Engineering 51 (2025) 158–170

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Power Systems Proactively Responding to Climate Change—Article

Frequent Pattern Growth-Based Identification of Critical Lines in Cascading Failures for Renewable-Dominant Hybrid AC/DC Power Systems

Tianhao Liu, Jiongcheng Yan, Yutian Liu *

Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan 250061, China

ARTICLE INFO

Article history:
Received 29 February 2024
Revised 12 June 2025
Accepted 25 June 2025
Available online 8 July 2025

Keywords:
Cascading failure
Risk assessment
Frequent pattern
Hybrid AC/DC power system
Renewable energy

ABSTRACT

In wind and solar renewable-dominant hybrid alternating current/direct current (AC/DC) power systems, the active power of high-voltage direct current (HVDC) system is significantly limited by the security and stability events caused by cascading failures. To identify critical lines in cascading failures, a rapid risk assessment method is proposed based on the gradient boosting decision tree (GBDT) and frequent pattern growth (FP-Growth) algorithms. First, security and stability events triggered by cascading failures are analyzed to explain the impact of cascading failures on the maximum DC power. Then, a cascading failure risk index is defined, focusing on the DC power being limited. To handle the strong nonlinear relationship between the maximum DC power and cascading failures, a GBDT with an update strategy is utilized to rapidly predict the maximum DC power under uncertain operating conditions. Finally, the FP-Growth algorithm is improved to mine frequent patterns in cascading failures. The importance index for each fault in a frequent pattern is defined by evaluating its impact on cascading failures, enabling the identification of critical lines. Simulation results of a modified Ningxia–Shandong hybrid AC/DC system in China demonstrate that the proposed method can rapidly assess the risk of cascading failures and effectively identify critical lines.

© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

To achieve carbon-neutrality goals, large-scale renewable energy is transmitted to load centers through line-commutated converter (LCC)-based high-voltage direct current (HVDC). However, due to the high uncertainty and susceptibility of wind- and solar-powered renewable energy devices, these systems are prone to cascading failures under external disturbances [1], such as extreme weather. Severe cascading failures threaten the safe and stable operation of the system, potentially leading to large-scale blackouts [2]. Therefore, cascading failure risk assessment is crucial for the operation of a system [3].

Cascading failure in a power system is the propagation of a sequence of dependent failures. In a renewable-dominant hybrid alternating current/direct current (AC/DC) power system, the interactions of cascading failures with HVDC and renewable energy can result in security and stability events [4]. Under the threat of these events, the HVDC system is unable to operate at the rated active power, which limits the economy of the system [5]. Therefore, efficient risk assessment of cascading failures and accurate identification of critical lines play an important role in preventing cascading failures and increasing system efficiency [6].

Regarding cascading failure risk assessment, Ref. [7] defines the value at risk (VaR) and conditional VaR (CVaR) for cascading failures based on the Oak Ridge National Laboratory (ORNL)–Power Systems Engineering Research Center (PSerc)–Alaska University (OPA) model, successfully introducing the probability distribution function curve into risk assessment. Ref. [8] proposes a risk assessment method that considers the multi-timescale characteristics of cascading failures, in which duplicated simulation is avoided by using a Markov tree search algorithm to increase efficiency. To reflect the dynamic process during severe failures, the transient low voltage induced by faults [9] and the frequency dynamic process caused by power disturbances [10] have been considered in risk assessment. For an AC/DC hybrid power grid, the risk of HVDC

* Corresponding author.
E-mail address: liuyt@sdu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.eng.2025.06.033
2095-8099/© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

successive commutation failure caused by cascading failures can be estimated using a stacked denoising autoencoder (SDAE) [11]. Using quantitative interaction indexes between AC and DC systems, Ref. [12] assesses the risk of cascading failures based on non-cooperative game theory. Machine learning methods can also be used for risk assessment: Ref. [13] uses an extreme learning machine to quickly estimate the total transfer capability limited by the dynamic security of a hybrid AC/DC system, and a physics-informed graph neural network has been utilized to evaluate the risk of cascading failures [14]. Although the above risk assessment methods consider the impact of security and stability events on subsequent failures, their risk definition is mostly centered around load loss. They cannot reflect the risk of the DC power being limited.

A dynamic thermal rating (DTR) system adjusts the transmission line capacity based on real-time weather conditions, increasing the transmission capability and mitigating node overvoltage [15]. By correcting the line ratings, DTR facilitates higher renewable energy integration and alleviates network congestion [16]. It also improves system resilience in N-1 security assessments [17] and enables a more accurate evaluation of voltage stability and fault risks under extreme weather [18]. DTR significantly boosts renewable utilization while reducing load shedding and curtailment [19]. Furthermore, by accounting for the effects of wind on both generation and transmission capacity, DTR can increase the reliability of wind-power-integrated networks [20]. However, existing cascading failure analyses often ignore DTR systems, leading to conservative risk assessments and greater deviations under special weather conditions.

Regarding critical line identification, fault chain theory and vulnerability indices are widely used in many vulnerability assessment methods [21]. Ref. [22] defines representative cascading failures from the perspective of the relationship between fault chains, which are identified by reinforcement learning to improve efficiency. Ref. [23] successfully applies the PageRank algorithm to critical line identification by mapping cascading failures to a directed weight graph. This mapping approach improves the applicability of the PageRank algorithm, although it leads to the loss of key information hidden in fault chains. Complex network theory has been used to analyze critical lines from the perspective of system structural vulnerability [24], with a focus on small-world networks [25] and scale-free networks [26]. These methods tend to identify devices with high betweenness as critical lines [27]. From the perspective of data mining, the association rules of cascading failures can be mined to identify critical cascading failure patterns [28], using algorithms such as Apriori [29] and PrefixSpan [30]. Frequent pattern growth (FP-Growth) can also be used to mine the relationship between the static security risk and the renewable energy characteristics [31]. Although these pattern-mining-based methods can use the information hidden in fault chains, it is difficult for them to distinguish the importance of each line in the patterns. In particular, when there are multiple types of security and stability events, the degree of impact of faults varies greatly. Therefore, the above methods still have shortcomings in determining critical lines.

To address the problems described above, we present a rapid risk assessment method for cascading failure in renewable-energy-dominant hybrid AC/DC power systems based on the gradient boosting decision tree (GBDT) and FP-Growth algorithms. First, the impact of cascading failures on HVDC transmission of renewable energy is analyzed at the receiving end and the sending end of a hybrid AC/DC system. Security and stability events—including successive commutation failure, overvoltage, and transient instability—are triggered by cascading failures and limit the active power of the HVDC system. Second, a cascading failure risk index is defined, with a focus on DC power limitation while considering the security and stability of the system. To overcome the problem of time-consuming calculations, GBDT is used to rapidly predict the maximum DC power limited by cascading failures. A GBDT update strategy significantly improves the model's generalization ability to the uncertainty of renewable energy. Finally, the FP-Growth algorithm is deployed to mine frequent patterns in cascading failures. By taking the subordinate relationship between frequent patterns and the degree of impact of faults into account, the importance of faults in frequent patterns is defined in order to accurately identify critical lines in cascading failures. Fault importance can effectively reflect the impact of lines on cascading failure.

The main contributions of this paper are as follows: ① A cascading failure risk index is defined based on the DC power being limited by security and stability events caused by cascading failures. ② The maximum DC power limited by cascading failures is predicted using GBDT with an update strategy, which significantly improves the speed and generalizability of risk assessment. ③ An improved FP-Growth algorithm is used to search for the association rules of cascading failures, making it possible to effectively identify critical lines in the mined frequent patterns.

The remainder of this paper is organized as follows: The impact of cascading failures on HVDC transmission of renewable energy is analyzed in Section 2. A risk assessment method for cascading failure based on GBDT is presented in Section 3. Critical line identification based on FP-Growth is proposed in Section 4. Case studies are presented in Section 5. Finally, Section 6 summarizes the conclusions.

2. Impact of cascading failures on HVDC transmission of renewable energy

In renewable-energy-dominant hybrid AC/DC power systems, cascading failures can trigger security and stability events, severely limiting the active power-transmission capability of HVDC systems. Therefore, we analyze the impact of cascading failures on security and stability events to explain the relationship between the maximum DC power and the cascading failures. This analysis provides a theoretical basis for subsequent risk assessment and critical line identification.

2.1. Successive commutation failure caused by low voltage

In LCC-HVDC systems, LCC commutation is inherently dependent on adequate commutation voltage from the AC grid. Insufficient commutation voltage can lead to HVDC commutation failure, which is usually modeled as a low extinction angle of the inverter. In a quasi-steady-state model, the extinction angle can be mathematically expressed as

$$\gamma = \arccos\left(\frac{\sqrt{2}\omega L_r I_d}{U_L} + \cos\beta\right) \tag{1}$$

where U_L is the commutation bus voltage, L_r is the equivalent commutation inductance, I_d is the DC, ω is the angular frequency, and γ and β represent the extinction and leading angles of the inverter, respectively.

A short-circuit fault in an AC grid can cause a decrease in the commutation bus voltage. Based on the expression of the extinction angle described in Eq. (1), insufficient commutation voltage leads to a decrease in the extinction angle. When the extinction angle drops below the minimum extinction angle, commutation failure occurs.

After a commutation failure, the reactive power demand of the converter increases significantly during the recovery of DC power [11]. Local compensation devices at converter stations are

insufficient, forcing reliance on the AC grid for part of the reactive power. However, the capacity of the AC grid to support the required reactive power is significantly dependent on the strength of the AC grid. If the AC grid is unable to provide sufficient reactive power support, the voltage of the converter bus will drop again, slowing the post-fault recovery and potentially leading to commutation failures. Subsequently, voltage collapse occurs and successive commutation failures happen.

