A Machine Learning-based Vulnerability Analysis for Cascading Failures of Integrated Power-Gas Systems

Shuai Li, *Student Member, IEEE*, Tao Ding, *Senior Member, IEEE*, Wenhao Jia, *Student Member, IEEE*, Can Huang, *Senior Member, IEEE*, João P. S. Catalão, *Senior Member, IEEE*, Fangxing Li, *Fellow, IEEE*

***Abstract*—This paper proposes a cascading failure simulation (CFS) method and a hybrid machine learning method for vulnerability analysis of integrated power-gas systems (IPGSs). The CFS method is designed to study the propagating process of cascading failures between the two systems, generating data for machine learning with initial states randomly sampled. The proposed method considers generator and gas well ramping, transmission line and gas pipeline tripping, island issue handling and load shedding strategies. Then, a hybrid machine learning model with a combined random forest (RF) classification and regression algorithms is proposed to investigate the impact of random initial states on the vulnerability metrics of IPGSs. Extensive case studies are carried out on three test IPGSs to verify the proposed models and algorithms. Simulation results show that the proposed models and algorithms can achieve high accuracy for the vulnerability analysis of IPGSs.**

***Index Terms*—Integrated power-gas system (IPGS), machine learning, vulnerability analysis, cascading failures**

I. Introduction

The advantages of natural gas power generation, including low emission, strong operation flexibility and high efficiency, have attracted rapid and continuous development in recent years [1]. Hence, there is a growing body of literature that both recognizes the importance of the planning and dispatch of integrated power-gas systems (IPGSs), modeling the effective integrated management and collaborative scheduling of IPGSs. At present, the dominant method of IPGS modeling is based on steady-state energy flow calculation considering different realistic constraints and adopting different algorithms to solve the optimal dispatch of IPGSs [2]. For instance, a novel quasi-dynamic simulation model was established in [3] by using the extensible tool SAInt to analyze the impact of the interdependence between the two systems on the operation stability of IPGSs. Based on the mixed-integer linear programming for the optimal IPGS dispatch model, [4]-[5] have investigated the interaction between the power system and gas network, considering the impact of the gas supply on the power system. To study the coupling models of IPGSs, [6]-[9] modeled gas-fired generators, power-to-gas (P2G) devices, electricity-driven gas compressors and gas storages, and evaluated the impacts on the planning and dispatch of IPGSs.

Meanwhile, the ever-increasing utilization of natural gas power generation has created multiple challenges for the resilience and security of IPGSs. A typical example is a widespread blackout in Texas on February 2021, where over half of the electricity supply came from natural gas. Due to the extreme winter weather, a sudden interruption of the gas supply occurred and natural gas power plants were subsequently shut down, leading to about 35.71% reduction of the power supply rapidly. Meanwhile, due to the shortage of reserve in the system, more than 25 million customers were eventually seriously affected [10]. Therefore, many studies have been conducted on the safety and reliability analysis for IPGSs. For instance, [11] and [12] developed an integrated model to evaluate the impact of the interdependency of power systems and natural gas systems on overall system security, which incorporated the natural gas system constraints into the solution of security-constrained unit commitment. Moreover, a sequential Monte Carlo-based reliability evaluation model for IPGSs considering P2G devices and gas storage was suggested in [13].

Furthermore, many novel models have been established to analyze the vulnerability and resilience of power systems [14]-[15]. Specifically, it verified cascading failure simulation (CFS) as an effective method for studying the mechanisms of local failure propagation and vulnerability analysis. For power systems, the IEEE Working Group on Prediction, Mitigation and Restoration of Cascading Failures (CFWG) has reported a variety of cascading failure modeling approaches [16]-[18]. All these models concentrated on certain assumptions to approximate the actual cascading failures in the power systems. Among them, some research modeled the cascading failure and evaluated the vulnerability metrics employing graph theory and complex network. Based on the complex network theory, [19] developed a cascading failure model considering the node overload failures and hidden failures of transmission lines. Reference [20] proposed a cascading failure model to examine the vulnerability of a specific type of complex network using an extended topological metric. Besides, some studies have utilized the dynamic models for cascading failure analysis, which can capture more dynamic mechanisms, like voltage collapse or transient stability [21]-[22]. Moreover, the steady-state power flow analysis methods have also been widely used to investigate the cascading failure. Based on the AC power

This work was supported in part by National Natural Science Foundation of China (Grant 51977166), in part by Natural Science Foundation of Shaanxi Province (Grant 2021GXLH-Z-059) and in part by the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 with Release Number LLNL-JRNL-827498.

S. Li, T. Ding and W. Jia are with the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049 China (tding15@mail.xjtu.edu.cn).

C. Huang is with Lawrence Livermore National Laboratory, Livermore, CA 94550, USA.

J. P. S. Catalão is with the Faculty of Engineering of the University of Porto (FEUP) and INESC TEC, Porto, Portugal.

F. Li is with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 31996, USA.

flow, [23]-[24] developed a stochastic ORNL-PSerc-Alaska (OPA) model. References [25]-[26] focused on self-organized criticality analysis in CFS by employing the AC power flow-based Manchester model [27] and CASCADE model [28]. In contrast to AC power flow, DC power flow based CFS is significant to balance the model complexity and vulnerability analysis efficiency. For instance, [29] proposed a DC optimal power flow based CFS considering both the topological integrity and the economic aspect. The important consistencies and discrepancies between DC power flow based CFS and the transient stability analysis based CFS was presented in [30].

However, these papers have not dealt with the impact of the natural gas system on cascading failures in IPGSs. In fact, the enhancement of the coupling characteristics of IPGSs contributes to a significant impact on the vulnerability analysis of cascading failures. A certain contingency occurring in one system may lead to serious failures of components in the other system, resulting in severe cascading failures propagating between the two systems [31]-[32]. Therefore, with the significantly increased coupling of IPGSs, it is urgent and critically important to investigate the cascading failure model for IPGSs. Recently, a considerable amount of literature have been published on the cascading failure model for IPGSs [33]-[36]. Based on a graph theory-based methodology, [33] studied the cascading failures in real interdependent power and natural gas networks with an efficient vulnerability analysis technique. Reference [34] proposed a cascading failure model to evaluate the structural robustness and performance against random failures of the IPGSs using graph theory. A novel reliability evaluation model of IPGSs considering the impact of cascading failures was built in [35] by employing the dynamic cascading analysis model and Monte Carlo simulation methods. Furthermore, a CFS model was set up in [36] to study the propagation of cascading failure caused by various contingencies, where the steady-state power flow and dynamic gas transmission were incorporated and combined together.

Nevertheless, the previous studies have only simply analyzed the impact of the gas system on the cascading failure of the power system, while neglecting to investigate the propagating process of cascading failures in an IPGS, particularly in the gas system. Moreover, few researchers have studied the impact the random changes in initial states on CFS for IPGSs (i.e., uncertain power and gas load levels, generator outputs, gas well supply and consumption of P2G devices). These uncertainties affect the cascading failure results, while leading to a high computational burden to perform CFS for all possible initial states of IPGSs in online applications. Recently, machine learning methods have been widely used in power systems [37]. To address the above challenges, a CFS model and a machine learning method for vulnerability analysis of IPGSs are proposed in this paper.

The main contributions can be summarized as follows:

(i) A steady-state energy flow-based CFS model is set up to investigate the propagating process of cascading failures under deterministic initial states, which considers generator and gas well ramping, transmission line and gas pipeline tripping, island issue handling and load shedding strategies. Meanwhile, the CFS model is performed to generate data for machine learning with initial states randomly sampled.

(ii) A hybrid machine learning model with the combined random forest (RF) classification and regression algorithms is proposed for vulnerability analysis of IPGSs under uncertain initial states. In this model, the RF classification is used to classify data by the number of events in a cascading failure, and then the RF regression is utilized for each class of data to characterize the relationship between random initial states and vulnerability metrics.

The remainder of the paper is organized as follows: Section II develops a CFS model for vulnerability analysis of IPGSs where the CFS procedure is presented in detail. Section III proposes a hybrid machine learning method with combined RF classification and regression for assessing the impacts of uncertain initial states on vulnerability metrics for cascading failures. Simulation results on three test systems are presented and discussed in Section IV. Section V presents the conclusion.

II. Vulnerability Analysis Model of Cascading Failures for IPGSs

Through CFS for IPGSs, we can identify the vulnerable part of the IPGS, which is important for the planning and dispatch. In this section, a CFS model based on steady-state power flow and gas flow calculation is established, which considers generator and gas well ramping, transmission line and gas pipeline tripping, island issue handling and load shedding strategies. Finally, two vulnerability metrics are proposed to evaluate the system vulnerability under cascading failures.

