temporal importance (WT). Here, WS measures the overall impact of a specific string's power generation sequence on the overall prediction, while WT characterizes the dynamic changes in contribution of a specific string at different time points, thereby providing a quantitative basis for analyzing model behavior.

For a given output Y_i, we compute the gradient $W(Y_i, X_i)$ with respect to input sequence X_i:

$$W(Y_i, X_i) = \frac{\partial Y_i}{\partial X_i} \tag{3}$$

The gradients of $W(Y_i, X_i)$ have the same dimensionality as X_i. For each feature $x_{i,t}^k$, we denote its corresponding gradient as $w_{i,t}^k$, establishing a one-to-one correspondence between gradients and features, where $t \in [1, T]$ and $k \in [1, D]$. As depicted in Eq.(4) and Eq.(5), we calculate the importance of the PV string k as the cumulative sum of gradients across all sequences and time steps. This importance measure is utilized to assess the model's sensitivity to variations among different PV strings.

$$WS^k = \sum_{i \in [1,N]} ws_i^k, k \in [1, D] \tag{4}$$

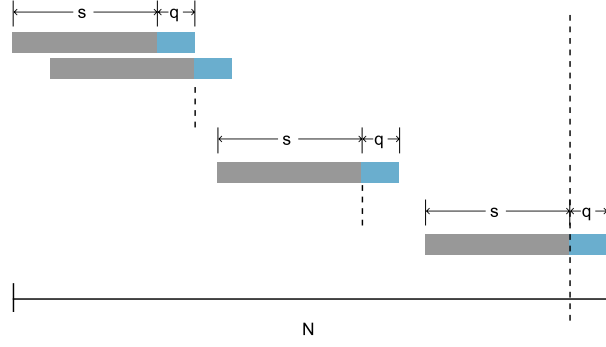$$ws_i^k = \sum_{i \in [1,N]} w_t^k, i \in [1, N] \tag{5}$$

Fig. 4: Overlap relationship between model inputs.

Based on the sliding window characteristics of the dataset (see Sec. 3.2), there exists an overlapping interval between adjacent input sequences. As shown in Fig. 4, assuming the length of the input sequence is s and the length of the output sequence as q. To quantify the importance of string k on the time scale, we construct a time bucket list $B^k = \{b_1, b_2, ..., b_{N-q}\}$ with length $N - q$. Each time bucket b_l collects all gradient values $w_{i,l}^k$ of string k corresponding to time l and renumbers them. Assuming there are m gradient values in b_l, the importance of PV string k at time l is:

$$WT_l^k = \frac{1}{m} \sum_{d \in [1,m]} |w_d^k|, w_d^k \in b_l \tag{6}$$

5.4 Anomaly detection

PV strings are characterized by highly variable power output, which can lead to significant deviations in time series forecasting models, potentially affecting their accuracy. The experts hope to explore the performance shortcomings of the predictive model and identify anomalous peaks in the prediction results. Unsupervised methods can identify these deviations based on statistical characteristics, offering strong interpretability. Additionally, their computational simplicity makes them well-suited for real-time processing of large-scale datasets. Drawing on the 3σ rule, we propose an unsupervised anomaly detection algorithm for PV power generation forecasting. The algorithm identifies significant deviations in predictions based on statistical characteristics and effectively utilizes large amounts of unlabeled data for anomaly detection. Beyond single time points, we also focus on persistent anomalies. By adjusting the threshold and persistence parameters, the algorithm can adapt to different operating conditions and anomaly definitions, demonstrating high flexibility. The algorithm is defined as follows:

13

$$r_i = |y_i - \hat{y_i}| \tag{7}$$

$$a_i = \begin{cases} 0, \text{ if } r_i \leq \mu + t \times \sigma \\ 1, \text{ if } r_i > \mu + t \times \sigma \end{cases} \tag{8}$$

Here, r_i represents the absolute deviation between the power generation y_i and the predicted value $\hat{y_i}$ of the current string at time i. Given a deviation sequence $R = \{r_1, r_2, ..., r_n\}$ of length n, we use its mean μ and standard deviation σ to construct threshold criteria, where the parameter t is the deviation multiplier with a default value of 3.

$$A_{ij} = \{a_i, a_{i+1}, ..., a_j\} \tag{9}$$
$$a_m = 1, m \in [i, j], j - i \geq s \tag{10}$$

When s (the default value is 10) or more consecutive points above the threshold are detected, the corresponding interval A_{ij} will be marked as an anomalous interval.

6 System overview

Based on the design requirements proposed in Sec. 4, we have designed and implemented a domain knowledge-driven visual analytics system for PV power forecasting. As shown in Fig. 5, the system consists of five views, including the data view (R1, R2), the detailed view (R1, R2, R3), the map view (R1), the compare view (R1), and the decision tree view (R4).

6.1 Data view (R1, R2)

The data view provides functionalities for selecting time series forecasting models and datasets, as well as choosing clustering algorithms, and displaying model performance and PV string information. By default, the system employs a DTW-based k-means algorithm with a predefined cluster count of 6. The clustering module incorporates various algorithms, including k-means, DBSCAN, and hierarchical clustering, while supporting customization of distance metrics and parameter settings. Users can tailor their choice of clustering algorithm with their specific data and requirements.