In an AC system, voltage support for the commutation bus is mainly provided by nearby power sources, and the support strength depends on their electrical connectivity. Transmission line outages weaken this connection, reducing the system support strength. As shown in Eq. (1), the extinction angle, which determines HVDC commutation failure, is closely related to the post-fault voltage drop. Reduced support strength leads to deeper voltage dips during AC faults, increasing the risk of HVDC commutation failure. Consequently, cascading line outages continuously degrade the system support, increasing the risk of successive commutation failures.

As indicated by Eq. (1), the extinction angle is related to the DC. Therefore, reducing the DC power—and thus lowering the required DC—is essential to prevent successive commutation failures; however, doing so reduces the capability to accommodate large-scale renewable energy across regions through HVDC transmission.

2.2. Sending-end overvoltage caused by commutation failure

To suppress commutation failure and accelerate commutation recovery, a voltage-dependent current order limiter (VDCOL) is used to limit the DC during failures. This significantly reduces the active and reactive power absorbed by the rectifier side converter from the AC system after commutation failure. Usually, the reactive power consumed by converters is provided by capacitors within the converter station. After a commutation failure, the capacitor cannot be cut off in time, and the redundant reactive power generated by the capacitor leads to overvoltage in the sending-end AC system [32]. The degree of overvoltage is related to the voltage support capacity of the AC system.

Renewable energy is highly dependent on power electronic components, which are susceptible to disturbances. Therefore, sending-end overvoltage disturbances caused by commutation failures can lead to cascading trip-offs of renewable energy. More specifically, continuous voltage disturbances have a more serious impact on doubly fed induction generator (DFIG)-based wind turbines (WTs). During normal operation, the rotor voltage amplitude of a DFIG is

$$V_{\mathrm{r}} = s\frac{L_{\mathrm{m}}}{L_{\mathrm{s}}}V_{\mathrm{s}} \tag{2}$$

where V_{s} is the stator voltage, L_{s} is the stator inductance, L_{m} is the mutual inductance, and s is the slip ratio.

It can be seen from Eq. (2) that the rotor voltage V_{r} is proportional to the slip s. Because $|s| < 0.3$, the rotor voltage is small during normal operation. When a fault causes a voltage disturbance, synchronous and static components of the flux will exist in the generator, since the stator flux cannot suddenly change [31]. The maximum rotor voltage can be approximated as

$$V_{\mathrm{r,max}} \approx \frac{L_{\mathrm{m}}}{L_{\mathrm{s}}}(|s|V_2 + (1-s)(V_1 - V_2)) \tag{3}$$

where V_1 and V_2 are the stator voltages before and after a disturbance, respectively.

In Eq. (3), the maximum rotor voltage is related to the slip and stator voltage and is significantly higher than the normal value. When $s < 0$, which indicates that the DFIG is functioning in a hyper-synchronous state, the rotor coil will induce a higher voltage because of the static component of the flux. The rotor control voltage is generated and limited by the rotor-side converter. Once the rotor voltage induced by the flux exceeds the output limit of the converter, the converter will lose control of the rotor current and the output power. When continuous low and high voltages occur due to cascading failures, the static components of the flux will be superimposed, increasing the risk of trip-off.

Cascading failures caused by power-flow transfer can weaken the system strength. Then, the degree of overvoltage caused by commutation failure is severe, increasing the risk of large-scale cascading tripping of renewable energy. To prevent the tripping of renewable energy caused by overvoltage, it is necessary to reduce the active power of the HVDC, which also reduces the reactive power required for the capacitor and suppresses the overvoltage at the sending end. This seriously limits the renewable energy power transmitted through HVDC.

A commutation failure caused by a short-circuit fault in the receiving-end system can change the operation conditions of the sending-end system due to power fluctuations, leading to cascading failures in the sending-end system. Therefore, cascading failures will transfer from the receiving end to the sending end. In addition, a short-circuit fault at the sending end can cause commutation failure at the receiving end under specific operation conditions [33], leading to power transfer at the receiving end and triggering cascading failures. With the participation of HVDC, cascading failures will thus propagate inter-regionally, increasing the failure range and limiting the DC power.

2.3. Equivalent power angle instability caused by low voltage

When phase-locked loop (PLL)-synchronized renewable energy farms (REFs) enter a low-voltage ride-through (LVRT), there is a risk of equivalent power angle instability [34]. Once synchronism is lost, the instability current will affect the point of common coupling (PCC) voltage, further affecting other REFs. An example system is shown in Fig. 1.

The PLL of the kth renewable energy adjusts the q-axis component of terminal voltage $U_{\mathrm{tq}k}$ to zero to realize a phase lock, as shown in Fig. 2(a).

In the system, the stable equilibrium point of the PLL for the kth REF exists only when $U_{\mathrm{tq}k} = 0$ and $\theta_k = \delta_k$ (where δ_k is actual power angle of the kth REF). When the fault occurs, the output current of the REF during LVRT can be considered a controlled current source, and the terminal voltage of the kth REF is

$$U_{\mathrm{t}k} = U_{\mathrm{f}} + \sum_m I_m Z_{\mathrm{P}} + I_k Z_k \tag{4}$$

where U_{f} is the fault voltage; I_m and I_k are the current of the mth and kth REF, respectively; Z_{P} is the common impedance; and Z_k is the line impedance of the kth REF.

According to Eq. (4), there is a coupling relationship between the terminal voltage of each REF because the current of each REF

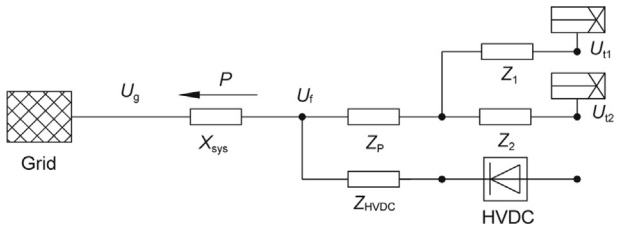

Fig. 1. An AC/DC example system. U_{g}: grid voltage; P: active power; X_{sys}: system equivalent reactance; U_{f}: voltage at the fault point; Z_{P}: common impedance; Z_{HVDC}: impedance of the HVDC station; Z_k: line impedance of the kth REF (here k = 1 and 2); $U_{\mathrm{t}k}$: terminal voltage of the kth REF.

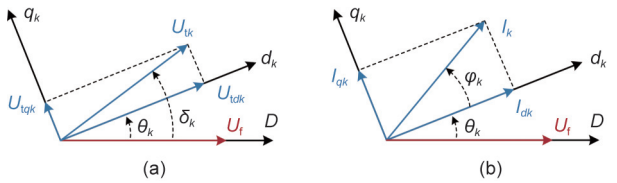

Fig. 2. Relationship between the (a) terminal voltage and the PLL angle and (b) current and the PLL angle. D: d-axis of the system; d_k: d-axis of the kth REF; q_k: q-axis of the kth REF; U_{tk}: terminal current of the kth REF; U_{tdk}: d-axis component of U_{tk}; U_{tqk}: q-axis component of U_{tk}; δ_k: actual power angle of the kth REF; θ_k: power angle output by the PLL of the kth REF; I_k: terminal current of the kth REF; I_{dk}: d-axis component of I_k; I_{qk}: q-axis component of I_k; φ_k: power factor angle of the kth REF.

has a common line impedance Z_P after the PCC. The REF terminal voltage is related to its output current. As shown by the relationship between the current and PLL in Fig. 2(b), the current phase depends on the PLL angle θ_k and the power factor angle φ_k, where the power factor angle has been determined. According to the relationship between the PLL and the voltage in Fig. 2(a), the PLL realizes phase-locking by tracking the terminal voltage, and U_{tqk} can be expressed as

$$U_{tqk} = -U_f \sin\theta_k + Z_k I_k \sin(\varphi_{Zk} + \varphi_k) + Z_P \sum_m I_m \sin(\varphi_{ZP} + \varphi_m + \theta_m - \theta_k) \quad (5)$$

where φ_{Zk} is the impedance angle of the kth REF, φ_{ZP} is the impedance angle of the common impedance, φ_m is the power factor angle of the mth REF, and θ_m is the power angle output by the PLL of the mth REF.

According to Eq. (5), the PLL angle, output current, and terminal voltage are coupled. The power angle stability is strongly related to the fault voltage, impedances, and power factors.

The active power transmitted through the equivalent line of the system can be expressed as

$$P = \frac{U_g U_f}{X_{sys}} \sin\delta \quad (6)$$

where U_g and U_f are the voltage magnitudes at both ends of the transmission line, δ is the power angle difference, and X_{sys} is the equivalent reactance of the system.

As shown in Fig. 3, as the transmission power of the equivalent line increases or the equivalent reactance increases, the transmission power's operating point approaches the static stability limit point A. At the same time, the rate of change of the power angle difference relative to the power ($\Delta\delta/\Delta P$) becomes greater. This means that the system damping of power angle changes is reduced. Under power disturbances with the same magnitude, the change in phase angle will be greater.

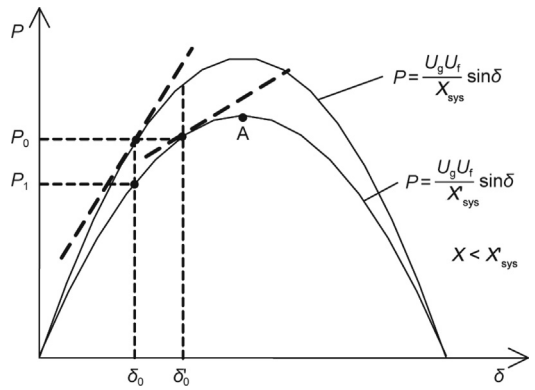

Fig. 3. The relationship between the power and the power angle difference.