A. Cascading Failure Simulation

For the CFS of IPGSs, different locations and types of initial failures will lead to different results. Here, we consider a situation in which the initial failure occurs on a power bus, which means that all components connected to this damaged bus will stop operating and be triggered by the system. The tripping of a gas pipeline is considered as an additional initial contingency.

Meanwhile, considering that the response and propagation speed of the power system to cascading failures are much faster than those of the natural gas system, we assume that the power system will quickly reach a steady state after a series of power cascading failure events (PCEs), and then analyze the impact of the failure on the natural gas system. In contrast, if a gas cascading failure event (GCE) in the natural gas system causes the redistribution of gas flow, we will immediately analyze the cascading failure of the power system through the change of gas flow at the coupling nodes of the IPGSs. In this way, the dynamic sequential process of cascading failures can be accurately described, and the final steady-state operation of the IPGS can be realized by the following operations.

(i) Generator and gas well ramping

When the output of power generators or gas wells is unbalanced with the power load or gas load due to a failure, the generators or gas wells will perform ramping according to the following two cases:

Case 1: If the total output of power generators or gas wells is larger than the power loads or gas loads, all generators or gas

wells will gradually reduce the output at a ramp rate of r_g or r_w at a period of Δt_P or Δt_G, respectively.

Case 2: If the total output of power generators or gas wells is less than the power loads or gas loads, all generators or gas wells will gradually increase the output at a ramp rate of r_g or r_w at a period of Δt_P or Δt_G, respectively.

It should be noted that when a generator or gas well reaches the maximum or minimum output, it will stop ramping and maintain the maximum or minimum output.

(ii) Energy flow redistribution and overload lines tripping

If the IPGS can balance the power or natural gas by ramping power and gas generation, the new distribution of power and gas flow is recalculated through the following power flow and flow calculation model.

1) DC power flow model:

$$\begin{cases} PF_l(t) = \left(\theta_i(t) - \theta_j(t)\right) / X_l \\ \theta_0(t) = \theta_0^{ref}, \qquad i, j \in \Omega_{BUS}, l \in \Omega_{LINE} \end{cases} \tag{1}$$

where $PF_l(t)$ is the power flow on the power transmission line l at time t. $\theta_i(t)$ is the voltage angle of bus i at time t. X_l is the reactance of transmission line l. $\theta_0^{ref}(t)$ is the voltage angle at the reference bus. Ω_{BUS} and Ω_{LINE} are the sets of power buses and transmission lines, respectively.

2) Gas flow model:

$$\begin{cases} GF_e^2(t)\,\mathrm{sgn}(\pi_m(t) - \pi_n(t)) = \tau_e(\pi_m^2(t) - \pi_n^2(t)) \\ \pi_0(t) = \pi_0^{ref} \\ \pi_n / \pi_m = \lambda_c, \quad m, n \in \Omega_{NODE}, e \in \Omega_{PIPE}, c \in \Omega_{COMP} \end{cases} \tag{2}$$

where Ω_{NODE}, Ω_{PIPE} and Ω_{COMP} are the sets of gas nodes, pipelines and compressors, respectively. $GF_e(t)$ is the gas flow on the gas pipeline e at time t. $\pi_m(t)$ is the nodal pressure of node m at time t. τ_e is the constant of pipeline e related to many physical factors. $\pi_0^{ref}(t)$ is the given pressure at the reference node. λ_c is the gas pressure boost ratio of the gas compressor c.

3) Coupling components model:

Gas-fired generators and P2G devices are two typical coupling components in IPGSs. Gas-fired generators use natural gas as a fuel to generate electricity, and the energy conversion meets the following requirements:

$$F_{f,m}(t) = \beta_f P_{f,m}(t) / GHV,\ f \in \Omega_{GAS}, m \in \Omega_{NODE} \tag{3}$$

where Ω_{GAS} is the set of gas-fired generators. GHV is the high calorific value of the natural gas. $F_{f,m}(t)$ and $P_{f,m}(t)$ are the gas consumption and electric power output at coupling node m. β_f is the consumption coefficient of the gas-fired generator f.

P2G devices consume electrical power to generate natural gas by electrochemical reactions. The energy conversion relationship can be expressed as:

$$F_{d,m}(t) = \eta_d P_{d,m}(t) / GHV, d \in \Omega_{P2G}, m \in \Omega_{NODE} \tag{4}$$

where Ω_{P2G} is the set of P2G devices. $F_{d,m}(t)$ and $P_{d,m}(t)$ are the electric power consumption and gas output at coupling node m at time t. η_d is the conversion efficiency of the P2G device d.

Tripping overload transmission lines or pipelines is the main reason to promote the continuous propagation of cascading failures. Through the power or gas flow calculation, the redistribution of power or gas flow is obtained. If the power or gas flow on the transmission lines or gas pipelines exceeds its transmission capacity, the relay will activate to trip the overload lines or pipelines to avoid damage to the system.

(iii) Generator or gas well tripping and load shedding

Due to the limited regulating capacity of generators and gas wells, if the power or gas balance cannot be satisfied within the given time T^{SP} or T^{SG}, the system will trip the generator, gas well or load. Specifically, there are two cases after T^{SP} or T^{SG} for dealing with the unbalance:

For Case 1, if there is a remaining surplus of output power supply or gas source, the system will trip the generators or gas wells in the merit order according to the importance of generators or gas wells from the small to the large, until $\sum_{i\in\Omega_{BUS}} P_{G,i}(t) \le \sum_{i\in\Omega_{BUS}} P_{L,i}(t)$ or $\sum_{m\in\Omega_{NODE}} F_{W,m}(t) \le \sum_{m\in\Omega_{NODE}} F_{L,m}(t)$.

For Case 2, if the output of power supply or gas source still cannot meet the load demand, the system will perform load shedding from the least important load until a new balance is established. Moreover, the quantity of load shedding will be determined by the following optimization model.

1) Optimal power load shedding model:

$$\min \sum_{i\in\Omega_{BUS}} C_i^P \left(P_{L,i}(t) - P_{L,i}^s(t)\right) \tag{5}$$

$$s.t. \quad \sum_{p\in\Omega_{CON}} P_{p,i}(t) + \sum_{f\in\Omega_{GAS}} P_{f,i}(t) - \sum_{d\in\Omega_{P2G}} P_{d,i}(t) - P_{L,i}^s(t) + \sum_{l\in\Omega_{LINE}^{in}} PF_{l,i}(t) - \sum_{l\in\Omega_{LINE}^{out}} PF_{l,i}(t) = 0, \quad i \in \Omega_{BUS}, l \in \Omega_{LINE} \tag{6}$$

$$PF_l(t) = \left(\theta_i(t) - \theta_j(t)\right) / X_l, \quad i, j \in \Omega_{BUS}, l \in \Omega_{LINE} \tag{7}$$

$$P_{p,i}(t - \Delta t_P) - r_p \Delta t_P \le P_{p,i}(t) \le P_{p,i}(t - \Delta t_P) + r_p \Delta t_P \qquad i \in \Omega_{BUS}, p \in \Omega_{CON} \tag{8}$$

$$P_{f,i}(t - \Delta t_P) - r_f \Delta t_P \le P_{f,i}(t) \le P_{f,i}(t - \Delta t_P) + r_f \Delta t_P \qquad i \in \Omega_{BUS}, f \in \Omega_{GAS} \tag{9}$$

$$0 \le P_{L,i}^s(t) \le P_{L,i}(t), i \in \Omega_{BUS} \tag{10}$$

where C_i^P is the weight of the power load shedding for load bus i. $P_{L,i}^s(t)$ is the load power at bus i and time t after load shedding. Ω_{CON} is the set of conventional generators. Ω_{LINE}^{in} and Ω_{LINE}^{out} are the sets of transmission lines flowing in or out of bus i. $P_{p,i}(t)$ and $P_{p,i}(t)$ are the output of conventional generator cg and gas-fired generator p at bus i and time t. r_p and r_f are the ramping rate of conventional generator p and gas-fired generator f.