As depicted in Fig. 5A1 and Fig. 5A2, the data view utilizes bar charts to illustrate the performance distribution of time series forecasting models, and radar charts to display information for all PV strings in the dataset. The bar chart's x-axis represents MSE values in increments of 0.1, while the y-axis shows the number of strings within each MSE range. At the top of the radar chart, a color-coded horizontal line represents the different PV string categories described in Sec. 5.2. The radar chart is ideal for the presentation of multidimensional data, as illustrated in Fig. 5A3. Its clockwise-rotating axes correspond to various metrics: the discrete rate of string-associated inverters,
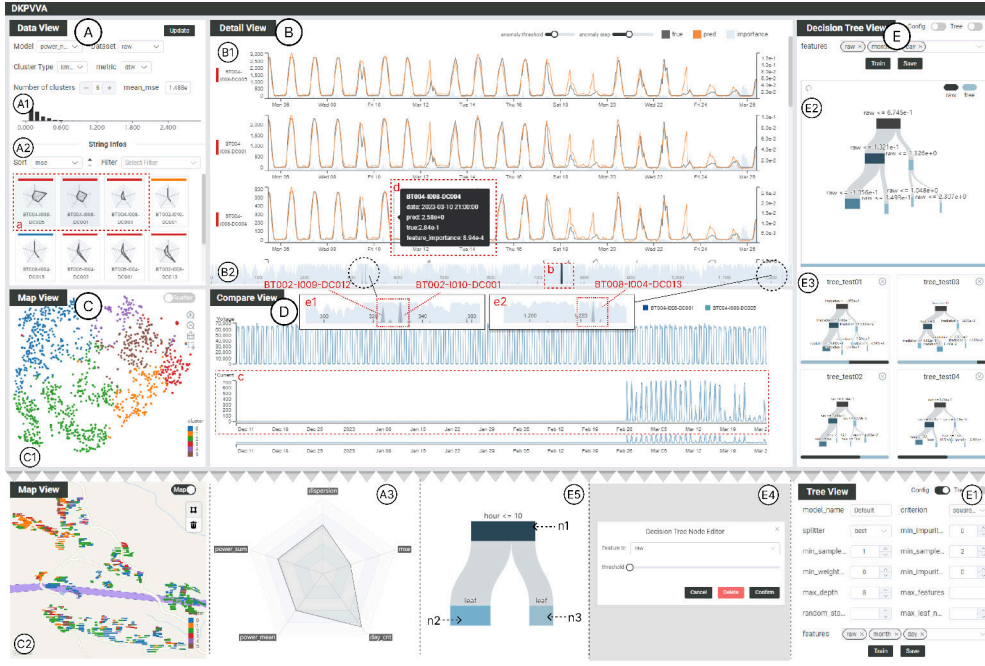
Fig. 5: System interface. (A) The data view provides model and dataset selection options along with an overview function; (B) the detail view displays power generation forecast results and string importance; (C) the map view visualizes the distribution of PV strings in geographical location and feature space; (D) the compare view displays historical time series of current and voltage for user-selected strings from the data view; (E) the decision tree view provides an interactive tool for constructing decision tree models, enabling experts to incorporate domain knowledge.

average MSE, number of days of power generation, average power generation, and total power generation. By using a uniform scale for all data ranges, from minimum to maximum values, the radar chart facilitates an comparison of data disparities among different strings.

The view provides sorting and filtering capabilities, allowing users to quickly identify PV strings of interest. When a string unit is clicked, experts can examine its historical performance in the compare view. As shown in Fig. 5a, the selected unit is highlighted with a background color that corresponds to the compare view.

6.2 Detail view (R1, R2, R3)

The detail view provides an in-depth analysis of the time series forecasting model. As shown in Fig. 5B, B1 displays the model's predictions for each PV string, arranged to correspond with the data view. Each string unit presents its name and category on the left, while featuring a line chart on the right. This chart illustrates both the forecasted (orange) and actual (dark gray) time series data. By overlaying an area chart on the

line chart, we can contextualize the string's importance over time. Specifically, the x-axis of the chart represents temporal information, while the left and right y-axes indicate the magnitude of power generation and temporal importance, respectively.

Fig. 5d shows that hovering over the line chart triggers a tooltip, displaying detailed data for the selected time point. An anomaly detection algorithm pinpoints forecast anomalies, while a slider at the top of the view enables users to adjust detection parameters. At the bottom, horizontal chart B2 visualizes string importance, with string numbers on the x-axis and importance levels on the y-axis. The chart uses color gradients to represent varying levels of importance, with darker shades indicating higher significance.

6.3 Map view (R1)

The map view provides two visualization modes: a scatter plot (Fig. 5C1) for the representation of PV strings in the feature space, and a map (Fig. 5C2) for their geographic distribution. Users can switch between these modes, with the scatter plot set as the default display.

We employ the t-SNE algorithm (Van der Maaten et al., 2008) to reduce the dimensionality of PV string power generation time series data and project it onto a two-dimensional scatter plot. The geographic view accurately depicts the spatial distribution of strings within the PV power station. String colors correspond to the categories identified through clustering as described in Sec. 5.2, allowing users to explore potential relationships between strings from both map and scatter plot perspectives. A brush selection tool enables users to highlight areas of interest on either view, with the system emphasizing the corresponding strings. Upon hovering over a specific string, the system provides visual highlighting and concurrently displays relevant information, including the string's name. Additionally, a zoom function facilitates focused, detailed analysis of string distribution patterns.

6.4 Compare view (R1)

Once a user selects a string of interest in the data view, the compare view in Fig. 5D displays historical time series for that string's current and voltage. Differently color-coded strings facilitate easy identification. Clicking the legend highlights the selected string's current and voltage sequences, facilitating easier comparison and inspection of string data. Current and voltage line charts share a common time axis, allowing users to zoom into specific time interval by selecting the area at the bottom. This feature enables detailed exploration of operational differences between strings. Similar to Fig. 5d, this view incorporates hover tooltips to reveal specific current and voltage values.

6.5 Decision tree view (R4)

The decision tree view offers capabilities for configuring model parameters, constructing decision trees, integrating forecasts, and saving models. As depicted in Fig. 5E1, a configuration button at the top toggles the display of the model settings panel. This

panel enables users to adjust various parameters, such as the tree's maximum depth, splitting criteria, and feature selection options.