Transmission line outages caused by cascading failures can increase the equivalent reactance of the system. According to Eq. (6), the power angle of U_f is prone to oscillations due to power disturbances. Thus, the synchronization instability risk of the renewable energy increases when the system enters a transient process, according to Eq. (5). This may even lead to a power angle instability of δ, due to crossing the static stability limit A. Therefore, to improve the stability of the system, the DC power must decrease to provide sufficient damping and a static stability margin. Thus, instability events caused by cascading failure can limit the maximum DC power.

The relationships and propagation sequence among successive commutation failures, overvoltage and power angle instability are as follows: Cascading failures typically cause voltage dips, which first trigger power angle instability and HVDC commutation failures. The commutation failures then induce overvoltage conditions in the sending-end AC grid, leading to the trip-off of renewable energy. In the receiving-end system, the weakened support strength from the cascading failures increases the risk of successive commutation failures, further propagating faults to the sending-end grid. In the sending-end system, cascading failures can also induce HVDC commutation failure by altering the DC through voltage fluctuations. Regardless of whether the commutation failures are initiated at the sending or receiving end, they cause alternating low voltage and overvoltage at the commutation bus, exacerbating the renewable energy trip-off. During the cascading failure process, both short-circuit-induced and commutation-failure-induced voltage dips contribute to power angle instability.

3. Risk assessment of cascading failures

Considering the strong nonlinear relationship between cascading failures and the maximum DC power, a rapid risk assessment method based on GBDT is proposed. A cascading failure risk index is defined, focusing on the limitation of DC power by security and stability events. To improve prediction accuracy and generalizability under renewable energy uncertainty, an update strategy is introduced into the GBDT. This method significantly increases the efficiency and robustness of risk assessment for cascading failures, supporting the subsequent critical line identification.

3.1. A risk index definition that focuses on DC power limitation

Security and stability events caused by cascading failures severely limit the renewable energy power transmitted through HVDC. Existing methods for assessing the risk of cascading failures primarily focus on the risk of load loss, which has a limited impact on increasing the DC power transmission capacity. In power system operation, the transfer capacity of the critical interface is assessed to reflect the security risk [35]. To increase the renewable energy power transmitted through HVDC, we define the risk index of cascading failures by focusing on the limitation of DC power, thereby establishing a foundation for identifying the critical line that is limiting the DC power.

As discussed in Section 2, the security and stability events caused by cascading failures are key limiting factors for DC transmission capacity. By reducing the DC power, it is possible to effectively avoid security and stability events. Therefore, the maximum DC power limited by cascading failures can be used as a factor for risk assessment. The line failure probability for a rated DC power is also considered. Thus, the risk index of cascading failures can be defined as

$$R = p \cdot (P_{rated,HVDC} - P_{limited,HVDC}) \quad (7)$$

where p is the failure probability, which is calculated according to a linear model [36], and $P_{\text{rated,HVDC}}$ and $P_{\text{limited,HVDC}}$ are the rated and limited DC power, respectively.

3.2. Calculation of maximum DC power based on the sensitivity method

Because it is not possible to analytically express the relationship between the maximum DC power and security events, it is necessary to obtain the maximum DC power through iterative calculation. If a binary search algorithm is used to test the maximum DC power of a 10 000 MW HVDC, a maximum of 10 iterations are necessary for convergence when the required precision is set to 10 MW. To improve the efficiency of calculating the maximum DC power, the calculation process is accelerated based on the sensitivity of the key features.

First, the key feature E is defined as the electrical quantity directly related to the security and stability events. Taking successive commutation failure as an example, the commutation bus voltage during a fault serves as the key feature. The threshold value of the key feature related to the event, E_{th}, is calculated based on historical data. For example, E_{th} is defined as the threshold commutation voltage, when successive commutation failure is considered. Then, the simplified sensitivity α is defined as the approximate sensitivity of the key parameter to the change in maximum DC power. Taking the converter bus voltage as an example, its sensitivity to the maximum DC power can be approximately considered as the short-circuit ratio at the commutation bus. Thus, the change in maximum DC power per iteration can be calculated as

$$\Delta P_{\text{limited,HVDC}} = \lambda \cdot (E - E_{\text{th}})/\alpha \tag{8}$$

where λ is the iteration step that is optimized based on historical calculation processes and varies for different events. In this way, statistics are used to eliminate the errors caused by approximations. If the number of overshoots exceeds the number of undershoots during the iterative process, λ is decreased; otherwise, λ is increased.

3.3. Fast risk assessment based on GBDT

In hybrid AC/DC systems, cascading failures at both the sending and receiving ends may have large-scale inter-regional impacts. Moreover, the integration of renewable energy and changes in load increase the uncertainty of operation scenarios. Thus, the size of the cascading failure set is very large. Using the above method to assess the risk of cascading failures requires a significant amount of time, making it difficult to meet the requirement of online application. To address this issue, we propose a fast online risk assessment method based on GBDT.

GBDT is an ensemble learning model characterized by efficient training [4]. It integrates multiple weak learners to form a more powerful predictive model. In the training process, each weak learner minimizes the model's prediction error by following the negative gradient of the loss function. Compared with a single decision tree, GBDT offers increased robustness and generalization capabilities. As illustrated in Fig. 4, the architecture of GBDT consists of an aggregator that combines the outputs from several weak learners; X represents the collection of training samples, while y denotes their corresponding labels. The strong learner F_N is composed of multiple weak learners h_n, each operating independently to perform prediction tasks. Their individual outputs are subsequently combined by an aggregator to produce the final prediction of the ensemble model, which can be formulated as

$$F_N(x_m) = \sum_{n=1}^{N} h_n(x_m) \tag{9}$$

where $F_N(\cdot)$ denotes the strong learner, x_m represents the mth sample in the input vector, $h_n(\cdot)$ refers to the nth weak learner, and N indicates the total number of weak learners integrated into $F_N(\cdot)$.

The strong learner F_n is constructed by aggregating all weak learners trained up to the nth iteration. Its mean squared error loss function is defined as

$$L_n = \frac{1}{2}\sum_{m=1}^{M}(y_m - F_n(x_m))^2 \tag{10}$$

where y_m denotes the label, and M indicates the total number of training samples.

The new learner, h_{n+1}, is trained by following the negative gradient direction of the loss function L_n. The label $y^{(n+1)}$ used for training new learner h_{n+1} is formulated as

$$y^{(n+1)} = \xi \cdot \sum_{m=1}^{M}(y_m - F_n(x_m)) \tag{11}$$

where ξ denotes the learning rate.

According to Eq. (11), the training label consists of two components: The first component is the learning rate, and the second component is the negative gradient of the loss function L_n, which is proportional to the residual defined as y_m–$F_n(x_m)$. As each weak learner is trained on the residuals, the overall prediction error progressively decreases. Ultimately, the strong learner is obtained by aggregating all weak learners.

To rapidly assess the risk of cascading failures, the risk assessment procedure involves training the GBDT offline, updating the weak learners in real time, and performing an online risk assessment of cascading failures. To improve the generalization ability, cascading failures under different operating conditions are generated offline using the Monte Carlo tree search method [4] and are processed into training samples. A training sample consists of input features and a sample label, where the input features contain the factors affecting the risk of cascading failures. We choose the active power of renewable energy, the HVDC system, synchronous generators, and loads; the terminal voltages of WTs and HVDC systems; and the topology of the AC system as the input features. The sample label y_m is the maximum DC power limited by cascading failure.

In the update process, real-time state estimation and scenario forecast data are used to acquire cascading failures as the correction set. Unlike the offline training stage, only a small number of high-risk cascading failures are needed to ensure computational efficiency. A Monte Carlo tree search with a depth-first strategy is employed to quickly identify high-risk cascading failures under the given real-time operating scenarios. The prediction accuracy of each weak learner is examined with the correction set. Ensemble learning is composed of a large number of weak learners, each of which has poor predictive ability and significant errors. However, the statistical result of a weak learner tested by the correction set can roughly reflect the accuracy of the weak learner. While updating the model, it is necessary to control the size of the tree structure, so learners with low prediction accuracy will be deleted. To update the GBDT, additional weak learners are trained to fit the correction set using Eq. (11). The update strategy provides the GBDT with a strong generalization ability for the uncertainty induced by renewable energy.

4. Identification of critical lines

To effectively identify the critical lines that dominate the propagation of cascading failures, we propose an improved FP-Growth algorithm. Based on the high-risk cascading failures obtained from the assessment, frequent patterns in cascading failures are mined,

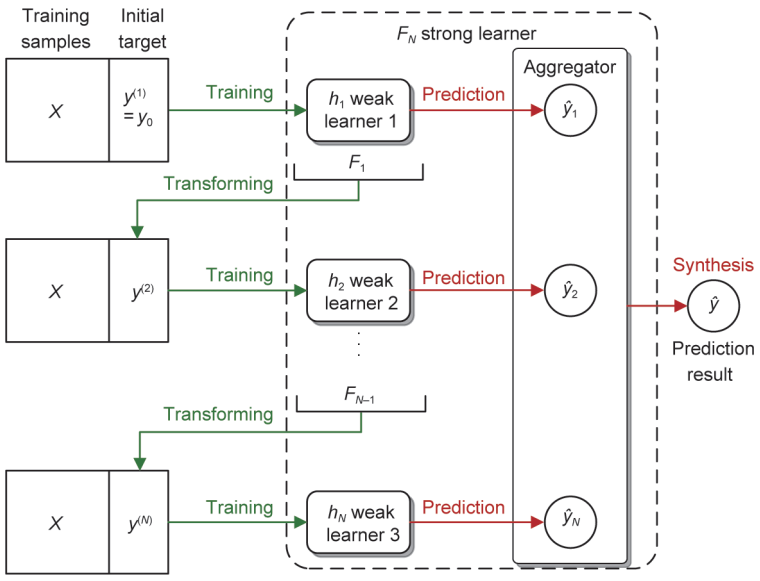

Fig. 4. Structure of the GBDT. $y^{(N)}$: label for training the nth weak learner; y_0: initial label; $\hat{y}_N$: predicted result of the nth weak learner; $\hat{y}$: predicted result of the GBDT.

and an importance index is defined based on the degree of impact of the faults within each pattern. By considering the subordinate relationship among the faults, the proposed method accurately identifies critical lines with a high risk contribution, thereby providing guidance for system reinforcement and protection strategies.