2) Optimal gas load shedding model:

$$\min \sum_{m\in\Omega_{node}} C_m^G \left(F_{L,m}(t) - F_{L,m}^s(t)\right) \tag{11}$$

$$s.t. \quad \sum_{w\in\Omega_W} F_{w,m}(t) + \sum_{d\in\Omega_{P2G}} F_{d,m}(t) - \sum_{f\in\Omega_{GAS}} F_{f,m}(t) - F_{L,m}^s(t) + \sum_{e\in\Omega_{PIPE}^{in}} GF_{e,m}(t) - \sum_{e\in\Omega_{PIPE}^{out}} GF_{e,m}(t) = 0, \ m \in \Omega_{NODE}, e \in \Omega_{PIPE} \tag{12}$$

$$GF_e^2(t) = \mathrm{sgn}(\pi_m(t) - \pi_n(t))\tau_e(\pi_m^2(t) - \pi_n^2(t)) \qquad e \in \Omega_{PIPE}, m, n \in \Omega_{NODE} \tag{13}$$

$$F_{w,m}(t)(t - \Delta t_G) - r_w \Delta t_P \le F_{w,m}(t) \le F_{w,m}(t - \Delta t_G) + r_w \Delta t_G \qquad m \in \Omega_{NODE}, w \in \Omega_W \tag{14}$$

$$0 \le F_{L,m}^{s}(t) \le F_{L,m}(t), \quad m \in \Omega_{NODE} \tag{15}$$

$$\underline{F}_f \le F_{f,m}(t) \le \overline{F}_f, f \in \Omega_{GAS}, m \in \Omega_{NODE} \tag{16}$$

$$\underline{F}_d \le F_{d,m}(t) \le \overline{F}_d, d \in \Omega_{P2G}, m \in \Omega_{NODE} \tag{17}$$

$$\underline{\pi}_m(t) \le \pi_m(t) \le \overline{\pi}_m(t), m \in \Omega_{NODE} \tag{18}$$

$$\pi_m^2(t) \le \gamma_c^2 \pi_n^2(t), \quad m,n \in \Omega_{NODE}, \ c \in \Omega_{COMP} \tag{19}$$

where C_m^G is the weight of the gas load shedding for load node m. $F_{L,m}^{s}(t)$ is the load power at node m and time t after load shedding. Ω_W is the set of gas wells. Ω_{PIPE}^{in} and Ω_{PIPE}^{out} are the sets of pipelines flowing into or out of node m. $F_{w,m}(t)$ and $F_{d,m}(t)$ are the gas output of gas well w and P2G device d at bus i and time t. γ_c is the maximum gas pressure boost ratio of the gas compressor c.

In the optimal gas load shedding model, (13) is a nonlinear formula. Therefore, the piecewise linearization method is utilized to linearize optimization model [38]. Firstly, $\Pi_m(t)=\pi_m^2(t)$ is introduced in (13) and (18). Then, $GF_e^2(t)$ in (13) can be linearized by the following formula:

$$\begin{aligned} GF_e^2(t) &= GF_{e,1}^2(t) + \sum_{k=1}^{N_{seg}-1} \left[GF_{e,k+1}^2(t) - GF_{e,k+1}^2(t) \right] \mu_{e,k}(t) \\ &= \tau_e(t)(\Pi_m(t) - \Pi_n(t)), e \in \Omega_{PIPE}, m,n \in \Omega_{NODE} \end{aligned} \tag{20}$$

$$\begin{aligned} &0 \le \mu_{e,k}(t) \le 1, \ \ \mu_{e,k+1}(t) \le \psi_{e,k}(t) \le \mu_{e,k}(t) \\ &e \in \Omega_{PIPE}, \ \ k = 1,2,...N_{seg}-1 \end{aligned} \tag{21}$$

where N_{seg} is the number of linear segments, $\mu_{e,k}(t)$ is the portion of segment k, and $\psi_{e,k}(t)$ is a binary variable satisfying the conditions: if $\psi_{e,k}(t)=1$, $\mu_{e,k}(t)=1$ and $0 \le \mu_{e,k+1}(t) \le 1$; if $\psi_{e,k}(t)=0$, $\mu_{e,k}(t)=\mu_{e,k+1}(t)=1$.

(iv) Islanding issues

In CFS, overload line tripping may cause islands in either the power system or the gas system. Each island can operate independently and continue to propagate cascading failures, and finally reach a steady-state operation point. However, when cascading failures propagate in each island, the response times $t_{P,pi}$ or $t_{G,gi}$ of different islands are different. In order to realize the synchronization of the simulation process of each island in the same PCE or GCE, we select the minimum value $t_{P,pi}^{min}$ or $t_{G,gi}^{min}$ of $t_{P,pi}$ or $t_{G,gi}$ in each PCE or GCE as the simulation time step to promote further CFS.

B. Vulnerability Metrics

In order to quantify the vulnerability of power buses or gas pipelines in cascading failures, the power and gas load shedding ratio are proposed as vulnerability metrics, which can be expressed as follows:

$$LOSS_P = \sum_{i \in \Omega_{BUS}} \left(P_{L,i}^0 - P_{L,i}^\infty \right) \Big/ \sum_{i \in \Omega_{BUS}} P_{L,i}^0 \tag{22}$$

$$LOSS_G = \sum_{m \in \Omega_{NODE}} \left(F_{L,m}^0 - F_{L,m}^\infty \right) \Big/ \sum_{m \in \Omega_{NODE}} F_{L,m}^0 \tag{23}$$

where $P^0_{L,i}$ and $P^\infty_{L,i}$ are the initial and final power load at bus i. $F^0_{L,m}$ and $F^\infty_{L,m}$ represent the initial and final gas load at node m.

C. Flowchart of CFS for IPGS

The above model describes the sequence process in a PCE or GCE. Combined with the assumed time scale of cascading failures in the two systems, we can obtain the whole flowchart of CFS for IPGS as shown in Fig. 1, and the specific steps are summarized as follows:

Step 1: Read the initial state parameters of IPGS and determine the initial failure location.

Step 2: Update the topology of the power system. For each unbalanced power system island, generator ramping, power load shedding or generator tripping are adopted to establish a new operating point with a power balance.

Step 3: Calculate power load shedding in the PCE and power redistribution via (5)-(10). Check the power flow on each transmission line: if the transmission line exceeds the limit, trip overload lines and go back to **Step 2**; otherwise, go to **Step 4**.

Step 4: Calculate the changes of gas flow at the coupling nodes according to the gas flow via (3)-(4). For each unbalanced gas island, gas well ramping, gas load shedding or gas well tripping is adopted to establish a new gas balance.

Step 5: Calculate gas load shedding in the GCE, gas redistribution via (11)-(21), and gas flow at the coupling nodes via (3)-(4). If the gas flow at the coupling nodes is different from that in **Step 3**, go back to **Step 2** and analyze the new possible cascading failures in the power system; otherwise, go to **Step 6**.

Step 6: Check the gas flow on each pipeline, if the pipeline exceeds the limit, trip the overload pipeline and go back to **Step 4**; otherwise, go to **Step 7**.

Step 7: Calculate the power and gas load shedding throughout the process of CFS and the vulnerability metrics of the power bus or gas pipelines in the IPGS.

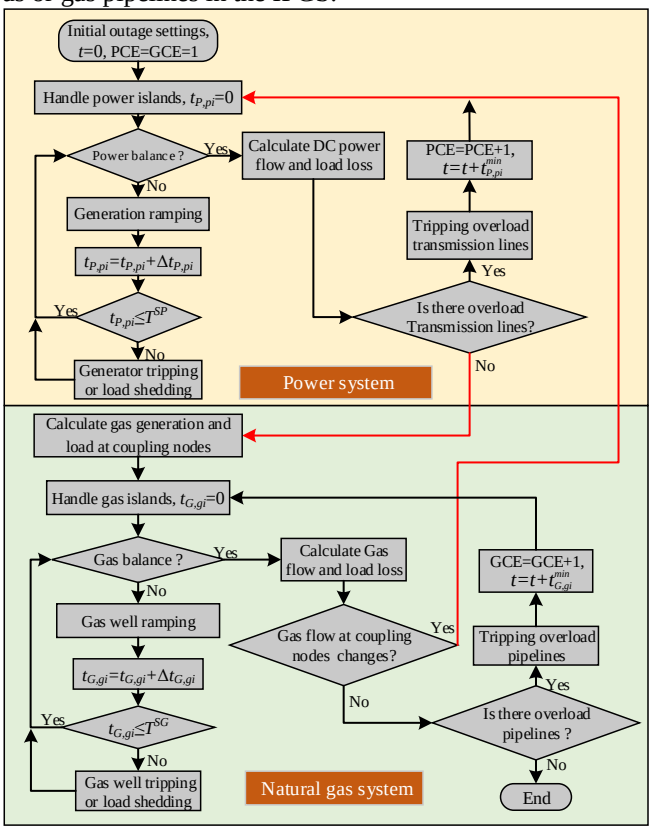


Fig. 1 Flowchart of CFS for IPGS

III. Cascading Failures for IPGS with Random Initial States based on Machine Learning Method

A. Data Collection and Preprocessing

We determined that the vulnerability metrics for the above CFS for IPGSs are greatly influenced by the initial system states, i.e., the uncertain power and gas load levels, generator outputs, gas well supply and consumption of P2G devices. However, the initial states may be randomly changed in the practical engineering problem, which may require a great number of CFSs to evaluate the impacts of these uncertain initial states on vulnerability metrics. This results in a heavy computational burden. Therefore, we employ the machine learning method to investigate the impact of uncertain initial states on the vulnerability analysis of CFS for IPGSs in this paper.