Decision trees, inherently binary in nature, are visualized using flow diagrams. Each node is represented by a rectangle, with parent and child nodes connected by edges. Edge width indicates the volume of data flowing between nodes—wider edges signify more data points. As depicted in Fig. 5E5, labels above parent nodes specify splitting rules, illustrating how nodes are divided based on certain criteria. To manage visual complexity in cases with numerous nodes, the system incorporates a drag-and-drop functionality for manual layout adjustments. A default display limited to the top three layers of the tree structure. These features help mitigate potential overlap and visual confusion within the limited view space.

Table 2: Feature names and their meanings in the decision tree dataset

Feature name	Meaning
raw	Power generation output forecasted by time series forecasting model
month	Sampling month
day	Sampling date
weekday	Day of the week
hour	Sampling Hour
temperature	Temperature
wind_speed	Wind speed
irradiation	Solar irradiance

Inspired by BaobabView (Van Den Elzen and Van Wijk, 2011) and iDT (Sarailidis et al., 2023), we developed an automatic algorithm for decision tree generation using the scikit-learn framework (Pedregosa et al., 2011). This algorithm enables the manual construction of decision trees, their automatic expansion, the modification of split points, and the pruning process, as detailed in Appendix. It enables experts to incorporate domain-specific knowledge into the tree construction process. When a user clicks on any node, a node editing window appears (Fig. 5E4), allowing for modification of partitioning attributes and boundary values, or node removal for pruning.

Upon clicking the train button, the system automatically expands and train the decision tree model, generating integrated forecasts based on the current tree structure and configuration. The trained tree structure is then synchronized and updated in the flow diagrams. Bands below the flow diagram display the average MSE for both the standalone time series forecasting model and the combined forecasting algorithm with the decision tree model. Users can view power generation forecast results using the decision tree model algorithm in the detail view by clicking the enable button in the upper right corner. As illustrated in Fig. 5E3, the system facilitates exploration of various decision tree structures, with snapshots of user-saved models displayed in a list at the bottom of the view.

7 Evaluation

We conducted two case studies and an expert interview to demonstrate the effectiveness of the visual analytics system. The experts involved in our evaluation are domain experts which mentioned in Sec.4. We illustrate these case studies from the experts' perspective, exploring the understanding of model prediction behavior and model optimization.

7.1 Case study I: exploring model interpretability

To validate the effectiveness of the visual analytics system in interpreting model prediction behavior, we used the PV power generation dataset constructed in Sec. 3.2. Based on the experimental configurations detailed in Sec. 3.3, we trained the iTransformer model and attempted to conduct an in-depth analysis of its performance using the system.

The accuracy of time series forecasting models in forecasting the power generation for various strings serves as a critical metric for expert model assessment. After selecting the iTransformer model and the analysis dataset, the data view reveals an average MSE of approximately 148.8 across the test set. Fig. 5A1 illustrates that most PV strings have a predicted MSE between 0.1 and 0.2. This distribution suggests that there may be a subset of PV strings with exceptionally high errors in their power generation forecasts.

To investigate further, the expert focuses on examining PV strings with significant forecast errors. As illustrated in Fig. 5a, three PV strings categorized as type 3 (specifically BT004-I008-DC001, DC005, and DC004) exhibit exceptionally high MSEs of 107137.34, 73461.06, and 4483.16 respectively. Their power generation days are approximately half of the maximum possible. Additionally, PV strings BT002-I010-DC001 and BT008-I004-DC013, with lower but still notable MSEs of 1.42 and 1.26, demonstrate minimal cumulative power output.

Intrigued by these large forecast errors, the expert aims to identify their underlying causes through a sequential exploratory analysis of these two categories of strings. The expert began by comparing actual and forecasted power generation sequences for PV strings BT004-I008-DC001, DC005, and DC004 using the detail view. These strings exhibit a high degree of overlap in their sequences, suggesting minimal discrepancies between forecasted and actual power generation. However, closer examination revealed that the vertical scales of the power generation time series for these strings (0-2000, 0-3000, and 0-800, respectively, as shown in Fig. 5B1) differed markedly from the typical range of -1 to 5 observed in most strings.

To investigate further, the expert added these three PV strings to the compare view (see Fig. 5c). This revealed that current data for these strings was zero before February 26, 2023. As described in Sec. 3.2, the power generation data underwent standardization based on the training set during dataset construction. The expert hypothesized that significant differences in power generation between the training and test sets for these strings may have led to an abnormally large standardized value in the test set, resulting in an excessively high calculated MSE.
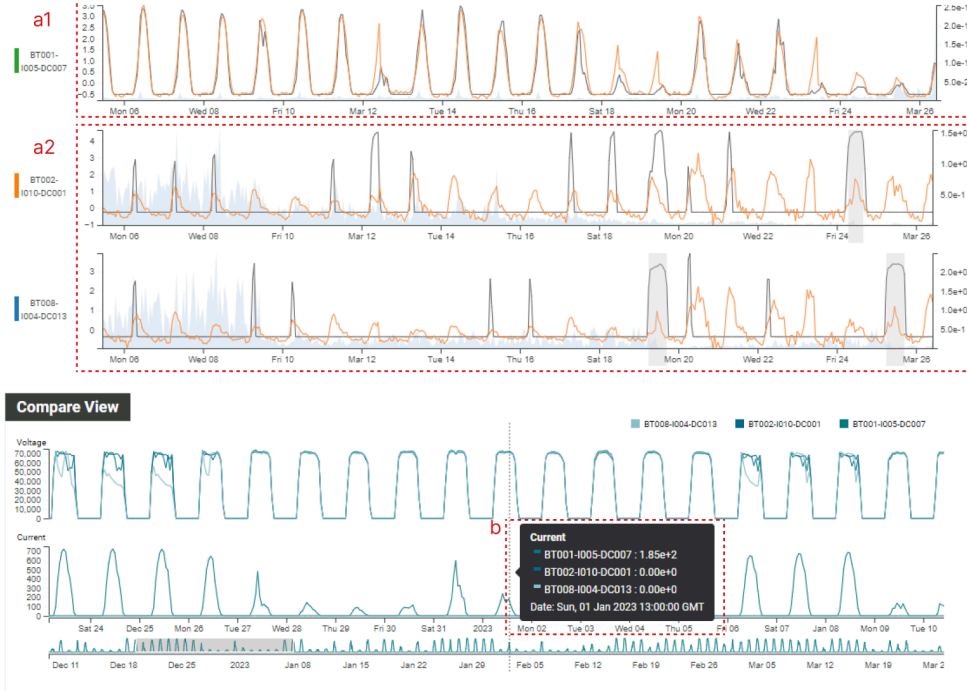
Fig. 6: The comparison of the forecasted power generation, historical current and voltage series of the string BT001-I005-DC007, BT002-I010-DC001, and BT008-I004-DC013.