4.1. Description of the cascading failure chain

Fault chains are used to describe the propagation of cascading failures. Traditional representations of fault chains focus on the relationships between failures. With large-scale renewable energy being transmitted through HVDC, the DC power is limited by various security and stability events. The inability to quantify the impact of failures on these events reduces the accuracy of critical line identification.

To address this problem, security and stability events caused by cascading failures and the impact factors of faults on these events are added to the fault chains. More specifically, for successive commutation failure and sending-end overvoltage, the impact factor is calculated from the electrical distance between the fault and the converter station [37]. For instability, the impact factor is calculated from the average electrical distance between the fault and instable devices. This calculation shows that the impact factor of a fault is strongly related to the form of the final event caused by the cascading failure.

Accordingly, a cascading failure l_1 is represented as l_1: ⟨a, b, c, d, e⟩, where a single regular lowercase letter represents a fault, such as a, b, c, d, e, f, g, and h. The last fault in the sequence denotes the security and stability events caused by the cascading failure, such as an HVDC block due to successive commutation failure. The probability of failures is represented as $[p_a, p_b, \ldots, p_e]$, and the impact factors are represented as $[k_a, k_b, \ldots, k_e]$. The risk of the cascading failure is R_1.

4.2. Critical line identification based on FP-Growth

Critical lines are identified as the lines that promote fault propagation or cause severe security and stability problems during cascading. A data mining algorithm is applied to extract the critical lines from a large number of high-risk cascading failures, identifying the key elements that drive fault propagation and increase the risk of cascading failure. Identifying the critical lines provides practical guidance for system operators by highlighting key areas for focused monitoring. It also offers valuable insights for preventive control, enabling more targeted and effective measures to strengthen system resilience against cascading failures.

The FP-Growth algorithm is an association rule mining algorithm that efficiently mines frequent patterns from large-scale data [38]. It consists of two steps: constructing an FP-Tree and mining frequent patterns based on the FP-Tree. The FP-Tree is a special data structure of the FP-Growth algorithm that stores information from the original dataset. With only two searches on the dataset, frequent patterns in the original dataset are compressed into the FP-Tree. Like a prefix tree, paths with the same prefix can be shared, which compresses the data to improve the efficiency of the algorithm.

To construct the FP-Tree, the cascading failure set is searched to record the frequency of each fault with the cascading failure risk as its weight. Failures are arranged in descending order of their weighted frequencies and stored as a frequent itemset. To reduce the sample space, failures with a weighted frequency lower than a minimum support are removed. For example, the frequent itemset L = {g: 6, a: 5, c: 5, d: 4, e: 4, f: 3} has a minimum support of 3, where this number is the weighted frequency of the fault.

During the second search of the dataset, each cascading failure is filtered and sorted according to the frequent itemset L. This step discards non-frequent failures and standardizes the failure order in the cascading failure chains. For example, cascading failure l_2 = ⟨f, e, b, a, g⟩ is modified to l_2' = ⟨g, a, e, f⟩ according to L.

The FP-Tree is built from the ordered cascading failures, containing paths, failure nodes, and their counts. Cascading failures are added to the tree sequentially. If a prefix path of the cascading failure exists in the tree, the count for the corresponding failures in the tree is incremented by the risk of the cascading failure. If not, a new line path is established at the current failure node. The formed FP-Tree is shown in Fig. 5.

To mine frequent patterns based on the FP-Tree, conditional pattern bases for each fault are first obtained. This is done by starting from each node representing the fault and tracing back to the root of the tree; all traced paths form the conditional pattern base of the failure. The result is shown in Table 1.

From the conditional pattern base of a failure, conditional FP-Trees for the failure are constructed, retaining only those paths

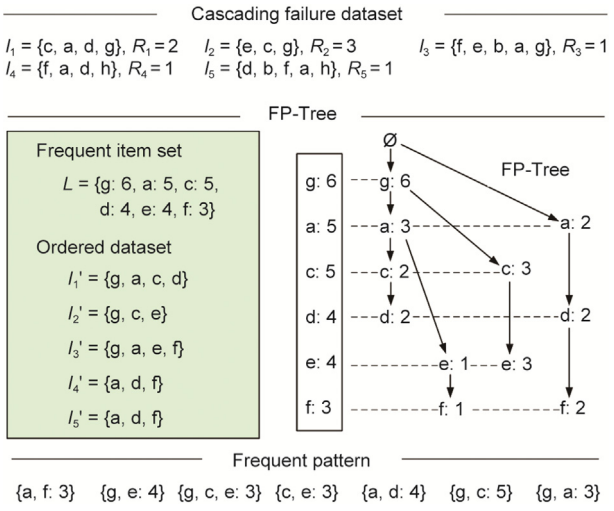

Fig. 5. Structure of the FP-Tree.

corresponding to the conditional pattern base and their counts. Thus, the conditional FP-Tree only includes cascading failures related to the failure, as shown in Table 1, and irrelevant cascading failures are discarded.

Finally, the frequent patterns of the failure are recursively obtained from the conditional FP-Tree by verifying whether the patterns meet the minimum support. The frequent patterns of each failure in the cascading failure set are identified by repeating this process. As shown in Table 1, some frequent patterns have a subordinate relationship; for example, {g, c: 5} is subordinate to {g, c, e: 3}.

After the construction of frequent itemsets, the PrefixSpan algorithm and confidence metric are employed to mine sequential rules. First, subsets are generated from a frequent itemset. Then, the PrefixSpan algorithm is applied to the database to mine frequent sequences composed of these subsets. Finally, sequential rules are filtered based on the confidence. The confidence of a sequential rule is defined as follows:

$$\text{confidence}(\text{set}_{0,i} \rightarrow \text{set}_{0,j}) = \frac{\sup(\text{set}_{0,i} \rightarrow \text{set}_{0,j})}{\sup(\text{set}_0)} \tag{12}$$

where set_0 is the frequent itemset, $\text{set}_{0,i}$ and $\text{set}_{0,j}$ are the subsets, i and j are the indices of these subsets, ($\text{set}_{0,i} \rightarrow \text{set}_{0,j}$) is the sequential rule, and sup(·) is the support.

4.3. Analysis of identification results

Given the definition of impact factor, different faults in a cascading failure have different impacts on the final security and stability events. Moreover, the same fault can exist in frequent patterns of multiple lengths in different cascading failures. Thus, the fault's importance varies across different cascading failures. However, all faults in the frequent patterns generated from the FP-tree have the same status, lacking information on their importance. To identify the critical lines of cascading failures, it is necessary to define the importance of a failure within a frequent pattern based on the frequency of occurrence and impact factors.

The average importance of fault x in the dataset is defined as

$$I(x) = \frac{1}{N_{\text{cf}}} \sum_{i}^{N_{\text{cf}}} (p_{x,l_i} \cdot k_{x,l_i} \cdot R_i) \tag{13}$$

where N_{cf} is the number of cascading failures, and p_{x,l_i} and k_{x,l_i} are the fault probability and impact factor of x in cascading failure l_1, respectively.

The count of a fault pattern, such as FP_1: {a, c}, in cascading failure l_1 is defined as

$$\text{Count}(l_1, \text{FP}_1) = \begin{cases} R_1, & \text{FP}_1 \in l_1 \\ 0, & \text{FP}_1 \notin l_1 \end{cases} \tag{14}$$

where R_1 is the risk of cascading failure l_1. If pattern FP_1 exists in cascading failure l_1, the count value is the cascading failure risk; otherwise, it is 0.

The frequent patterns obtained from the FP-Tree have a subordinate relationship. We define a frequent pattern as a long frequent pattern (LFP) if it is not subordinate to any other frequent pattern, such as {g, c, e: 3} in Table 1; otherwise, it is defined as a subordinate frequent pattern (SFP), such as {g, c: 5}. To increase the efficiency of critical line identification, only LFPs are retained in the frequent pattern set; the influence of SFPs is integrated into LFPs.

The importance of frequent pattern FP_m in the LFP_n is defined as the average count of this frequent pattern across all cascading failures, represented as

$$I(\text{LFP}_n, \text{FP}_m) = \frac{1}{N_{\text{cf}}} \sum_{l_i \in S_{\text{LFP}_n}} \text{Count}(l_i, \text{FP}_m) \tag{15}$$

where S_{LFR_n} is the subset of cascading failures containing LFP_n.

The importance of fault x in LFP_n is the average importance of all frequent patterns containing fault x in LFP_n, represented as

$$I(\text{LFP}_n, x) = I(x) \cdot \sum_{\text{FP}_m \in U_x} I(\text{LFP}_n, \text{FP}_m) \tag{16}$$

where U_x is the set of all cascading failure patterns containing fault x.