Since the probability of N-k contingencies in an IPGS is low, the amount of corresponding historical data is extremely limited. The following steps are performed to collect sufficient experimental data: 1) Choose a scenario with a random initial state within a certain range; 2) Conduct the CFS following the flowchart depicted in Fig. 1 and calculate the vulnerability metrics for the given initial state; 3) Obtain the input features of each scenario x_i (i.e., uncertain power and gas load levels, generator outputs, gas well supplies and P2G consumption), classification labels y_i (i.e., the different numbers of PCEs and GCEs), and output labels z_i (i.e., vulnerability metrics).

In order to unify the dimension and ensure the reliability of training results, the z-score standardization algorithm is employed to map all features and labels to (0,1), given by:

$$x_i^* = \frac{x_i - \mu}{\sigma}, \quad i = 1,2...,N_{data} \tag{24}$$

where N_{data} is the number of data sets, μ and σ are the mean value and standard deviation, respectively.

According to the cascading failure model in Section II, the cascading failure process of an IPGS consists of a series of PCEs and GCEs. Under different initial states, the load loss may increase significantly due to the increasing numbers of PCEs and GCEs, and the corresponding vulnerability metrics may take on a hierarchical structure, resulting in an uneven data distribution. Directly conducting regression training would lead to regression inaccuracy due to the uneven distribution of the data. Therefore, this paper proposes a hybrid machine learning model with combined classification and regression, where an RF classification algorithm classifies the data into several classes according to the numbers of PCEs and GCEs. For each class, the RF regression algorithm is used for regression training to investigate the impact of random initial states on vulnerability metrics of IPGSs.

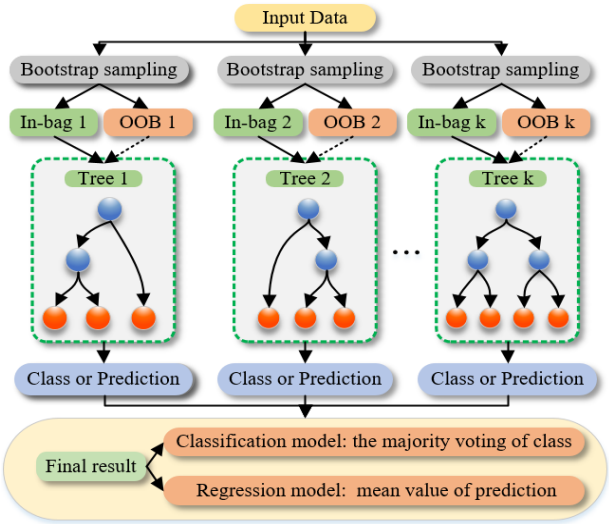


Fig.2 Schematic diagram of RF classification and regression model

B. Random Forest Classification and Regression Model

As shown in Fig. 2, the RF is an ensemble algorithm based on decision trees, which can be used for either classification or regression [39].

(i) RF classification and regression model

For the M data $D = \{(x_i, y_i, z_i) \mid i = 1,2,...,M\}$, where $x_i \in R^{N_f}$ is the initial state feature set, y_i is the classification label, and z_i is the regression label. First, we perform the RF classification model by the following steps:

Step 1: Adopt the bootstrap sampling method to select k_c sub training sets (in-bag data) to construct k_c independent classification decision trees. The unsampled data is called out-of-bag (OOB) data.

Step 2: For each node of trees, randomly select N_{try} features as splitting features, and select the optimal splitting method.

Step 3: Traverse all trees to get the classification result of each tree, and take the voting for all trees as the final RF classification result.

Then, for the data of class j after the above classification $\{(x_i, y_{i,j}, z_i) \mid i = 1,2,...,M_j\}$, perform the RF regression model. Compared with the RF classification model, the proposed RF regression model replaces the classification decision trees with k_r independent regression decision trees, and takes the average prediction value of all regression trees as the final regression prediction value.

(ii) RF feature ranking and selection

In the vulnerability analysis for an IPGS, there are many initial state features (i.e., power and gas load levels, generator outputs, gas well supplies and P2G consumption). However, some features irrelevant to labels may lead to overfitting problems in either classification or regression training due to data redundancy, and eventually decrease the accuracy of classification and regression prediction [40]. Hence, it is necessary to rank and select features of the classification and regression training.

For RF classification, we use OOB data to evaluate the importance of features, according to the following formula:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3119237, IEEE Transactions on Power Systems

$$Import_c = \frac{1}{k_c}\sum_{p=1}^{k_c}\left(R_p^{OOB} - R_{p,f}^{OOB}\right),\ f = 1,2,...,N_f \quad (25)$$

where N_f is the number of features. R_1^{OOB}, R_2^{OOB},…, $R_{k_c}^{OOB}$ and $R_{1,f}^{OOB}$, $R_{2,f}^{OOB}$,…, $R_{k_c,f}^{OOB}$ are the numbers of data correctly classified of each classification tree before and after the rearrangement of the *f*-th feature.

For the RF regression model after the classification, it is also necessary to calculate and rank the importance of features. Compared with the RF classification model, the RF regression model calculates the regression prediction error E_1^{OOB}, E_2^{OOB},…, $E_{k_r}^{OOB}$ and $E_{1,f}^{OOB}$, $E_{2,f}^{OOB}$,…, $E_{k_r,f}^{OOB}$ before and after the rearrangement of the *f*-th feature. Then, the importance of the *f*-th feature in the RF regression can be formulated as:

$$Import_r = \frac{1}{k_r}\sum_{q=1}^{k_r}\left(E_q^{OOB} - E_{q,f}^{OOB}\right),\ f = 1,2,...,N_f \quad (26)$$

By this method, the importance of features for both RF classification and RF regression can be ranked in descending order until the accuracy of classification and regression reaches the maximum value. At this time, the optimal number of features can be determined.

(iii) RF classification and regression metrics

Generally, we use the cross-validation method to divide the data into a training set and a test set, utilize the training set to train the model, and verify the accuracy of the model on the test set. Since the bootstrap method used in the RF algorithm samples the same number of training subsets as the original data set for replacement from the original data, we can use OOB data as the test set to evaluate the model without additional cross validations or independent test sets.

In order to evaluate the RF classification model, the confusion matrix is introduced here, including: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). However, the number of PCEs and GCEs will lead to an imbalanced distribution of vulnerability metrics data. To solve the imbalanced multi-classification problem, this paper introduces an F1 score and uses OOB data to evaluate the performance of RF classification. The F1 score is the harmonic mean of Precision (PRE) and Recall (REC), which can evaluate multi-classification problems. For each class *j*, PRE, REC and the F1 score can be calculated as:

$$PRE = \frac{\sum_{j\in N_C} TP_j}{\sum_{j\in N_C} TP_j + FP_j} \quad (27)$$

$$PEC = \frac{\sum_{j\in C} TP_j}{\sum_{j\in C} TP_j + FN_j} \quad (28)$$

$$F1 = \frac{2\left(PRE \times REC\right)}{PRE + REC} \quad (29)$$

where *Nc* is the number of classes.

For the RF regression model, the coefficient of determination R^2 is employed to verify the accuracy of prediction, given by:

$$R^2 = 1 - \sum_{p=1}^{N_{OOB}}(y_p - \hat{y}_p)^2 \Big/ \sum_{p=1}^{N_{OOB}}(y_p - y_{mean})^2 \quad (30)$$

where N_{OOB} is the number of OOB data, y_{mean} is the mean value of all true values in the OOB data. y_p and $\hat{y}_p$ are the mean values and predicted values of the *p*-th OOB data.

The important parameters affecting the performance of RF algorithm mainly include: the number of decision trees, the maximum depth of decision trees and the minimum number of samples for leaf nodes. In this paper, the grid search method is used to obtain all possible combinations of four important parameters in a certain range. Then, all parameter combinations are trained by RF classification and regression algorithm to obtain the best parameter combination with the highest prediction accuracy.

C. Flowchart of the Proposed Machine Learning Model

The procedure of the proposed machine learning method combining an RF classification and an RF regression algorithm is presented in Fig. 3 with the specific steps summarized as:

Step 1: Obtain sufficient data through the CFS method detailed in Fig. 1 under random initial states of the IPGS. Perform the data preprocessing via the z-score standardization algorithm according to (22).