To Address this issue, the expert re-standardized the data for these three PV strings using the test set. This significantly reduced their MSEs to 0.1464, 0.1424, and 0.1375, respectively. The analysis revealed that despite the data anomalies in these PV strings, the time series model's forecasting performance remained relatively accurate. This discovery highlights the iTransformer model's robust generalization capabilities, demonstrating its ability to apply patterns learned from other PV strings to forecast these outliers effectively.

Following this observation, the expert noted in Fig. 5B2 that the importance of various PV strings was closely aligned, with most values ranging from -1.54e-4 to -2.08e-5. However, a few strings exhibit notably high importance (as seen in Fig. 5b). Tooltip annotations reveals that the strings in the dark area were BT004-I008-DC004, DC005, and DC001, with importance values of 6.68e-4, 4.44e-4, and 2.77e-4, respectively. The expert noted that these high-importance string identifiers corresponded with the anomalous strings identified earlier. This suggests that the model may place greater emphasis on anomalous data, potentially making it more influential in the model's decision-making process than normal data.

The expert then examined the power generation forecast results for PV strings BT002-I010-DC001 and BT008-I004-DC013 through the detail view (as depicted in

19

Fig. 6a2). The anomaly detection algorithm identified and highlighted the anomalous intervals where the actual power generation of these strings exhibited sharp fluctuations, deviating significantly from the forecasted values. Meanwhile, the string with the lowest MSE, BT001-I005-DC007, exhibited a markedly different power generation pattern (as illustrated in Fig. 6a1).

For a more detailed analysis, the expert incorporated data from these three PV strings into the compare view (as depicted in Fig. 6b). At 13:00 on January 1, 2023, the operating current of BT001-I005-DC007 was measured at 0.0185, while BT002-I010-DC001 and BT008-I004-DC013 had zero operating currents. Moreover, the current values for these two strings remained near zero throughout the observation period. By examining the historical voltage time series, the expert noted similar voltage patterns among the three strings. This led to the inference that PV strings BT002-I010-DC001 and BT008-I004-DC013 might be unable to generate power normally, possibly due to damage, shading, or other issues.The expert also suspected a potential open circuit fault affecting their operation.

The expert then conducted a detailed zoom analysis on the horizontal line chart. This analysis revealed that the feature importance of PV strings BT002-I010-DC001 and BT008-I004-DC013 was notably higher than that of other strings, at 1.28e-4 and 8.64e-5 respectively (as shown in Fig. 5e1, e2). Based on prior analysis experience, the expert quickly identified string BT002-I009-DC012 in the data view, which also exhibited a high feature importance of 6.59e-5. Comparative analysis revealed that its current and voltage patterns were similar to those of BT002-I010-DC001 and BT008-I004-DC013. This finding supported the expert's hypothesis that the model is more sensitive to anomalous data than to normal data. Furthermore, experts observed that the temporal importance of PV string exhibits increasing importance with temporal proximity, indicating enhanced model sensitivity to recent temporal data.

After this analysis, the expert identified various patterns in the PV strings' power generation. As a result, he decided to further investigate the distribution of the power generation data. The expert randomly selected areas within the map view and analyzed the data for strings categorized as 0 (blue) and 3 (red). Fig. 7 shows significant differences in the radar charts among strings of different categories, while strings within the same category display similar patterns. Furthermore, the expert noted that the historical time series of current and voltage for category 3 strings also showed a lack of current during the initial operational phase. This observation suggests that the clustering algorithm effectively identifies strings with similar power generation patterns.

As shown in Fig. 5C2, PV strings in the upper left corner of the PV power station are mainly classified as category 1 (orange), while those on the far right are classified as category 2 (green). This spatial distribution of categories suggests a correlation: geographically adjacent PV strings often exhibit similar power generation profiles.