By substituting Eq. (16) into Eq. (15), it can be deduced that

$$\begin{aligned} I(\text{LFP}_n, x) &= \frac{1}{N_{\text{cf}}} I(x) \cdot \sum_{\text{FP}_m \in U_x} \sum_{l_i \in S_{\text{LFP}_n}} \text{Count}(l_i, x) \\ &= \frac{1}{N_{\text{cf}}} I(x) \cdot \sum_{l_i \in S_{\text{LFP}_n}} \sum_{\text{FP}_m \in U_x} \text{Count}(l_i, x) \end{aligned} \tag{17}$$

From Eq. (14), it can be seen that the count values under the same cascading failure l_i are all R_i. The number of frequent patterns in U_x containing fault x is $2^{\zeta-1} - 1$, where ζ is the total number of frequent items in the cascading failure. Taking LFP: {g, c, e} as example, ζ = 3. The frequent patterns containing fault g are {g, c}, {g, e}, and {g, c, e}, fulfilling the above relationship. This makes it possible to simplify the calculation of the importance of fault x in LFP_n, which is expressed as

$$I(\text{LFP}_n, x) = \frac{1}{N_{\text{cf}}} I(x) \cdot \sum_{l_i \in S_{\text{LFP}_n}} \left(2^{\zeta-1} - 1\right) R_i \tag{18}$$

In this way, it is possible to determine the importance of different faults in a single LFP. By ranking the faults according to their importance, the critical lines in cascading failures can be identified.

Table 1
Result of frequent pattern mining.

Fault	Conditional pattern base	Conditional frequent pattern tree	Frequent pattern
f	{g, a, e: 1}, {a, d: 2}	{a: 3}	{a, f: 3}
e	{g, a: 1}, {g, c: 3}	{g: 4}, {g, c: 3}, {c: 3}	{g, e: 4}, {g, c, e: 3}, {c, e: 3}
d	{g, a, c: 2}, {a: 2}	{a: 4}	{a, d: 4}
c	{g, a: 2}, {g: 3}	{g: 5}	{g, c: 5}
a	{g: 3}	{g: 3}	{g, a: 3}

5. Case study

In this section, a modified Ningxia–Shandong hybrid AC/DC power system is utilized to validate the proposed cascading failure risk assessment and critical line identification methods. The simulation results demonstrate the accuracy, rapidity, and scalability of the proposed approach. All simulations were conducted on a personal computer equipped with an Intel 3.20 GHz central processing unit (CPU) and 16 GB random access memory (RAM).

5.1. Cascading failures in the Ningxia–Shandong hybrid AC/DC system

The Ningxia–Shandong hybrid AC/DC system has 268 buses, 92 machines, and 462 AC lines. Two AC power grids are connected through a Ningdong (ND) DC link with a capacity of 4000 MW. Twelve 400 MW wind farms are integrated into the Ningxia grid, and six 400 MW wind farms are integrated into the Shandong grid. The proportion of the renewable energy installed capacity to the total installed generation capacity exceeds 60% in the area near both converter stations. Therefore, this system is a renewable-dominated hybrid AC/DC power system. In particular, when the power output from renewable energy is abundant, renewable energy generation plays a dominant role in the dynamic interactions during cascading failures. The simplified structure of the hybrid AC/DC system is shown in Fig. 6.

A cascading failure search method based on a Monte Carlo tree search [4] is employed to generate a set of cascading failure samples. The sample size is set to 10 000. The initial faults are randomly selected based on the load ratio of each line, and four-stage cascading failures are generated. The probability of these line faults is calculated using a linear model [36]. The cascading failures lead to security and stability events that meet the following constraints. The threshold for overvoltage causing the renewable energy sources to trip is 1.3 per unit (p.u.). The minimum voltage for judging a commutation failure is set to 0.8 p.u., and the time threshold for judging forced HVDC blocking is set to 1 s [39]. Regarding the transient stability, the maximum divergence between any two rotor angles should not exceed 180 degrees.

Three typical types of cascading failures are shown in Table 2. The maximum DC power limited by each cascading failure is calculated by the proposed sensitivity-based method. The error tolerance is set to 10 MW. Taking cascading failure l_2 as an example, when the DC power of the ND DC link decreases to 2378.46 MW, the WT trip-off caused by overvoltage is eliminated. The failure probability of l_2 is about 0.107. The cascading failure risk is calculated based on the maximum DC power and the cascading failure probability according to Eq. (7). The risk of l_2 is 173.04.

During cascading failure propagation, an early-stage fault cannot trigger security events because the AC grid structure is strong. Thus, the primary impact of early-stage faults is to change the structure of the AC grid. As more lines trip, the structure of the AC grid becomes weaker, allowing short-circuit faults to cause security events [11]. Therefore, during cascading failure propagation, the preceding stages of faults are considered as line trips, while the final stage is a three-phase short-circuit fault. This is followed by an examination of whether the system experiences any security events after the short-circuit fault.

In cascading failure l_2, the cascading failure sequence begins with a random initial fault on transmission line 128–130, causing a power flow redistribution that subsequently overloads and trips lines 102–103 and 121–122. The cascading line outages cause the power flow on line 134–136 to increase substantially to more than twice its initial loading, which significantly increases the risk of a short-circuit fault on the line. The detailed power-flow redistribution is shown in Table 3. The weakened system then suffers a critical three-phase short-circuit fault on line 134–136, which triggers HVDC commutation failure, ultimately leading to severe overvoltage and consequent trip-off of renewable generation in the sending-end system.

To validate the effectiveness of the proposed risk index, the system security under cascading failure l_2 is assessed using two operating conditions with a DC power of 2400 and 2300 MW, respectively. Furthermore, to verify the influence of renewable energy on the dynamic characteristics of the system, a simulation is also carried out for a conventional-power-source-dominated system without WTs. The simulation results for the wind farm voltage at bus 128 are illustrated in Figs. 7 and 8. The cascading failure sequence begins with an initial fault at 1 s, with subsequent faults occurring at 2 s intervals. Fig. 8 presents a detailed comparison of the voltage dynamics under different DC power conditions.

At 7 s, a three-phase short-circuit fault occurs on line 134–136, causing a voltage dip at the sending end. The fault is cleared by line

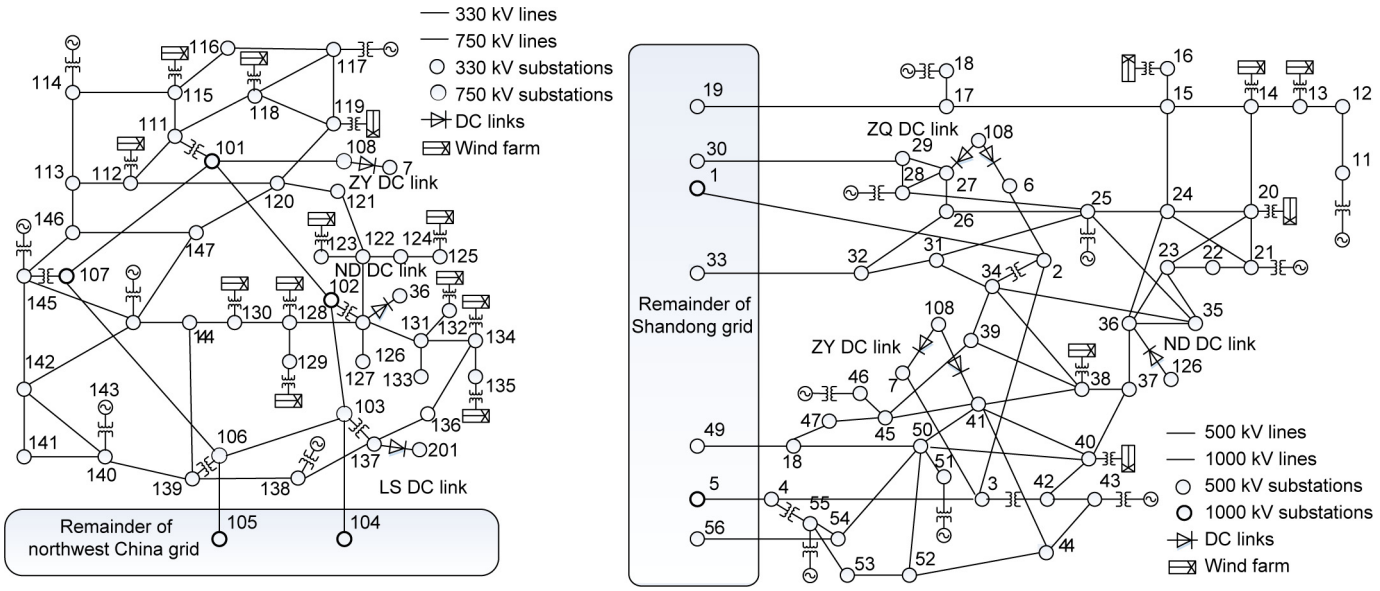

Fig. 6. The Ningxia–Shandong hybrid AC/DC system. ZY: Zhaoyi HVDC project; ZQ: Zhaqing HVDC project; ND: Ningdong HVDC project.

Table 2
Risk of three typical types of cascading failures.

Index	Cascading failure chain	Maximum DC power (MW)	Cascading failure risk
l_1	Line 23–36, line 36–37, line 23–35, line 25–35, successive commutation failures	1849.87	215.62
l_2	Line 128–130, line 102–103, line 121–122, line 134–136, overvoltage leads to WT trip-off	2378.46	173.04
l_3	Line 102–103, line 121–122, line 130–144, line 126–131, WT instability at bus 129 and 130	2631.84	104.85

protection at 7.1 s, leading to a rapid voltage recovery. However, at 7.14 s, the voltage disturbance at the sending end induces a DC overcurrent, resulting in a commutation failure at the receiving end of the DC system. This then causes the sending-end AC voltage to first drop and then rise sharply.