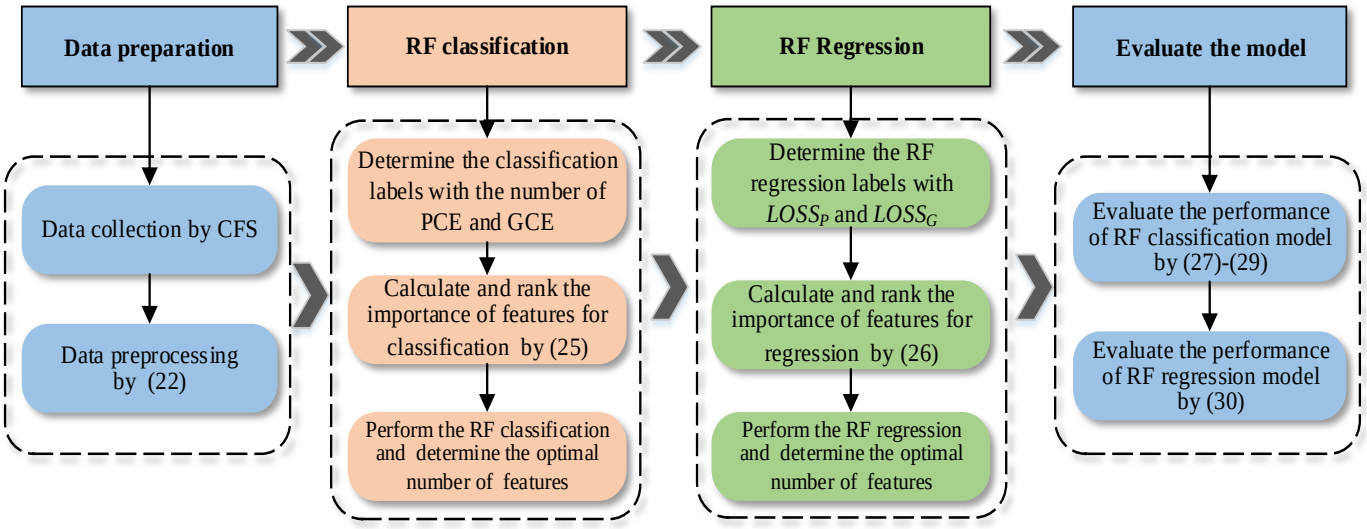


Fig. 3 Flowchart of the proposed machine learning method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3119237, IEEE Transactions on Power Systems

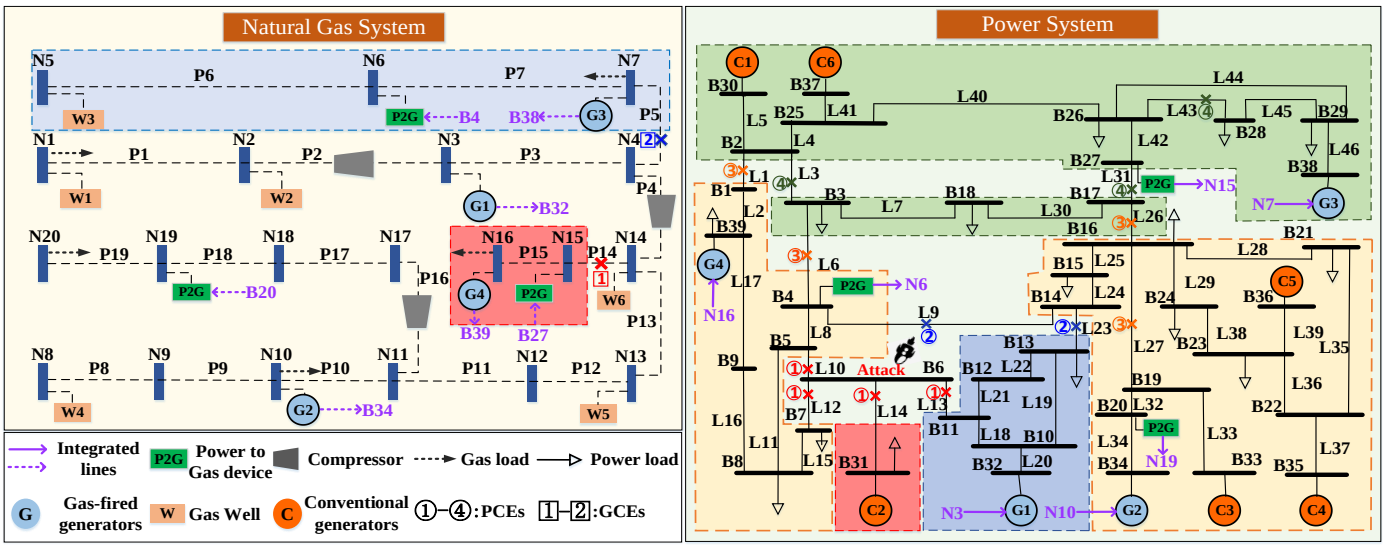


Fig. 4 CFS on the 39-bus-20-node System at bus 6

Step 2: Calculate and rank the importance of features for RF classification according to (23). Conduct RF classification and determine the optimal number of features according to the results of features ranking and classification training.

Step 3: For each class of data classified in **Step 2**, calculate and rank the importance of features for RF regression according to (24). Conduct RF regression and determine the optimal number of features according to the results of features ranking and regression training.

Step 4: Use the OOB data as a test set to evaluate the performance of the RF classification model via (26)-(28) and the RF regression model via (29).

IV. Case Studies

A. *CFS on the 39-bus-20-node System*

In order to demonstrate the performance of the proposed CFS model, an IPGS consisting of an IEEE 39-bus power system and a Belgium 20-node gas system [41] is introduced, the topology of which is depicted in Fig. 4. The two systems are coupled by 4 gas-fired generators and 3 P2G devices. The gas-fired generators on buses B32, B34, B38 and B39 are supplied by gas system at nodes N3, N10, N7 and N16. The three P2G devices driven by electricity on buses N4, N20 and N27 will inject gas into nodes N6, N19 and N15 of the gas system. In the process of CFS, the following parameters are set: 1) The generator ramp rate r_g and gas well ramp rate r_w at each time slot Δt_P and Δt_G are set as 5% of the total capacity of the corresponding generator and gas well. 2) The importance of generators, gas wells and load is proportional to the corresponding capacity. 3) The transmission capacity of each pipeline is 1.5 times the gas flow under the rated initial state.

First, we take the initial failure that occurred at bus 6 as an example to analyze the propagation process of the cascading failure in the IPGS. The results are shown in Fig. 4 and Table I. In Fig. 4, the numbers with different colors in the two systems represent the tripping of transmission lines and pipelines under different cascading failure events (PCEs and GCEs). Dashed boxes with different colors correspond to power or gas islands formed in different PCEs and GCEs. With the increase of the number for PCEs and GCEs, new islands are constantly formed, and the cascading failure process is promoted. The whole cascading failure process consists of four PCEs and two GCEs, where GCE1 occurs between PCE3 and PCE4, and GCE2 occurs after PCE4. After the initial failure occurs, the three transmission lines (i.e., L10, L12 and L13) and one transformer (i.e., L14) connected to the bus B6 are tripped, resulting in two islands and 0.82% $LOSS_P$. The island with only bus B31 can achieve power balance through generator ramping. For another island with the other 37 buses, the transmission lines L9 and L23 are tripped after the redistribution of power flow, resulting in 3.17% $LOSS_P$ and further splitting into two new islands. The island composed of buses B10, B11, B12, B13 and B32 can reach power balance without any load loss. The island with the other 32 buses splits into three islands due to the tripping of transmission lines L1, L6, L26 and L27, leading to 9.95% $LOSS_P$. Subsequently, all islands reach balance temporarily, and no transmission lines exceed the capacity. Thus, this cascading failure stops.

At this time, the gas power flow is calculated according to the power flow at the coupling node, and it can be found that gas pipeline P14 is overloaded and tripped, resulting in two gas islands (i.e., one island with nodes N15 and N16, and the island with the other nodes) and 5.34% $LOSS_G$. According to the redistribution of gas flow, the power flow can be calculated through the coupling nodes immediately. Then, transmission lines L3, L31 and L43 are overloaded and tripped, leading to two new islands and 17.27% $LOSS_P$. At this time, all power islands are balanced again. By calculating the power change at the coupling nodes, the redistribution of natural gas flow is obtained. The gas pipeline P5 is overloaded and tripped, resulting in two gas islands (i.e., the island with nodes N5, N6, and N7, and the island with the other nodes) and 7.15% $LOSS_G$. Then, all islands in the IPGS achieve balance and cascading failures end, resulting in 31.21% $LOSS_P$ and 12.49% $LOSS_G$ in total. It can be observed that the cascading failure in an IPGS can propagate between the two systems through the energy flow variation at the coupling nodes.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3119237, IEEE Transactions on Power Systems