7.2 Case study II: exploring model performance optimization

Based on the analysis of case I, the expert identified anomalies in the PV string power generation data that significantly affected the model's performance. As a result, the expert eliminated strings showing standardized anomalies and those with negligible
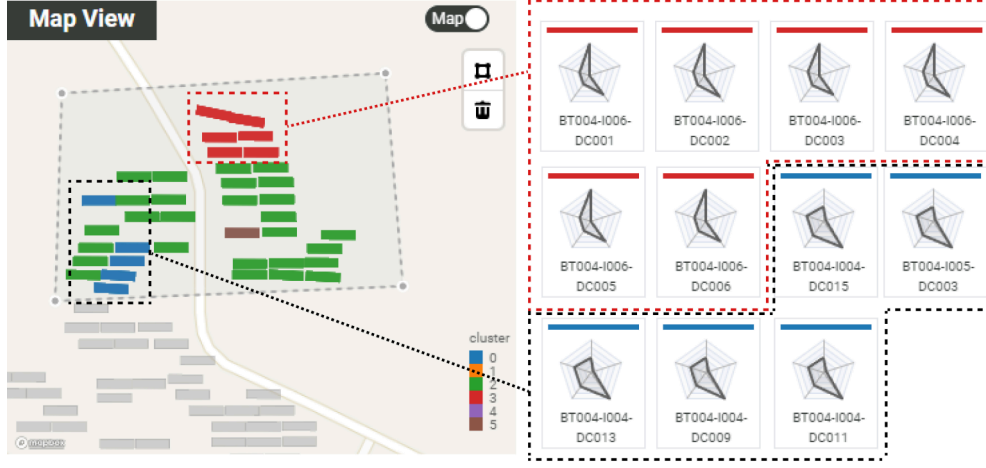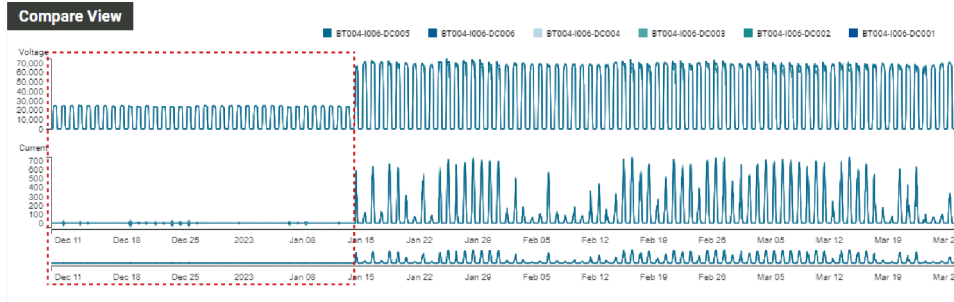
Fig. 7: Data analysis of PV strings within the selected area. (a) Strings of category 3 and category 0 show different radar chart patterns; (b) Strings of category 3 all exhibit a lack of power generation.

power generation. Subsequently, the dataset and model were reconstructed following the procedures outlined in Sec. 3. Fig. 8 illustrates the new model's performance, showing an reduced average MSE of 0.268, with most string MSEs ranging between 0.1 and 0.2. Upon sorting by descending MSE, the expert identified string BT006-I004-DC002 as having the highest MSE, approximately 1.29.

Fig. 9c shows a detail view with a horizontal bar chart displaying uniform color distribution, indicating that the new model maintains consistent sensitivity across different strings, without bias towards any particular string. This consistency suggests that the model processes all input data uniformly. In summary, eliminating anomalous data not only minimized prediction errors but also improved the model's balanced forecasts across various strings, resulting in more reliable forecasting outcomes.
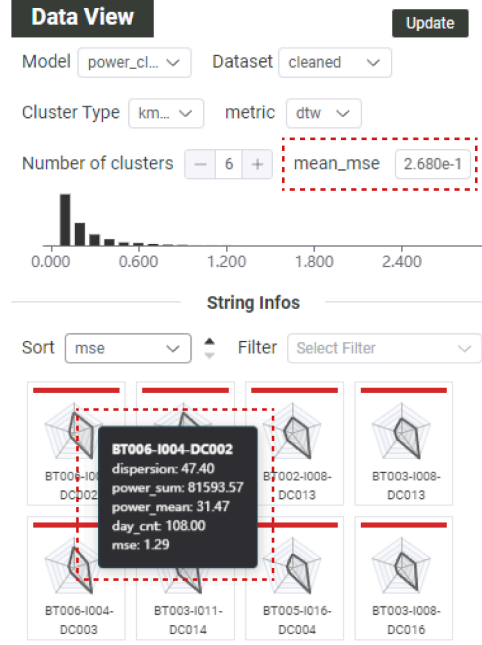
Fig. 8: The data view snapshot for case II illustrates the model's average MSE and the MSE specific to the string BT006-I004-DC002.

To gain a deeper understanding of the time series forecasting model's performance variability across different time scales, the expert meticulously analyzed the model's predicted power generation against actual values for various strings. It was observed that during certain time frames, the model's predictive accuracy uniformly declined across different strings. Fig. 9 shows three time intervals (labeled a1, a2, and a3) where a simultaneous trough in the actual power generation of the strings was observed, accompanied by significant deviations in the model's forecasts. This consistent and widespread low power generation is likely due to overcast and rainy weather conditions. The expert's analysis indicates substantial potential for improving the model's ability to capture PV power generation volatility.

Furthermore, the expert found that the time series forecasting model occasionally produces forecasts inconsistent with domain knowledge. As depicted in Fig. 9b, from 6 PM on March 6, 2023, to 6 AM the next day, PV strings should cease power generation after sunset, correctly represented by a flat line in the actual power generation data. However, the model's forecast for this period incorrectly displays an oscillating pattern. This observation highlights the time series forecasting model's limitations in comprehending the physical processes of PV power generation, highlighting the necessity for further model refinement and the incorporation of domain knowledge.

The expert employed a method that incorporates decision trees to improve the precision of forecasts. Initially, the expert concentrated on temporal features, selecting five features—*raw*, *month*, *day*, *weekday*, and *hour*—using a feature selector. With
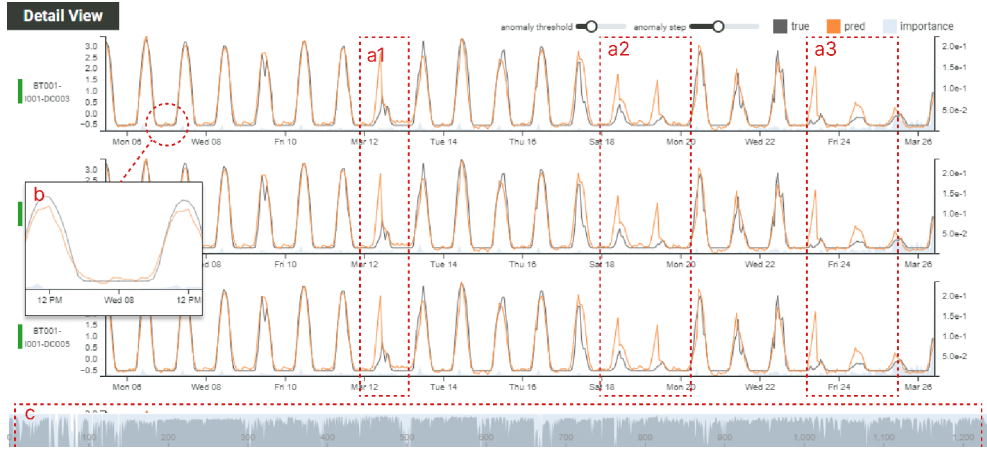
Fig. 9: Analysis of the forecast results of the deep time series forecasting model in case II. (a) The model's forecast results have large deviations under overcast and rainy weather conditions; (b) When there is no solar radiation, the model's forecast results that do not conform to domain knowledge; (c) The distribution of feature importance of PV strings.