Due to the larger reactive power disturbance under the condition of 2400 MW DC power, the resulting voltage fluctuations are more significant. At 7.23 s, the sending-end voltage reaches its peak. Under the 2300 MW condition, the maximum overvoltage is 1.284 p.u., which does not trigger overvoltage protection of the renewable energy sources. In contrast, under the 2400 MW condition, the overvoltage exceeds 1.3 p.u., activating the overvoltage protection and causing the renewable energy sources to trip from the grid. The loss of reactive power support from the high-voltage ride-through capability leads to a further voltage rise, reaching up to 1.326 p.u.

In contrast, for a conventional-power-source-dominated system without WTs, the strong voltage support capability of conventional generators results in smaller voltage disturbances following a fault. As a result, commutation failure in the HVDC system is avoided, and overvoltage-induced cascading trip-off of renewable energy at the sending-end system does not occur. This finding indicates that the large-scale integration of renewable energy significantly changes the dynamic characteristics of the power system.

These results demonstrate that the proposed risk index effectively reflects the constraints imposed by cascading failures on DC transmission capacity and provides a reliable assessment of the impact of such faults on the stability of a hybrid AC/DC power system.

5.2. Performance validation of GBDT

To verify the effectiveness of the fast risk assessment method based on GBDT, an SDAE method is used for comparison. The cascading failure samples and the features of the power system are taken as inputs, and the maximum DC power values limited by each cascading failure are taken as sample labels to train the models. The 10 000 cascading failure samples are randomly divided into a training set of 8000 samples and a testing set of 2000 samples. The hyper-parameters of the GBDT-based method,

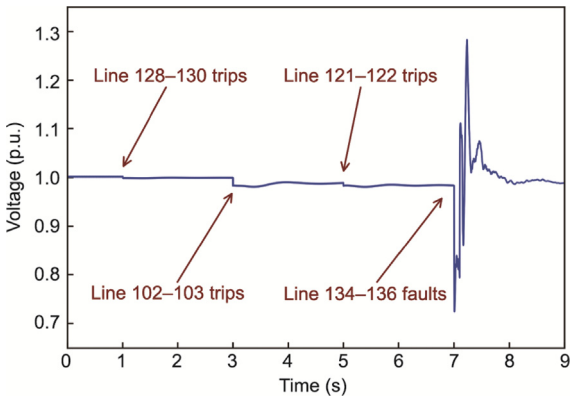

Fig. 7. Voltage of the wind farm at bus 128 under the condition of 2300 MW DC power.

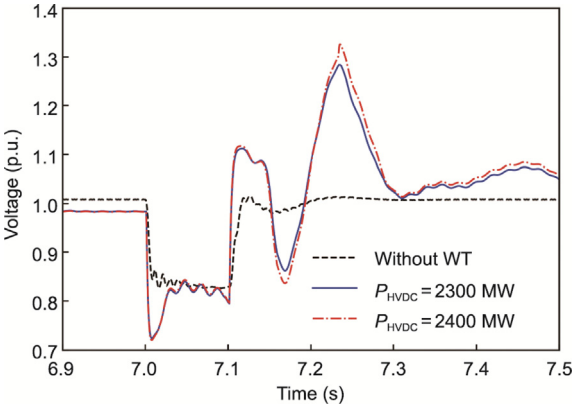

Fig. 8. Detailed comparison of wind farm voltage dynamics at bus 128.

which include the learner number, learning rate, and maximum depth, are set to 1000, 0.05, and 5, respectively.

Next, the training set is used to train the GBDT and SDAE models. The testing set is used to validate the performance of the trained models. A comparison of the prediction errors is shown in Table 4. The results show that the GBDT can predict the maximum DC power with high accuracy, and the accuracy is within the error tolerance. It can be seen that using a trained GBDT to predict risks has a high accuracy with a low average error, and a stable performance with a low root mean square error.

Time consumption is a critical index for the online risk assessment of cascading failures. The time each method requires to process 10 000 samples is shown in Table 5. The tests are performed on a personal computer with an Intel 3.20 GHz CPU and 16 GB RAM. The results show that the proposed sensitivity-based method can significantly improve the efficiency of offline risk calculation compared with the binary search method. However, for risk assessment with a large sample set, the offline methods still require a significant amount of time. Both methods spend most of the time

Table 3
Active power-flow distribution during the cascading failure l_2.

Fault stage	Line active power (MW)			
	Line 128–130	Line 102–103	Line 121–122	Line 134–136
Initial state	661.24	1669.60	448.94	411.15
Line 128–130 trips	0	1894.14	608.66	471.02
Line 102–103 trips	0	0	805.18	718.77
Line 121–122 trips	0	0	0	849.86

Table 4
Prediction errors for the maximum DC power according to the GBDT and SDAE methods.

Method	Average error (MW)	Root mean square error (MW)
GBDT	3.523	3.974
SDAE	12.634	18.429

Table 5
Time consumption by the GBDT and SDAE methods.

Process	Time (s)
Binary search method	13 732.81
Sensitivity method	4 264.27
GBDT training	308.59
GBDT prediction	11.52
SDAE prediction	45.74

on offline training and realize fast online risk assessment, which can meet the requirements of online applications. Thanks to the parallel computing capability of GBDT, its evaluation speed is much higher than that of the SDAE method.

In terms of computational efficiency, the GBDT model can filter out irrelevant input features and extract high-level features. The security events in power systems exhibit local characteristics, which are typically influenced by a small subset of key features. Therefore, the GBDT model shows strong adaptability to power systems of varying sizes.

The offline training time complexity of the GBDT model is $O(Kd|\varphi|_0 \log M)$ [40], where K is the number of shallow trees, d is the tree depth, $|\varphi|_0$ is the number of input features, and M is the number of training samples. As the system size S increases, the number of input features increases exponentially, which can be modeled as $|\varphi|_0 = \kappa \cdot S^{\mu}$, where κ and μ are coefficients. In practical implementations, the tree depth d is typically fixed based on the task complexity. The numbers of trees K and training samples M are increased to maintain the prediction accuracy. The growth rate of K is generally less than $\log|\varphi|_0$. The number of training samples tends to scale linearly with the system size.

Based on these assumptions, the approximate offline training time complexity can be expressed as $O(S^{\mu}(\log S)^2)$, which is lower than $O(S^{\mu+1})$. Since the weak learners in the GBDT model are shallow trees with limited decision features per tree, the online assessment time complexity is independent of the number of input features. The online assessment time complexity is $O(Kd)$ [40], which can be approximated as $O(\log S)$, and is lower than linear growth for a system size $O(S)$.

In conclusion, although the offline training time of the model increases with the exponential growth in feature dimensionality as the system size expands, the online assessment remains highly efficient. This ensures the model's feasibility for online applications in large-scale power systems.

To verify the generalizability of the proposed method for wind power uncertainty, cascading failure samples in 10 new wind power scenarios are used for sequential updating models. A training set of 800 samples and a testing set of 200 samples are generated for each scenario. Additionally, updating strategies are considered. The average prediction errors of the maximum DC power limited by the cascading failures for different wind power scenarios are shown in Fig. 9. The results demonstrate that GBDT with the updating strategy has a good generalizability for uncertain scenarios. Moreover, the updating processes of GBDT and SDAE take 54 and 3253 s, respectively, demonstrating the effectiveness of GBDT and its real-time updating strategy.

5.3. Performance validation of FP-Growth

To evaluate the performance of the FP-Growth algorithm, a comparative analysis is conducted using the Apriori and PrefixSpan algorithms. Both FP-Growth and Apriori are frequent itemset mining algorithms, while PrefixSpan is a frequent sequence mining algorithm. A sequence takes the order of items into account, whereas an itemset does not. A dataset comprising 10 000 cascading failures is used for pattern mining. The performance comparison is shown in Table 6.

The results indicate that the FP-Growth algorithm achieves significantly higher computational efficiency than Apriori. This improvement is mainly due to FP-Growth requiring only two full scans of the transaction database and exhibiting a time complexity of $O(m_t n_t + f_t)$, where m_t is the number of transactions, n_t is the average number of items per transaction, and f_t is the time required for recursive mining. In contrast, the Apriori algorithm necessitates multiple database scans and, in the worst-case scenario, has a time complexity of $O(2^{\sigma} m_t)$, where σ is the number of distinct items. Nevertheless, FP-Growth consumes more memory, as it stores the complete FP-Tree in memory. Therefore, while both algorithms yield identical frequent itemsets, FP-Growth is more efficient when sufficient memory resources are available.

A further comparison is conducted between frequent itemset mining and frequent sequence mining algorithms. All items (i.e., transmission lines) are renumbered based on the weights of their frequent 1-itemsets. A subset of the results for 4-item patterns is illustrated in Fig. 10, where the dark-colored bars represent frequent sequences and the light-colored bars denote frequent itemsets. The selected itemsets correspond to ranks 1, 4, 6, 9, 11, and 13 in all item sets. Additionally, the frequent sequences obtained via PrefixSpan are ranked by weight, as shown in Fig. 11.