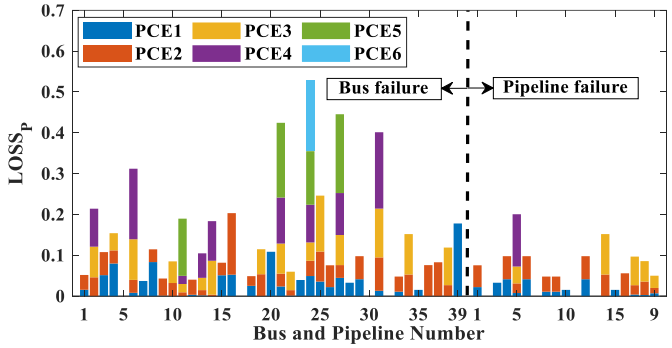


Fig. 5 $LOSS_P$ for all buses and pipelines

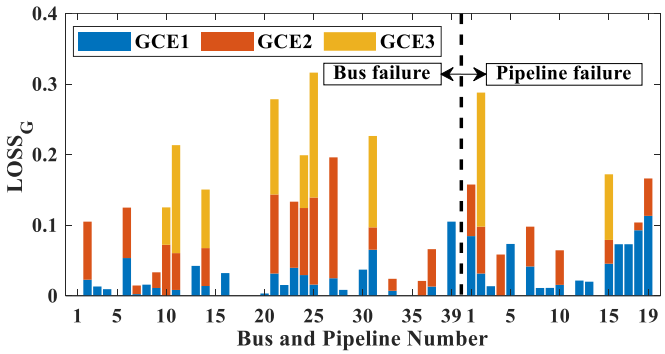


Fig. 6 $LOSS_G$ for all buses and pipelines

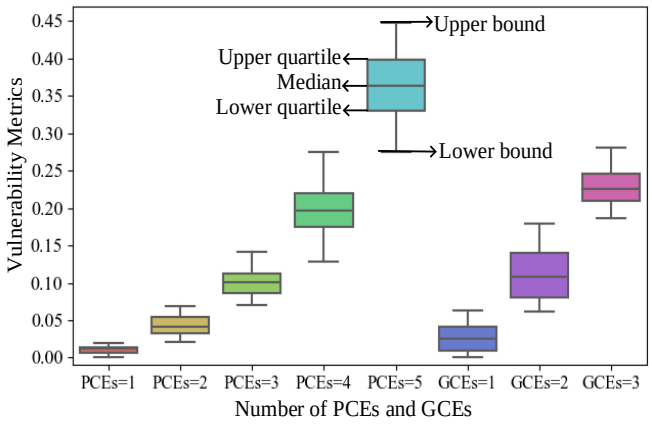
TABLE I ISLANDS AND LOAD LOSS OF CFS AT BUS 6

PCE or GCE	Overload Lines	islands	Buses or nodes in each island	$LOSS_P$ or $LOSS_G$
PCE 1	L10,L12, L13,L14	2+1	B31 Other buses	0.82%PL
PCE 2	L9,L23	3+1	B31 B10,B11,B12,B13,B32 Other buses	3.17%PL
PCE 3	L1, L6, L26,L27	5+1	B31 B10,B11,B12,B13,B32 B1,B4,B5,B7, B8,B9,B39 B2,B3,B17,B18,B25, B26,B27,B28,B29,B30, B37,B38 Other buses	9.95%PL
GCE 1	P14	5+2	31 B10,B11,B12,B13,B32 B1,B4,B5,B7, B8,B9,B39 B2,B3,B17,B18,B25, B26,B27,B28,B29,B30, B37,B38 Other buses N15,N16 Other nodes	5.34%GL
PCE 4	L3,L31, L43	6+2	31 B10,B11,B12,B13,B32 B1,B4,B5,B7, B8,B9,B39 B3,B17,B18 B2,B25,B26,B27, B28,B29,B30,B37,B38 Other buses N15,N16 Other nodes	17.27%PL
GCE 2	P5	6+3	31 B10,B11,B12,B13,B32 B1,B4,B5,B7,	7.15%GL
			B8,B9,B39 B3,B17,B18 B2,B25,B26,B27, B28,B29,B30,B37,B38 Other buses N15,N16 N5,N6,N7 Other nodes	

All buses and gas pipelines are traversed to investigate the vulnerability metrics of each PCE and GCE under the given initial state, as shown in Fig. 5 and Fig. 6. It can be found that the initial failures of some buses or gas pipelines will not cause cascading failure or any load loss (e.g., buses B5, B17 and pipelines P11 and P13). There are at most 6 PCEs and 3 GCEs in the cascading failures of all buses and gas pipelines, and the bus or gas pipelines with a larger number of PCEs and GCEs have a higher vulnerability in general (e.g., buses B24, B31 and pipelines P21, P25). However, some buses or gas pipelines directly connected with power load or gas load (e.g., buses B3, B20, B39 and pipeline P19) with only one PCE or one GCE may still cause a large $LOSS_P$ or $LOSS_G$. Finally, according to the cumulative results of $LOSS_P$ and $LOSS_G$, the vulnerability of each bus and gas pipeline can be obtained corresponding to $LOSS_P$ and $LOSS_G$, respectively.

B. The proposed machine learning model on the 39-bus-20-node system

The vulnerability analysis is performed by the proposed hybrid machine learning method for different initial states on the 39-bus-20-node system, and the following steps are constructed to get sufficient data: 1) Data sets with 10,000 initial states are obtained by randomly changing the power and gas load levels, generator outputs, gas well supply and consumption of P2G devices (i.e., 40 features in total) within the range of 0.8-1.2 times of the rated initial state; 2) A CFS is constructed for each data set to calculate the number of PCEs and GCEs and the vulnerability metrics. Taking bus B6 as an example, the distribution of vulnerability metrics data is depicted in Fig. 7. There are at most 5 PCEs and 3 GCEs in cascading failures for the bus B6 under different initial states. The vulnerability metrics of the bus B6 are significantly different under different initial states, and with the number of PCEs and GCEs increasing, the vulnerability metrics under the cascading failure increase sharply and obviously present a hierarchical distribution.



This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3119237, IEEE Transactions on Power Systems

Fig. 7 Data distribution of vulnerability metrics under different initial states

According to the data distribution of bus B6 shown in Fig. 7, we take the number of PCEs and GCEs as the classification label, calculate the importance of features, and conduct the RF classification by increasing the number of features according to their importance. The classification results are shown in Fig. 8. By utilizing the RF classification method with an optimal number of features (i.e., 24 for PCE and 22 for GCE), we can classify the data with an accuracy of 97.4% and 98.3%, and compared with all features (i.e., 40), the F1 score is improved by 8.9% for PCE and 11.5% for GCE, respectively.

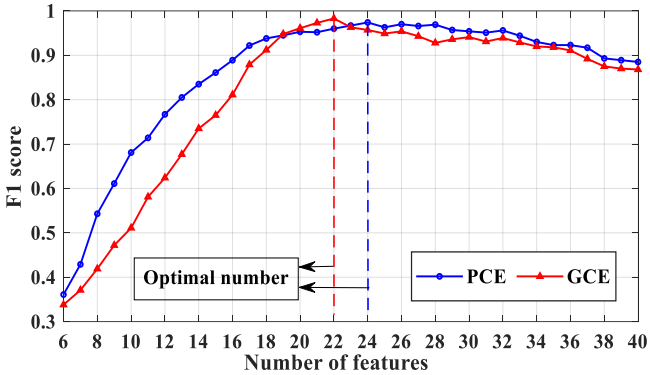


Fig. 8 Performance of RF classification with different number of features

According to the classification results, RF feature selection and regression are carried out for each class of data. The results of the proposed combined method of RF classification (RFC) and RF regression (RFR) with PCEs=2 and GCEs=2 and the results of the direct RFR method are shown in Fig. 9. We found that selecting the optimal number of features by feature importance calculation improves the accuracy of regression. For example, for the data with PCEs=2, the regression accuracy with the optimal number of features (i.e., 19) is improved from 92.3% to 98.4% in comparison to all features (i.e., 40). In addition, compared with the direct regression method, the proposed machine learning method greatly improves the regression prediction accuracy of vulnerability metrics. The accuracy of the proposed method for GCEs = 2 (i.e., 99.3%) is almost two times higher than the direct regression method. This is because the RF classification can effectively reduce the impact of uneven distribution of data on the prediction accuracy.

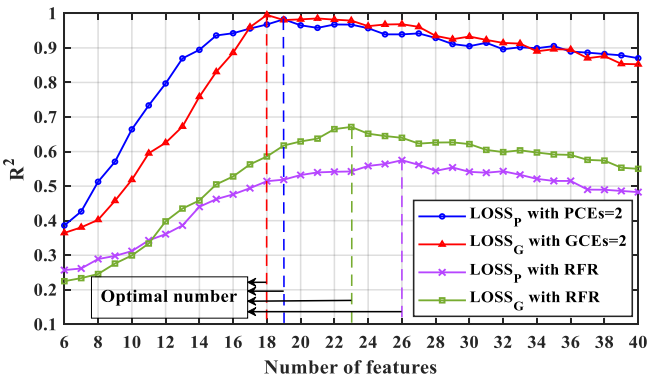


Fig. 9 Performance of RF regression with different number of features

All buses and gas pipelines are traversed to get their corresponding vulnerability prediction model by the proposed method. For a certain initial state in a real integrated power-gas system, the proposed machine learning method is used to quickly predict the vulnerability metrics of each bus and pipeline online. Then, take the components with high vulnerability metrics as the vulnerable critical components in this operation state, and assist the planners and operators to take corresponding measures to enhance the resistance of those vulnerable critical components.