these, the expert constructed a decision tree named *time_tree*. However, the integrated forecasts' MSE was approximately 0.3982, falling short of expectations. For deeper analysis, the expert examined the integrated forecast results in the detail view (as shown in Fig. 10a2). The expert observed that while previous discrepancies with domain knowledge were addressed, significant deviations persisted in overall model forecasts.

Reviewing the *time_tree*'s first three layer (Fig. 10a1) revealed that the initial two layers split on the raw predicted power generation, with thresholds of 0.674, 0.134, and 1.37. The third layer used both *hour* and *raw*, with split values of 6.5, 16.5, 15.5, and 2.345.This structure reflects the intricate relationship between PV string power generation and solar irradiance, which peaks at noon and diminishes during sunrise and sunset. The expert refined the tree's third layer, setting the the second and third nodes' split values to 18 to better reflect dusk generation characteristics. After retraining, the MSE improved to 0.3703, though further enhancement remained possible. The expert deduced that the current feature set might not fully capture PV power generation domain knowledge, necessitating additional refinement.

Drawing from the previous insights, the expert developed another model, *all_tree*, incorporating all available features. The first three layers of this model (as shown in Fig. 10b1) primarily use the irradiation feature for node splits. Consequently, the model's MSE significantly dropped to about 0.1485, signifying a marked improved forecast accuracy. Comparing Fig. 10a2 and b2 reveals that *all_tree* has more effectively captured power generation temporal characteristics during cloudy and rainy conditions than *time_tree*. The expert saved this decision tree in the bottom list for further investigation and refinement.
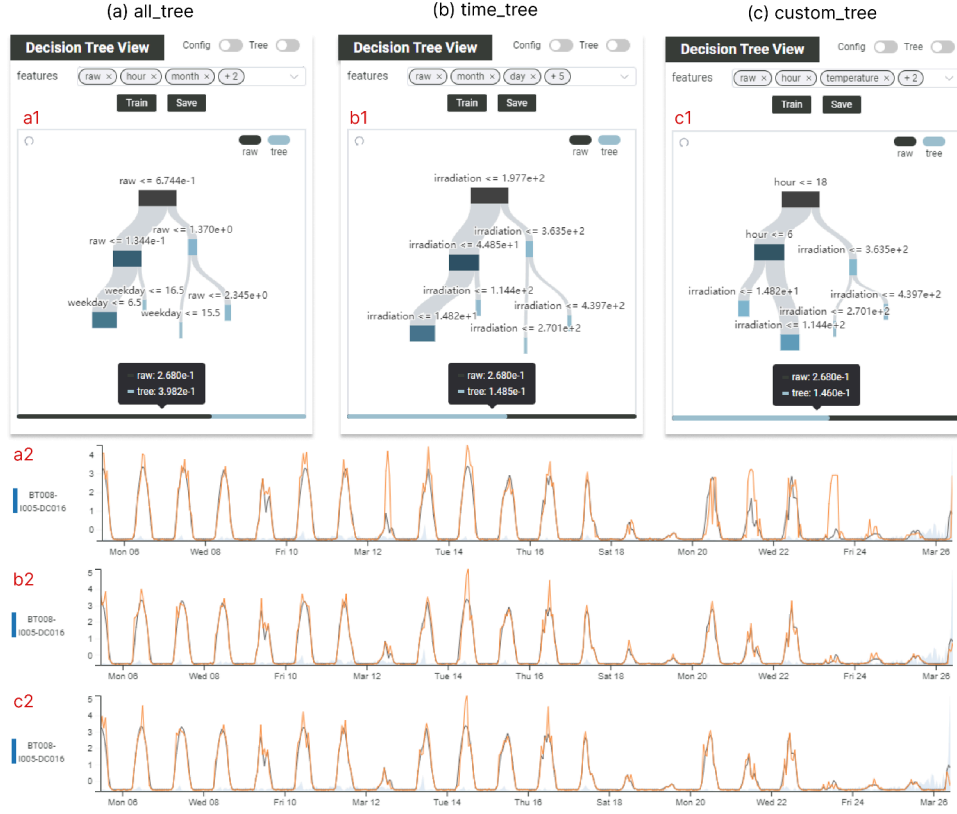
23

Fig. 10: Decision tree model construction and prediction results. (1) a1, b1, c1 are the construction results of *time_tree*, *all_tree*, and *custom_tree*, respectively; (2) a2, b2, c2 are the forecast results obtained by applying *time_tree*, *all_tree*, and *custom_tree*, respectively.

Based on domain expertise in PV power generation, key features directly affecting solar irradiance were identified: *raw, hour, temperature, wind_speed,* and *irradiation.* These variables were used to construct the decision tree model, which named *custom_tree.* The expert manually designed its initial structure, using *hour* as the splitting feature with thresholds of 18 and 6 for the root and left subtree nodes, respectively. These splits reflect the strong correlation between solar irradiance and time, with 6 AM and 6 PM marking sunrise and sunset. The model's initial three layers after training are shown in Fig. 10c1, achieving an MSE of 0.1460.

To evaluate and compare the three models, the expert randomly selected PV string BT008-I005-DC016. As observed in Fig. 10a2, b2, and c2, all models showed marked improvement in predictive performance during periods of solar irradiance absence, with the *custom_tree* model exhibiting optimal MSE performance.