The PrefixSpan algorithm focuses on mining sequential patterns but lacks the ability to aggregate similar cascading failures into pattern clusters. In contrast, FP-Growth identifies pattern clusters formed by multiple similar sequences, effectively capturing the patterns in similar failure combinations. For instance, the critical pattern in which arbitrary combinations of failures at lines 0, 2, and 4 lead to a failure at line 1 can be represented by sequences such as ⟨0, 2, 4, 1⟩, ⟨2, 0, 4, 1⟩, ⟨2, 4, 0, 1⟩, ⟨4, 0, 2, 1⟩, and ⟨4, 2, 0, 1⟩. The sequence ⟨0, 4, 2, 1⟩ is not a frequent one in the database. These sequences exhibit numerous permutations and relatively uniform weight distribution. FP-Growth effectively captures the role of this pattern cluster, assigning the itemset {0, 1, 2, 4} a rank of 4 among all frequent item sets, which corresponds to a rank of 2 among the selected item sets shown in Fig. 10. In contrast, the sequence ranks obtained via PrefixSpan are 151, 107, 95, 185, and 163 among all sequences, with corresponding ranks of 31, 19, 16, 41, and 36 in Fig. 11. This result demonstrates the superior pattern-recognition capabilities of FP-Growth compared with PrefixSpan.

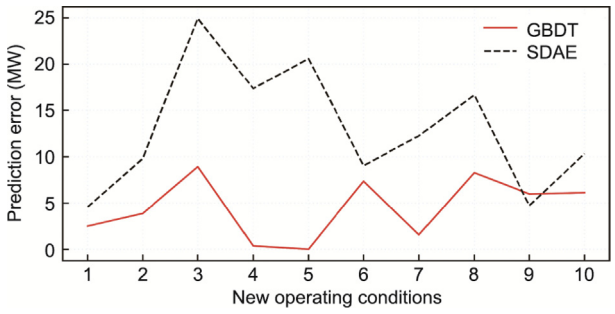

Fig. 9. Prediction errors in different wind power scenarios.

Table 6
Performance comparison of the FP-Growth, Apriori, and PrefixSpan algorithms.

Method	Time (s)	Memory usage (MB)	Mining result
FP-Growth	0.534	17.22	Frequent itemsets
Apriori	404.167	5.71	Same as FP-Growth
PrefixSpan	0.156	0.76	Frequent sequences

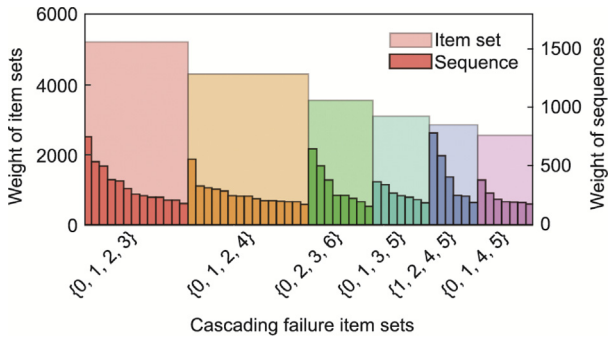

Fig. 10. Results of cascading failure 4-item sets.

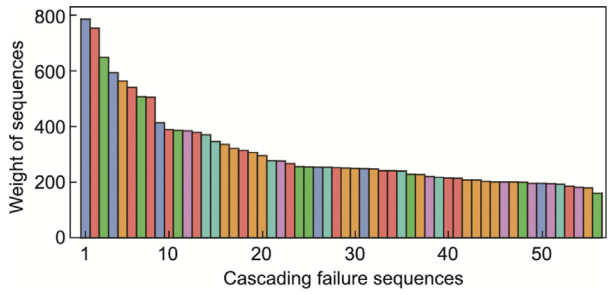

Fig. 11. Results of cascading failure 4-item sequences.

Because PrefixSpan does not merge sequences with different orders of the same items, it produces a large number of similar sequences, hampering the identification of critical failure components. Moreover, the 293 frequent sequences mined by PrefixSpan correspond to only 71 unique item sets, all of which are subsets of the 125 item sets mined by FP-Growth. This result shows that the screening scope of frequent sequence mining is a subset of that of frequent itemset mining.

In the context of cascading failure prevention and control, pattern clusters provide more effective guidance for decision optimization. This enables the development of unified strategies capable of mitigating multiple cascading failure chains.

After the frequent item sets have been identified, the PrefixSpan algorithm and the confidence metric are further employed to mine sequential rules. First, all non-empty subsets are generated from each frequent itemset. Then, the PrefixSpan algorithm is applied to the database to extract frequent sequences formed by these subsets. Finally, the sequential rules are filtered based on a confidence measure.

For example, given the frequent itemset set_0 = {0, 1, 2, 4}, its subsets $set_{0,i}$ include {0}, {1}, {2}, {4}, {0, 1}, {0, 2}, . . ., {0, 2, 4}, {1, 2, 4}, totaling 16 combinations. By applying the PrefixSpan algorithm and filtering based on confidence, the resulting sequential rules are listed in Table 7. This method effectively extracts sequential rules.

5.4. Identification of critical lines

First, the samples in the cascading failure set are formatted. The probability, impact factor, and risk of each cascading failure are calculated. The minimum support is set to 2400 to construct a frequent itemset L = {ND DC link sending-end overvoltage, ND DC link successive commutation failure, line 126–128, line 126–131, line 102–103, line 34–35, line 128–130, line 36–37, line 121–122, line 101–102, line 24–36, line 25–35, line 131–134, . . .}. In the frequent itemset, most of the elements have a high load rate, are in the critical power transmission section of the system, or are close to the converter bus.

All the cascading failures are compressed and ordered according to the frequent itemset. The FP-Tree is built based on the ordered dataset. It contains all the patterns that satisfy the minimum support in cascading failures, so its structure is extremely complex. Then, the conditional FP-Tree for a given fault is constructed from the FP-Tree, thereby compressing the FP-Tree into a tree with patterns related only to the given fault. In this example, the conditional FP-Tree of fault line 131–134 is displayed in Fig. 12.

Based on the conditional FP-Tree, a conditional frequent pattern set corresponding to fault line 131–134 can be constructed. In the conditional FP-Tree, the same short pattern can exist in different long patterns. It is difficult to obtain intuitive information from the tree structure and identify critical lines.

The proposed identification result analysis method is used to calculate the importance of LFPs. This makes it possible to obtain the importance of each line in the LFPs, thereby accurately identifying the critical lines of cascading failures. The importance indexes for some LFPs are shown in Table 8.

Taking LFP_1 = {a, b, c, e} as an example, the importance of each line in the pattern $I(LFP_1, x)$ is {a: 31.53, b: 36.35, c: 16.52, e: 22.15}. According to this result, it can be seen that, in LFP_1, the impact of fault b on event e is the most severe; that is, fault line 126–128 makes the greatest contribution to the ND DC link sending-end overvoltage in the cascading failure caused by LFP_1. This can be explained from the system perspective as follows. Line 126–128 has a high load ratio and strong voltage support for the commutation bus. Once a fault in line 126–128 occurs, it weakens the support strength of the power grid; at the same time, the flow transfer will further affect the remaining lines of the transmission section. This result shows that the proposed method can calculate the importance of each line based on the frequent pattern and its impact on final security events.

The impact of the DTR system [16] is also considered. Under a high wind scenario with substantial wind power generation, traditional fixed-setting protection strategies tend to underestimate the transmission line's overload capacity, leading to overly conservative risk assessments. In actual power systems equipped with DTR, strong wind conditions not only increase renewable generation but also improve the thermal dissipation of transmission lines, thereby increasing their dynamic ampacity.

By establishing a correlation between renewable generation and the transmission capacity of nearby lines, the results indicate that the probability of cascading failures can be reduced by 27%. Future work may further incorporate factors such as overload duration following a disturbance and the initial conductor temperature, in order to refine the influence of DTR. Through appropriate modeling, the current-carrying capability of transmission lines can

Table 7
Results of sequential rules from frequent item sets.

Sequential rule	Confidence
{0, 2, 4} → {1}	0.325
{0, 1, 2} → {4}	0.269
{2, 4} → {0, 1}	0.251
{1, 4} → {0, 2}	0.197
{2} → {0, 4} → {1}	0.154
{1} → {2} → {0} → {4}	0.141

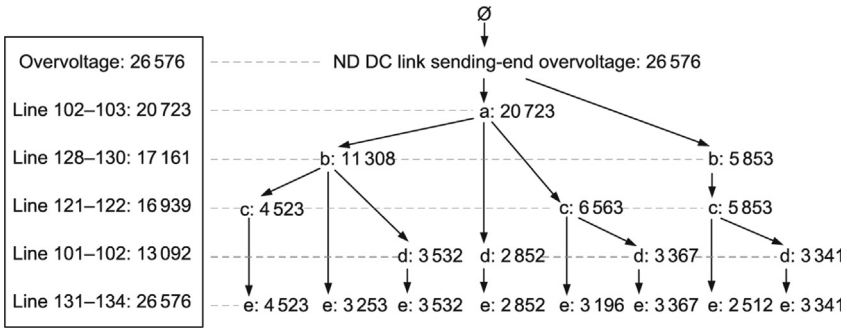

Fig. 12. The conditional frequent pattern tree of fault line 131–134.

Table 8
Importance indexes for LFPs causing ND DC link sending-end overvoltage.

LFP	Importance of each line in LFP
{a, b, c, e}	a: 31.53, b:36.35, c: 16.52, e: 22.15
{a, b, d, e}	a: 29.87, b:38.73, d: 12.61, e: 20.93
{a, c, d, e}	a: 33.21, c:14.35, d: 18.13, e: 21.62

be more accurately evaluated, reducing the error in cascading failure risk estimation and increasing overall system security.