C. Comparisons on other IPGSs

The proposed machine learning method is also applied to a 73-bus-40-node system [11], a 300-bus-135-node system and a 2383-bus-582-node system [42]. The results of average accuracy and average computation time for all buses and gas pipelines for each IPGS are shown in Fig. 10. Compared with the direct regression method, we found that the proposed method improves the final regression accuracy significantly by 40.4%, 52.1%, 59.7% and 60.2% for $LOSS_P$ and 37.2%, 48.7% and 51.2% 53.7% for $LOSS_G$ in the four IPGSs respectively. Moreover, the proposed method can be applied to large-scale IPGSs. Larger IPGSs have more obvious improvements while sacrificing computation time for the RF classification.

Finally, in order to verify the effectiveness of the proposed machine learning method, some other classic machine learning algorithms are employed here, including three classification algorithms: the decision tree classification (DTC) algorithm, the support vector machine classification (SVC) algorithm and the extreme gradient boosting classification (XGBC) algorithm, and three corresponding regression algorithms: the decision tree regression (DTR) algorithm, the support vector machine regression (SVR) algorithm and the extreme gradient boosting regression (XGBR) algorithm. All algorithms select the optimal feature set based on the same RF feature importance calculation method. In order to guarantee a fair comparison, the grid search method is employed to get all possible combinations of the important parameters affecting each algorithm. Then, the corresponding algorithm is used to train data with all parameter combinations to obtain the best parameter combination with the highest prediction accuracy.

The average accuracy and average computation time of $LOSS_P$ and $LOSS_G$ are shown in Table II. We found that the average accuracies of the four direct regression methods are less than 65%. However, the hybrid methods with combined classification and regression greatly improve the regression accuracy at the cost of increased computation time. For example, the accuracy of the four hybrid methods increased by 27.2%, 35.6%, 32.4% and 52.1% on the 73-bus-40-node system, respectively. The accuracies of the hybrid methods based on the SVR and DTR algorithms are no more than 85%, which is far lower than the RF algorithm utilized in this paper. Compared with the RF algorithm integrating multiple decision trees, the algorithm structures of SVR and DTR are relatively simple. They are difficult to deal with the overfitting problem in the process of classification and regression under multi-dimensional features. Therefore, the calculation time of the hybrid methods based on SVR and DTR is short, but the prediction accuracy is low. Furthermore, compared with RF

algorithm, the decision trees of XGBR is not independent. XGBR algorithm needs to learn the residual of the previous tree and minimize the objective function to generate a new decision tree. The algorithm structure of XGBR is more complex. Therefore, the hybrid method based on XGBR also achieves a high accuracy, but the computation time increases steeply in comparison with the other algorithms (i.e., more than twice as long as RF). Therefore, the proposed method in this paper selects the RF classification and regression method to predict the vulnerability metrics under different initial states in the cascading failure of IPGSs.

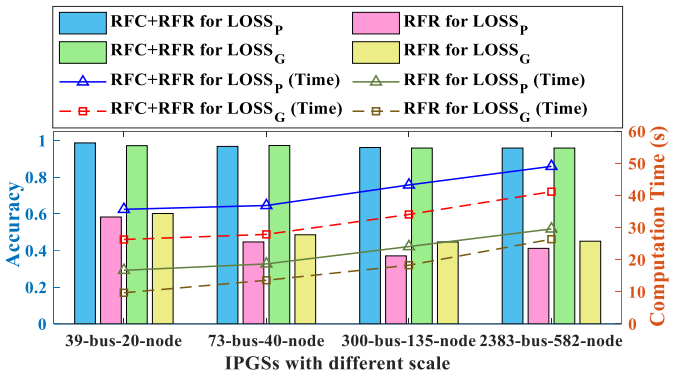


Fig. 10 Comparison of accuracy and computation time on several IPGSs

V. Conclusions

The recent February 2021 Texas winter blackout called the need of investigating the mutual impact and potential cascading failures among different energy infrastructures like electric power systems and natural gas networks. With this motivation, this paper proposed a CFS method for vulnerability analysis of IPGSs and a hybrid machine learning method to investigate the impact of random initial states on vulnerability metrics. The CFS method considering generator and gas well ramping, transmission line and gas pipeline tripping, island issue handling and load shedding strategies was conducted to study propagation process of cascading failures between two system, and generate data for machine learning with initial states randomly sampled. A hybrid machine learning model with the combined RF classification and regression algorithms and feature selection method was established to quantify the relationship between the initial states and the vulnerability metrics. Simulation results showed that the proposed hybrid method can significantly improve learning accuracy in comparison to the direct regression method, while slightly sacrificing the computation time. Furthermore, the proposed hybrid method based on RF algorithms presented better performance than other algorithms considering accuracy and computation time comprehensively.

TABLE II Comparison between other machine learning methods with the proposed method

Methods	39-Bus-20-Node System		73-Bus-40-Node System		300-Bus-135-Node System		2383-Bus-582-Node System	
	Accuracy	Time (s)	Accuracy	Time (s)	Accuracy	Time (s)	Accuracy	Time (s)
DTR	39.3%	9.8	37.7%	10.6	29.8%	12.5	27.0%	21.4
SVR	42.8%	15.8	40.7%	19.7	42.2%	21.0	43.1%	26.2
XGBR	52.3%	53.1	63.5%	68.1	57.2%	90.3	58.5%	115.3
RFR	58.6%	16.7	44.7%	18.2	37.1%	24.1	44.3%	27.1
DTC+DTR	65.6%	21.3	61.9%	25.0	57.0%	27.2	57.0%	33.4
SVC+SVR	79.1%	29.3	76.3%	31.1	73.6%	39.2	73.6%	50.2
XGBC+XGBR	96.7%	89.8	95.9%	109.7	96.9%	135.9	96.9%	159.8
RFC+RFR	98.7%	35.7	96.8%	36.9	96.2%	43.3	96.2%	49.1

References

[1] A. Quelhas, E. Gil, J. D. McCalley and S. M. Ryan, "A Multiperiod Generalized Network Flow Model of the U.S. Integrated Energy System: Part I—Model Description," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 829-836, May 2007.

[2] T. Ding, Y. Hu and Z. Bie, "Multi-Stage Stochastic Programming With Nonanticipativity Constraints for Expansion of Combined Power and Natural Gas Systems," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 317-328, Jan. 2018.

[3] K. A. Pambour *et al*., "SAInt - A novel quasi-dynamic model for assessing security of supply in coupled gas and electricity transmission networks," *Appl. Energy*, vol. 203, pp. 829-857, Oct. 2017.

[4] C. M. Correa-Posada and P. Sánchez-Martín, "Integrated Power and Natural Gas Model for Energy Adequacy in Short-Term Operation," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3347-3355, Nov. 2015.

[5] T. Ding, Y. Xu, W. Wei and L. Wu, "Energy Flow Optimization for Integrated Power–Gas Generation and Transmission Systems," *IEEE Trans. Industr. Inform.*, vol. 16, no. 3, pp. 1677-1687, Mar. 2020.

[6] S. Chen, Z. Wei, G. Sun, K. W. Cheung and Y. Sun, "Multi-Linear Probabilistic Energy Flow Analysis of Integrated Electrical and Natural-Gas Systems," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1970-1979, May 2017.

[7] S. Clegg and P. Mancarella, "Integrated Modeling and Assessment of the Operational Impact of Power-to-Gas (P2G) on Electrical and Gas Transmission Networks," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1234-1244, Oct. 2015.

[8] Z. Bao, Q. Zhang, L. Wu and D. Chen, "Cascading Failure Propagation Simulation in Integrated Electricity and Natural Gas Systems," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 5, pp. 961-970, Sept. 2020.

[9] Z. Zhang, T. Ding, Q. Zhou, Y. Sun, F. Chi, "A review of technologies and applications on versatile energy storage systems," *Renew. Sust. Energ. Rev.*, vol. 148, no. 3, 2021.

[10] D. Hagan *et al*., "The Texas Blackout: A Regulatory Retrospective and Future Outlook," Mar. 16, 2021. [Online]. Available: https://www.whitecase.com/publications/alert/texas-blackout-regulatory-retrospective-and-future-outlook.

[11] C. Shao, M. Shahidehpour, X. Wang, X. Wang and B. Wang, "Integrated Planning of Electricity and Natural Gas Transportation Systems for Enhancing the Power Grid Resilience," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4418-4429, Nov. 2017.

[12] T. Li, M. Eremia and M. Shahidehpour, "Interdependency of Natural Gas Network and Power System Security," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1817-1824, Nov. 2008.