7.3 Expert feedback

Following the case analysis, we conducted one-on-one structured interviews with three previously mentioned PV experts. We began by outlining the system's core capabilities, demonstrating its practical applications, and illustrating the processes of exploring deep time series forecasting models and constructing decision tree models. We then invited the experts to use the system to pinpoint model performance deficiencies, investigate strings of interest, and analyze the performance disparities of time series forecasting models across these strings. We asked the experts to verbalize their thoughts while interacting with the system and meticulously documented their observations. Finally, we sought the experts' feedback on the system's design and usability.

The experts' feedback is summarized as follows:

- Regarding system usability, the experts first acknowledged the detail view's positive impact on understanding the behavior of time series forecasting models. EA stated, "The process of exploring and analyzing power generation time series is straightforward and efficient", emphasizing how the detail view significantly enhances performance analysis at the string level. EC also acknowledged the value of diagrams illustrating string importance, noting that they provide targeted insights that effectively guide dataset and model optimization. The integration of decision trees was also praised, as this feature enables experts to combine their domain knowledge with time series forecasting models, thus improving the accuracy of power generation forecasts.
- From a visual design standpoint, the system uses line charts, radar charts, and maps to display string current and and voltage time series, multi-dimensional string characteristics, and geographical strings distribution, respectively. Experts found these visualizations simple and intuitive, striking a balance between ease of use and comprehensibility, while facilitating the comparison of multi-dimensional data across different strings. Furthermore, the system's comprehensive annotation features have significantly enhanced the user experience.

Unlike traditional forecasting model construction methods, this work improves the interpretability of time series forecasting models, allowing experts to calibrate them more accurately. In addition to the aforementioned feedback, experts have provided suggestions to address the system's limitations:

- Experts acknowledge the importance of clustering in identifying distinct patterns in string power generation. However, the current design lacks an intuitive display of patterns for each category, hindering users' comprehension of various power generation modes. Furthermore, EB suggests that the geographical information in the map view is underutilized, recommending "a more in-depth analysis of the correlation between the map and clustering results within the map view."
- Currently, the system only displays the top three layers of the decision tree structure. While the system supports manual decision tree construction, the limited context constrains the full application of domain knowledge. Experts suggest implementing functionalities such as collapse and expansion to allow users to view additional

layers of the decision tree structure when necessary. Moreover, they propose that the system offer pre-defined parameter configuration templates or tree structures to help users quickly select appropriate decision tree configurations, thereby improving the efficiency of decision tree construction.

- The system lacks support for real-time modifications to the dataset or training capabilities of deep time series forecasting models. Updating the dataset and rebuilding the model based on analysis results typically requires 2 to 3 hours. Experts suggest that integrating an automated training process for deep time series forecasting modules could enhance model construction efficiency and improve the system's practical applicability.

8 Conclusion

This study introduces an innovative PV forecasting approach that integrates a deep time series forecasting model with decision tree models, significantly improving prediction accuracy. Building on this foundation, we have developed a domain-knowledge-driven visual analytics system for PV time series forecasting. The system's effectiveness in evaluating and optimizing forecasting models has been empirically validated through rigorous case studies and user experiments.

While the current system demonstrates PV forecasting and analysis capabilities in controlled settings, we acknowledge several practical implementation challenges. Real-world PV monitoring systems frequently encounter sensor malfunctions, data transmission errors, and quality issues that require robust handling mechanisms. Future work should focus on developing automated data quality assessment and correction pipelines to handle missing data, sensor failures, and configuration errors. Additionally, integration with existing PV monitoring infrastructure would benefit from standardized data interfaces and error handling protocols. These enhancements would improve the system's practical applicability while maintaining its core forecasting and analytical capabilities.

Moving forward, we aim to continually enhance and refine the user experience. Planned improvements include implementing interactive connections between the map and compare views, enabling users to add strings to the compare view through direct selection, thereby reducing operational complexity. We also intend to incorporate visual representations of power generation patterns, such as radar charts integrated with data clusters, to enhance the semantic richness of the map view and strengthen support for decision tree construction. Moreover, future research should focus on developing an incremental learning mechanism that enables continuous model updates using streaming data, allowing the forecasting system to automatically adapt to PV array degradation and other temporal changes in system performance, thereby maintaining prediction accuracy over extended operational periods. Finally, a key focus of our future research is to explore the extension of this system's capabilities to a wider range of time series forecasting scenarios.

References

Ali, M., Alqahtani, A., Jones, M.W., Xie, X.: Clustering and Classification for Time Series Data in Visual Analytics: A Survey. IEEE Access **7**, 181314–181338 (2019) https://doi.org/10.1109/ACCESS.2019.2958551

Abdel-Nasser, M., Mahmoud, K.: Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Computing and Applications **31**, 2727–2740 (2019) https://doi.org/10.1007/s00521-017-3225-z

Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D.: A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renewable and Sustainable Energy Reviews **124**, 109792 (2020) https://doi.org/10.1016/j.rser.2020.109792

Bogl, M., Aigner, W., Filzmoser, P., Lammarsch, T., Miksch, S., Rind, A.: Visual Analytics for Model Selection in Time Series Analysis. IEEE Transactions on Visualization and Computer Graphics **19**(12), 2237–2246 (2013) https://doi.org/10.1109/TVCG.2013.222

Hohman, F., Kahng, M., Pienta, R., Chau, D.H.: Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. IEEE Transactions on Visualization and Computer Graphics **25**(8), 2674–2693 (2019) https://doi.org/10.1109/TVCG.2018.2843369

Hao, J., Shi, Q., Ye, Y., Zeng, W.: TimeTuner: Diagnosing Time Representations for Time-Series Forecasting with Counterfactual Explanations. IEEE Transactions on Visualization and Computer Graphics **30**(1), 1183–1193 (2024) https://doi.org/10.1109/TVCG.2023.3327389