6. Conclusions

Based on the GBDT and FP-Growth algorithms, this work proposed a rapid cascading failure risk assessment method for renewable-energy-dominant hybrid AC/DC power systems. The method was validated on a renewable-dominant hybrid AC/DC system—namely, the modified Ningxia–Shandong interconnected power grid in China. Three main conclusions were drawn from theoretical analyses and simulation results:

(1) Cascading failures can trigger security and stability events in renewable-energy-dominant hybrid AC/DC power systems. These events interact with HVDC systems and significantly limit the DC power with security constraints.

(2) The proposed cascading failure risk index, which focuses on the limitation of DC power, effectively reflects the impact of cascading failures on HVDC systems. The proposed risk assessment method based on GBDT accurately and quickly predicts the maximum DC power.

(3) The improved FP-Growth algorithm exhibits high efficiency in mining frequent patterns from cascading failures. By considering the contribution of faults to the cascading failures, determining the importance index of faults helps in identifying critical lines in cascading failures. The results show that the importance of a fault varies with the final events and cascading failures.

Future work will focus on the early warning of high-risk cascading failures and on preventive and blocking control of cascading failures based on identified critical lines.

CRediT authorship contribution statement

Tianhao Liu: Writing – original draft, Validation, Methodology, Data curation, Conceptualization. **Jiongcheng Yan:** Writing – original draft, Visualization, Resources, Methodology, Formal analysis. **Yutian Liu:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Key Research and Development Program of China "Key technologies for system stability and HVDC transmission of large-scale renewable energy generation base without conventional power support (2022YFB2402700)," and the project of the State Grid Corporation of China (52272222001J).

References

[1] Stanković AM, Tomsovic KL, Caro FD, Braun M, Chow JH, Čukalevski N, et al. Methods for analysis and quantification of power system resilience. IEEE Trans Power Syst 2023;38(5):4774–87.
[2] Liu Y, Fan R, Terzija V. Power system restoration: a literature review from 2006 to 2016. J Mod Power Syst Clean Energy 2016;4(3):332–41.
[3] Vaiman M, Bell K, Chen Y, Chowdhury B, Dobson I, Hines P, et al. Risk assessment of cascading outages: methodologies and challenges. IEEE Trans Power Syst 2012;27(2):631–41.
[4] Liu T, Liu Y. Online cascading failure searching based on gradient boosting decision tree. In: Proceedings of 2023 IEEE Belgrade PowerTech; 2023 Jun 25–29; Belgrade, Serbia. IEEE; 2023. p. 1–6.
[5] Zhang X, Chen C. Maximum available power of multi-infeed HVDC system analysed by sensitivity method. IET Gener Transm Distrib 2014;8(3):473–9.
[6] Liu L, Wu H, Li L, Shen D, Qian F, Liu J. Cascading failure pattern identification in power systems based on sequential pattern mining. IEEE Trans Power Syst 2021;36(3):1856–66.
[7] Mei S, He F, Zhang X, Wu S, Wang G. An improved OPA model and blackout risk assessment. IEEE Trans Power Syst 2009;24(2):814–23.
[8] Yao R, Huang S, Sun K, Liu F, Zhang X, Mei S, et al. Risk assessment of multi-timescale cascading outages based on Markovian tree search. IEEE Trans Power Syst 2017;32(4):2887–900.
[9] Yao R, Huang S, Sun K, Liu F, Zhang X, Mei S. A multi-timescale quasi-dynamic model for simulation of cascading outages. IEEE Trans Power Syst 2016;31(4):3189–201.
[10] Song J, Cotilla-Sanchez E, Ghanavati G, Hines PDH. Dynamic modeling of cascading failure in power systems. IEEE Trans Power Syst 2016;31(3):2085–95.
[11] Zhu Y, Liu T, Li C, Liu Y. Fast probability estimation of HVDC successive commutation failure caused by AC grid cascading failures. Int J Electr Power Energy Syst 2022;135:107618.
[12] Zhu D, Cheng W, Duan J, Wang H, Bai J. Identifying and assessing risk of cascading failure sequence in AC/DC hybrid power grid based on non-cooperative game theory. Reliab Eng Syst Saf 2023;237:109359.
[13] Yan J, Li C, Liu Y. Insecurity early warning for large scale hybrid AC/DC grids based on decision tree and semi-supervised deep learning. IEEE Trans Power Syst 2021;36(6):5020–31.
[14] Zhu Y, Zhou Y, Wei W, Wang N. Cascading failure analysis based on a physics-informed graph neural network. IEEE Trans Power Syst 2023;38(4):3632–41.
[15] Su Y, Teh J. Two-stage optimal dispatching of AC/DC hybrid active distribution systems considering network flexibility. J Mod Power Syst Clean Energy 2023;11(1):52–65.

[16] Lai CM, Teh J. Network topology optimisation based on dynamic thermal rating and battery storage systems for improved wind penetration and reliability. Appl Energy 2022;305:117837.
[17] Lai CM, Teh J, Alharbi B, AlKassem A, Aljabr A, Alshammari N. Optimisation of generation unit commitment and network topology with the dynamic thermal rating system considering N-1 reliability. Electr Power Syst Res 2023;221:109444.
[18] Yang L, Teh J, Alharbi B. Optimizing distributed generation and energy storage in distribution networks: harnessing metaheuristic algorithms with dynamic thermal rating technology. J Energy Storage 2024;91:111989.
[19] Lawal OA, Teh J. Dynamic line rating forecasting algorithm for a secure power system network. Expert Syst Appl 2023;219:119635.
[20] Teh J, Cotton I. Reliability impact of dynamic thermal rating system in wind power integrated network. IEEE Trans Reliab 2016;65(2):1081–9.
[21] Wang A, Luo Y, Tu G, Liu P. Vulnerability assessment scheme for power system transmission networks based on the fault chain theory. IEEE Trans Power Syst 2011;26(1):442–50.
[22] Zhang Z, Yao R, Huang S, Chen Y, Mei S, Sun K. An online search method for representative risky fault chains based on reinforcement learning and knowledge transfer. IEEE Trans Power Syst 2020;35(3):1856–67.
[23] Ma Z, Shen C, Liu F, Mei S. Fast screening of vulnerable transmission lines in power grids: a pagerank-based approach. IEEE Trans Smart Grid 2019;10 (2):1982–91.
[24] Fang J, Su C, Chen Z, Sun H, Lund P. Power system structural vulnerability assessment based on an improved maximum flow approach. IEEE Trans Smart Grid 2018;9(2):777–85.
[25] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998;393(6684):440–2.
[26] Bao Z, Cao Y, Ding L, Han Z, Wang G. Dynamics of load entropy during cascading failure propagation in scale-free networks. Phys Lett A 2008;372 (36):5778–82.
[27] Bompard E, Pons E, Wu D. Extended topological metrics for the analysis of power grid vulnerability. IEEE Syst J 2012;6(3):481–7.
[28] Wang C, Dong Y, Sun P, Lu Y. Cascading failure path prediction based on association rules in cyber-physical active distribution networks. In: Proceedings of 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion; 2020 Dec 11–14; Macau, China. IEEE; 2020. p. 458–64.
[29] Ji X, Li D, Tian X, Hou Y, Meng X. Research on early warning algorithm of active distribution network operational risk based on improved apriori association rules. In: Proceedings of 2021 China Automation Congress; 2021 Oct 22–24; Beijing, China; 2021. p. 2721–6.
[30] Liu Y, Huang S, Mei S, Zhang X. A fast searching method for cascading failure pattern based on prefixspan algorithm. In: Proceedings of 2018 International Conference on Power System Technology; 2018 Nov 6–8; Guangzhou, China; 2018. p. 345–50.
[31] Sun L, Lv Y, Li L, Sun B. An operation risk analysis method based on feature association rule mining algorithm for power system with high proportion renewable energy. In: Proceedings of 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference; 2022 Dec 16–18; Chongqing, China. IEEE; 2022. p. 1858–62.
[32] Jin X, Nian H. Overvoltage suppression strategy for sending AC grid with high penetration of wind power in the LCC-HVDC system under commutation failure. IEEE Trans Power Electr 2021;36(9):10265–77.
[33] Hong L, Zhou X, Xia H, Liu Y, Luo A. Mechanism and prevention of commutation failure in LCC-HVDC caused by sending end AC faults. IEEE Trans Power Del 2021;36(1):473–6.
[34] Chen S, Yao J, Liu Y, Pei J, Huang S, Chen Z. Coupling mechanism analysis and transient stability assessment for multiparalleled wind farms during LVRT. IEEE Trans Sustain Energy 2021;12(4):2132–45.
[35] Yan J, Li C, Liu Y, Yu D, Jia Z. Incremental model evolution for power system security early warning based on knowledge distillation and active learning. IEEE Trans Ind Inform 2024;20(11):12958–68.
[36] Henneaux P. Probability of failure of overloaded lines in cascading failures. Int J Electr Power Energy Syst 2015;73:141–8.
[37] Qi H, Li C, Liu Y, Sun R, Yan J. Continuous feature representation of fault location based on electrical distance for generalizable data-driven dynamic security assessment. IEEE Trans Power Syst. In press.
[38] Han J, Pei J, Yin Y, Mao R. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Min Knowl Discov 2004;8 (1):53–87.
[39] Zhu Y, Li W, Liu Y. Propagation model and blackout risk quantitation analysis for cascading failures in AC/DC hybrid power systems. In: Proceedings of 2019 IEEE Power & Energy Society General Meeting; 2019 Aug 4–8; Atlanta, GA, USA. IEEE; 2019. p. 1–5.
[40] Chen T, Guestrin C, Assoc Comp M. Xgboost: A scalable tree boosting system. In: Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD); 2016 Aug 13–17; San Francisco, CA, USA; 2016. p. 785–94.