[13] Z. Zeng, T. Ding, Y. Xu, Y. Yang and Z. Dong, "Reliability Evaluation for Integrated Power-Gas Systems With Power-to-Gas and Gas Storages," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 571-583, Jan. 2020.

[14] T. Ding, Y. Lin, G. Li and Z. Bie, "A New Model for Resilient Distribution Systems by Microgrids Formation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 4145-4147, Sept. 2017.

[15] T. Ding, Z. Wang, W. Jia, B. Chen, C. Chen and M. Shahidehpour, "Multiperiod Distribution System Restoration With Routing Repair Crews, Mobile Electric Vehicles, and Soft-Open-Point Networked Microgrids," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4795-4808, Nov. 2020.

[16] R. Baldick et al., "Initial review of methods for cascading failure analysis in electric power transmission systems: IEEE PES CAMS task force on understanding, prediction, mitigation and restoration of cascading failures,"*Proc. IEEE Power Energy Soc. General Meeting*, 2008, pp. 1–8.
[17] M. Papic et al., "Survey of tools for risk assessment of cascading outages," in *Proc. IEEE Power and Energy Soc. General Meeting*, 2011, pp. 1–9.
[18] M. Vaiman et al., "Risk assessment of cascading outages: methodologies and challenges," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 631–641, May 2012.
[19] W. Fan, Z. Liu, P. Hu, and S. Mei, "Cascading Failure Model in Power Grids Using the Complex Network Theory," *IET Gener. Transm. Dis.*, vol. 10, no. 15, pp. 3940-3949, Nov. 2016.
[20] M. Ouyang, "Comparisons of purely topological model, betweenness based model and direct current power flow model to analyze power grid vulnerability," *Chaos: Interdiscipl. J. Nonlin. Sci.*, vol. 23, no. 2, p. 023114, 2013.
[21] J. Song, E. Cotilla-Sanchez, G. Ghanavati and P. D. H. Hines, "Dynamic Modeling of Cascading Failure in Power Systems," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2085-2095, May 2016.
[22] R. Fitzmaurice, E. Cotilla-Sanchez, and P. Hines, "Evaluating the impact of modeling assumptions for cascading failure simulation," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2012, pp. 1–8.
[23] J. Qi and S. Mei, "Blackout model considering slow process and SOC analysis," in *Proc. 2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–6.
[24] S. Mei, F. He, X. Zhang, S. Wu and G. Wang, "An Improved OPA Model and Blackout Risk Assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 814-823, May 2009.
[25] D. Xianzhong and S. Sheng, "Self-Organized Criticality in Time Series of Power Systems Fault, Its Mechanism, and Potential Application," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1857-1864, Nov. 2010.
[26] S. Mei, Y. Ni, G. Wang and S. Wu, "A Study of Self-Organized Criticality of Power System Under Cascading Failures Based on AC-OPF With Voltage Stability Margin," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1719-1726, Nov. 2008.
[27] D. Kirschen, D. Jayaweera, D. Nedic, and R. Allan, "A probabilistic indicator of system stress," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1650–1657, Aug. 2004.
[28] I. Dobson, B. Carreras, and D. Newman, "A probabilistic loading-dependent model of cascading failure and possible implications for blackouts," in *Proc. 36th Annu. Hawaii Int. Conf. System Sciences*, 2003, 2003, vol. 19, no. 01, pp. 15–32.
[29] J. Liu, H. Zhang, W. Qiao and L. Qu, "DC (Optimal) Power Flow-Based Models for Simulation and Mitigation of Overload Cascading Failures," *2019 North American Power Symposium (NAPS)*,2019, pp.1-5.
[30] J. Yan, Y. Tang, H. He and Y. Sun, "Cascading Failure Analysis With DC Power Flow Model and Transient Stability Analysis," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 285-297, Jan. 2015.
[31] C. M. Correa-Posada and P. Sánchez-Martın, "Security-Constrained Optimal Power and Natural-Gas Flow," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1780-1787, Jul 2014.
[32] C. Shao, M. Shahidehpour, X. Wang, X. Wang and B. Wang, "Integrated Planning of Electricity and Natural Gas Transportation Systems for Enhancing the Power Grid Resilience," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4418-4429, Nov. 2017.
[33] J. Beyza, G. J. Correa-Henao and J. M. Yusta, "Cascading Failures in Coupled Gas and Electricity Transmission Systems," *2018 IEEE ANDESCON*, Santiago de Cali, pp. 1-6, 2018.
[34] J. Beyza and J. M. Yusta, "Robustness assessment of the expansion of coupled electric power and natural gas networks under cascading failures," *IET Gener. Transm. Dis.*, vol. 12, no. 21, pp. 5753-5760, Nov. 2018.
[35] M. Bao, Y. Ding, C. Shao, Y. Yang and P. Wang, "Nodal Reliability Evaluation of Interdependent Gas and Power Systems Considering Cascading Effects," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4090-4104, Sept. 2020.
[36] Z. Bao, Z. Jiang, L. Wu, "Evaluation of Bi-Directional Cascading Failure Propagation in Integrated Electricity-Natural Gas System," *Int. J. Electr. Power Energy Syst.*, vol. 121, Oct. 2020.
[37] H. Zang, *et al.* "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *Int. J. Electr. Power Energy Syst.* vol. 118, 2020.
[38] C. M. Correa-Posada and P. Sánchez-Martín, "Integrated Power and Natural Gas Model for Energy Adequacy in Short-Term Operation," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3347-3355, Nov. 2015.
[39] H. Su and T. Liu, "Enhanced-Online-Random-Forest Model for Static Voltage Stability Assessment Using Wide Area Measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6696-6704, Nov. 2018.
[40] X. Peng *et al*., "Random Forest Based Optimal Feature Selection for Partial Discharge Pattern Recognition in HV Cables," *IEEE Trans. Power Deliv.*, vol. 34, no. 4, pp. 1715-1724, Aug. 2019.
[41] D. De Wolf and Y. Smeers, "The gas transmission problem solved by an extension of the simplex algorithm," *Manage. Sci.*, vol. 46, no. 11, pp. 1454–1465, Nov. 2000.
[42] M. Schmidt, *et al*., "GasLib—A library of gas network instances," *Data*, vol. 2, no. 4, pp. 40, Mar. 2017.

Shuai Li (S'20) received the B.S. degree from the School of Electrical Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently working toward the M.S. degree at Xi'an Jiaotong University. His major research interests include power system optimization and resilience.

Tao Ding (SM'19) received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2015. He is currently a Professor in the State Key Laboratory of Electrical Insulation and Power Equipment, the School of Electrical Engineering, Xi'an Jiaotong University. His current research interests include electricity markets, power system economics and optimization methods, and power system planning and reliability evaluation. Dr. Ding is an editor of IEEE Transactions on Power Systems, IEEE Power Engineering Letters, IEEE Systems Journal, IET Generation Transmission & Distribution and CSEE JPES.

Weihao Jia (S'20) received the B.S. degree from the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2020. He is currently working toward the M.S. degree at Xi'an Jiaotong University. His major research interests include power system optimization and renewable energy integration.

Can Huang (S'13-M'16-SM'18) received the B.S.E.E degree from Hohai University, Nanjing, China, in 2008, the M.S.E.E. degree from Southeast University, Nanjing, China, in 2011, and the Ph.D. degree in Electrical Engineering from the University of Tennessee, Knoxville, TN, USA, in 2016. Now he is a Research Staff with Lawrence Livermore National Laboratory, Livermore, CA, USA. His current research interests include smart sensors, data analytics, and machine learning for energy and power systems, cyber-physical systems, and Internet of Things.

João P. S. Catalão (Senior Member, IEEE) received the M.Sc. degree from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2003, and the Ph.D. degree and Habilitation for Full Professor ("Agregação") from the University of Beira Interior (UBI), Covilha, Portugal, in 2007 and 2013, respectively. Currently, he is a Professor at the Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal, and Research Coordinator at INESC TEC. He was also appointed as Visiting Professor by North China Electric Power University (NCEPU), Beijing, China. His research interests include power system operations and planning, distributed renewable generation, power system economics and electricity markets, demand response and smart grid.

Fangxing Li (F'17) is also known as Fran Li. He received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree from Virginia Tech, Blacksburg, VA, USA, in 2001. Currently, he is the James W. McConnell Professor in electrical engineering and the Campus Director of CURENT at the University of Tennessee, Knoxville, TN, USA. His current research interests include renewable energy integration, demand response, distributed generation and microgrid, energy markets, and power system computing. Since 2020, Prof. Li has been serving as the Editor-In-Chief of *IEEE Open Access Journal of Power and Energy (OAJPE)* and the Chair of IEEE PES Power System Operation, Planning and Economics (PSOPE) Committee.