Jin, S., Lee, H., Park, C., Chu, H., Tae, Y., Choo, J., Ko, S.: A Visual Analytics System for Improving Attention-based Traffic Forecasting Models. IEEE Transactions on Visualization and Computer Graphics **29**(1), 1102–1112 (2023) https://doi.org/10.1109/TVCG.2022.3209462

Kitaev, N., Kaiser, L., Levskaya, A.: Reformer: The efficient transformer. In: Proceedings of the International Conference on Learning Representations (2020). https://openreview.net/forum?id=rkgNKkHtvB

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: itransformer: Inverted transformers are effective for time series forecasting. In: Proceedings of the International Conference on Learning Representations (Spotlight) (2024). https://openreview.net/forum?id=JePfAI8fah

Lai, C., Li, J., Chen, B., Huang, Y., Wei, S.: A Survey on Photovoltaic Power Output Forecasting Techniques. Proceedings of the Chinese Society of Electrical Engineering **34**(6), 1201–1217 (2019) https://doi.org/10.19595/j.cnki.1000-6753.tces.180326

Liang, Y., Li, S., Yan, C., Li, M., Jiang, C.: Explaining the black-box model: A survey of local interpretation methods for deep neural networks. Neurocomputing **419**, 168–182 (2021) https://doi.org/10.1016/j.neucom.2020.08.011

Liu, G., Zhu, L., Wu, X., Wang, J.: Time series clustering and physical implication for photovoltaic array systems with unknown working conditions. Solar Energy **180**, 401–411 (2019) https://doi.org/10.1016/j.solener.2019.01.041

Muhlbacher, T., Linhardt, L., Moller, T., Piringer, H.: TreePOD: Sensitivity-Aware Selection of Pareto-Optimal Decision Trees. IEEE Transactions on Visualization and Computer Graphics **24**(1), 174–183 (2018) https://doi.org/10.1109/TVCG.2017.2745158

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

Phan, Q.-T., Wu, Y.-K., Phan, Q.-D.: An Approach Using Transformer-based Model for Short-term PV generation forecasting. In: Proceedings of the International Conference on Applied System Innovation (ICASI), pp. 17–20 (2022). https://doi.org/10.1109/ICASI55125.2022.9774491

Qu, J., Qian, Z., Pei, Y.: Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. Energy **232**, 120996 (2021) https://doi.org/10.1016/j.energy.2021.120996

Sun, D., Feng, Z., Chen, Y., Wang, Y., Zeng, J., Yuan, M., Pong, T.-C., Qu, H.: DFSeer: A Visual Analytics Approach to Facilitate Model Selection for Demand Forecasting. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2020). https://doi.org/10.1145/3313831.3376866

Strobelt, H., Gehrmann, S., Behrisch, M., Perer, A., Pfister, H., Rush, A.M.: Seq2seq-Vis: A visual debugging tool for sequence-to-sequence models. IEEE Transactions on Visualization and Computer Graphics **25**(1), 353–363 (2019) https://doi.org/10.1109/TVCG.2018.2865044

Strobelt, H., Gehrmann, S., Pfister, H., Rush, A.M.: LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. IEEE Transactions on Visualization and Computer Graphics **24**(1), 667–676 (2018) https://doi.org/10.1109/TVCG.2017.2744158

Streeb, D., Metz, Y., Schlegel, U., Schneider, B., El-Assady, M.: Task-based visual interactive modeling: Decision trees and rule-based classifiers. IEEE Transactions on Visualization and Computer Graphics **28**(9), 3307–3323 (2022) https://doi.org/10.1109/TVCG.2020.3045560

Sharda, S., Singh, M., Sharma, K.: RSAM: Robust Self-Attention Based Multi-Horizon Model for Solar Irradiance Forecasting. IEEE Transactions on Sustainable Energy **12**(2), 1394–1405 (2021) https://doi.org/10.1109/TSTE.2020.3046098

Shen, Q., Wu, Y., Jiang, Y., Zeng, W., Lau, A.K.H., Vianova, A., Qu, H.: Visual Interpretation of Recurrent Neural Network on Multi-dimensional Time-series Forecast. In: Proceedings of the IEEE Pacific Visualization Symposium (PacificVis), pp. 61–70 (2020). https://doi.org/10.1109/PacificVis48177.2020.2785

Sarailidis, G., Wagener, T., Pianosi, F.: Integrating scientific knowledge into machine learning using interactive decision trees. Computers & Geosciences **170**, 105248 (2023) https://doi.org/10.1016/j.cageo.2022.105248

Teoh, S.T., Ma, K.-L.: PaintingClass: Interactive construction, visualization and exploration of decision trees. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 667–672 (2003). https://doi.org/10.1145/956750.956837

Teoh, S.T., Ma, K.-L.: StarClass: Interactive Visual Classification Using Star Coordinates. In: Proceedings of the SIAM International Conference on Data Mining, pp. 178–185 (2003). https://doi.org/10.1137/1.9781611972733.16

Van Den Elzen, S., Van Wijk, J.J.: BaobabView: Interactive construction and analysis of decision trees. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 151–160 (2011). https://doi.org/10.1109/VAST.2011.6102453

Van der Maaten, Laurens and Hinton, Geoffrey: Visualizing data using t-SNE. Journal of Machine Learning Research **9**(86), 2579–2605 (2008)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30, pp. 6000–6010 (2017)

Wu, H., Wu, J., Xu, J., Wang, J., Long, M.: Flowformer: Linearizing transformers with conservation flows. In: Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 24226–24242 (2022)

Xu, K., Yuan, J., Wang, Y., Silva, C., Bertini, E.: mTSeer: Interactive Visual Exploration of Models on Multivariate Time-series Forecast. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2021).

https://doi.org/10.1145/3411764.3445083 . Article No. 23

Zhang, J., Verschae, R., Nobuhara, S., Lalonde, J.-F.: Deep photovoltaic nowcasting. Solar Energy **176**, 267–276 (2018) https://doi.org/10.1016/j.solener.2018.10.024

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325